

The Use of Diaries as a Qualitative Research Method to Investigate Teachers' Perception and Use of Rating Schemes

Jyi-yeon Yi

Chongshin University

Yi, J. (2008). The use of diaries as a qualitative research method to investigate teachers' perception and use of rating schemes. *Journal of Pan-Pacific Association of Applied Linguistics*, 12(1), 1-10.

Whilst diary study is used for pedagogical purposes, course evaluation and basic research on language learners, this study aims to explore the possibility of using it to investigate how teachers perceive and use rating schemes. Three English teachers who worked at various high schools in Korea rated 224 scripts written by 112 Korean high school students for this study. The teachers assessed the scripts twice, first according to their subjective holistic scoring and then using the FCE scale for writing assessment, and they kept diaries on their rating process for each assessment. The analysis of their diaries shows the teachers' rating patterns and tendencies, problems with the rating schemes and their understanding of the rating schemes. It can be concluded from these findings that diaries can be employed with regard to assessment, that is to reveal raters' perception of rating schemes, to investigate the validity of the assessment and to identify the aid and guidance that they might need for assessment.

Key Words: rating schemes, qualitative research method, diary study, validation

1 Introduction

1.1 Definition of diary

Diary studies have attracted attention from researchers who are interested in gathering qualitative data, especially since Bailey (1983), as reviewed in Howell-Richardson and Parkinson (1988). Krishnan and Lee (2002) define diaries as first-person observations of experiences that are recorded over a period of time. While they refer to such records as diaries (e.g. Bailey, 1983; Howell-Richardson & Parkinson, 1988; Parkinson *et al.*, 2003) which tend to be associated with 'confessions' or 'baring the soul' "highlighting the unavoidable tension between writing a record of personal relevance and having it read by a tutor" (Jarvis, 1992: 135), some other researchers prefer to call them journals (e.g. Krishnan & Lee, 2002) or records (e.g. Jarvis, 1992) which is related with 'public' consumption because they are designed to be read by others. Regardless of what these recordings of his or her thought, feeling and reflection are called, the studies using these recordings are classified as one of ethnographical studies in that they are intended to

reveal existing phenomenon and generate hypothesis (Bailey, 1983; Woodfield & Lazarus, 1998). This study will use the term “diary” as it appears to be more widely employed.

1.2 Background and purpose of the study

Howell-Richardson and Parkinson (1988) observe that the literature on diary studies shows that they can be used for up to sixteen different purposes, which can be categorised into three main groups: pedagogical purposes, course evaluation and basic research. In the first group, diaries can be used as effective channels of communication between teachers or trainee teachers and learners discussing their language learning process or lessons given by the teachers/trainee teachers. This type of diary is written by learners (e.g. Bailey, 1983; Jarvis, 1992; Parkinson *et al.*, 2003), teachers as learners of a language (e.g. Woodfield & Lazarus, 1998), or interactively by both teachers/trainee teachers and learners (e.g. Gray, 1998). The general aim of diary-keeping in this context is to help learners to be aware of how they learn. Reflection by teachers can make them reflect on their language learning process or teaching methods/experiences, establish links between theory and practice in the learning and teaching of second languages, and make changes in their teaching methods. Finally, the interactive diaries written by both trainee teachers and learners benefit both parties: providing the trainee teacher with valuable feedback on their teaching that can help them plan effective classes, and giving learners the opportunity to reflect on their learning process.

The aim of the second purpose, course evaluation, is to be useful to both course developers and teachers. It includes “attempting to re-balance group dynamics by moving students (between or within classes)” and “evaluation decisions taken at course-director level, including change of teacher” (Howell-Richardson & Parkinson, 1988: 75). Krishman and Lee (2002), who administered a diary study for this purpose, found that learners expressed more anxiety when they moved from their home country to the host country; had a language agenda in that they had ideas about what they wanted to learn; and that their attitude towards learning was affected by the learning environment (teacher, other learners, activities and materials). Therefore, the study suggested that course developers and teachers should design courses and adjust activities and materials to meet the learners’ needs and expectations. Halbach (1999) and Jarvis (1992) also used diaries for this purpose, particularly for the evaluation of a teacher training course; noting that diaries can enable teachers to reflect on the theory of second language learning suggested in the course and link it with their own learning experience.

The third purpose involves using diaries for basic research, discovering, among other things, what language learners do outside class, how they feel in terms of learning-related anxiety, and what they remember from their class (Howell-Richardson & Parkinson, 1988).

It is, however, worth noting that diaries could be kept by teachers for *other* purposes than those mentioned above. One of them would be in regard to assessment. This study, therefore, attempts to explore such a possibility: the use of

The Use of Diaries as a Qualitative Research Method

diaries as a qualitative research method for investigating one of assessment-related aspects, i.e., how teachers perceive and use rating schemes for assessment. Through revealing raters' perception and use of rating schemes, this study aims to suggest diary study as one of ways to investigate the validity which is the central issue in testing and to identify the aid and guidance that they might need for assessment.

This study intends to tackle this question across two cases with regard to the use of rating schemes: assessing according to their own judgment without any formal rating scales; assessing using a rating scale provided with.

2 Study

2.1 Methodology of the study

The scripts for this study were obtained by asking 112 high school students in 1st and 2nd year at a foreign language high school in Korea to do two writing tasks (an informal letter to a foreign friend suggesting places to visit in Korea, and a formal essay explaining the advantages and disadvantages of using the Internet). Their teachers voluntarily helped obtain these scripts between September and October 2003. The tasks and topics were decided on the basis of the content of the English Writing course at their school, and the scripts obtained were typed up by the researcher and then handed over to three teachers as raters for this study.

The three teachers in this study, known Teachers A, B and C, were Korean teachers of English at different high schools. They had worked as English teachers for twelve years, five years and thirty-four years respectively, and had never acted as professional raters.

The workload was adjusted for each teacher, so Teachers A and B kept diaries on the assessment of eighty scripts and Teacher C on sixty-four scripts. They were invited to use two scoring schemes for the assessment: subjective holistic scoring¹ and the FCE rating scale for writing assessment (FCE Handbook, 2001).² For subjective holistic scoring, the teachers were allowed to use their own assessment features and criteria for each band, provided there was a total of six bands in the scoring scheme, with Band 6 as the highest. The FCE rating scale, on the other hand, was specifically designed for the FCE writing assessment. These two kinds of rating schemes were chosen to investigate the possibility of the use of diaries across the two rating schemes.

As the teachers had never been asked to keep diaries on an assessment, they were provided with guidelines and instructions on what they should include and

¹ According to Hamp-Lyons (1991), subjective scoring has two types: either using a holistic scale; or assessing according to a teacher's own subjective judgement without any formal holistic scale, of which the former is more popular than the latter nowadays. To differentiate between two, I will name the latter as subjective holistic scoring in this study.

² Given that the FCE test was devised for intermediate level EFL/ESL learners, and that the students from whom the scripts were obtained were generally around this level, the scale was chosen for the students in this study.

focus on, and given some of the most informative diaries that they had kept during the trial stage as a reference point. They were asked to make a diary entry just after assessing and scoring each script.

2.2 Data analysis

I followed the process outlined in the literature (e.g. Halbach, 1999; Lakshmy & Lee, 2002), reading the obtained diaries and trying to find salient features and patterns in them. Some patterns were found in relation to the purpose of this study, and grouped depending on what diaries could reveal under the following three headings: rating patterns and tendencies; problems with the employed rating schemes; teachers' understanding of the rating scheme.

2.3 Findings

2.3.1 Rating patterns and tendencies

It was found that diaries could reveal which assessment features the teachers focused on in the case of subjective holistic scoring. For example, all the teachers in this study paid attention to content and grammar, as can be seen in (1.1). They also considered different features depending on the proficiency level of the scripts: for example, length and/or grammar when assessing low-level scripts such as Band 2 scripts, (see (1.2)), organization and intelligibility for intermediate level scripts in Bands 3 and 4 (see (1.3)), and expressions and sentence structure that looked natural and like native English for Bands 5 and 6 (see (1.4)). They also showed central tendency in rating, that is, avoiding assigning the lowest (i.e. Bands 1 and 2) and highest (i.e. Band 6) bands (see (1.5)). When using the FCE rating scale, they sometimes considered their own subjective criteria even though they were using the scale. For example, Teacher A considered 'paragraphing' and 'balance between paragraphs', although they were explicitly not included in the scale (see (1.6)).

(1.1)

Band 3

This script could be good because it is fairly long and well organized. But as I looked at it more closely, I found a major error in it. It is that this script deals with the myth of Ulsan Rock rather than the required content from the prompt--places in Korea which are worth visiting. Additionally, it has basic grammatical errors and does not include connectors which can be seen in other students' scripts. Therefore, I think Band 3 is most appropriate for this script.³

³ All the diary entries obtained for this study were originally made in Korean. I have translated them into English.

The Use of Diaries as a Qualitative Research Method

(1.2)

Band 2

There are grammatical errors in almost every clause. Most of them are caused by the use of inappropriate vocabulary, and consequently there are many clauses which are unintelligible. Additionally, the writer makes errors in the use of verbs. That is, since he/she does not know the exact meaning of the verbs he/she makes errors by either using a preposition wrongly or omitting it after a verb. What is worse, it is too short. I suppose there are few aspects that would help this put script in a good band.

(1.3)

Band 3

This script is at the very middle level, I think. Whilst communication is relatively good and it is well organized on the whole, it has errors in terms of differentiating between singular vs. plural, tense and word classes. The errors, however, are local rather than global, so they do not affect the communication of clauses. So I marked it as Band 3.

(1.4)

Band 6

In addition to the content, this script is absolutely excellent. The use of vocabulary and idioms is very fluent and at university level. Although there are errors in the use of idioms like “take into account” and “make good use of”, they appear to be mistakes. Generally the writer has a good command of advanced expressions.

(1.5)

Band 3

Although the writer seemed to try to write smoothly as his/her thoughts flowed, what s/he has actually produced is a script that looks illogical. What is worse, the writer does not attend to paragraphing and organisation at all. The length is also insufficient and it is filled with many grammatical errors. I marked it as Band 3, but to be honest, I would like to put it in the lower band, Band 2.

(1.6)

Band 4

The most noticeable characteristic of this script is that there is no paragraphing.....

Band 4

When I had a closer look at this script, I found that as the argument proceeded, the points in the latter part got shorter and shorter. The first two paragraphs look good, but the third is shorter and not developed

enough, and the fourth is even worse. These last two paragraphs don't look sufficiently developed. I suppose it might be because the writer wanted to finish more quickly, but it has meant that the writer did not maintain a balance between paragraphs.

2.3.2 Problems with the rating schemes

The diaries helped reveal the problems with each rating scheme. In this study, when the teachers were assessing according to their own subjective criteria, they sometimes considered an aspect of the scripts that did not correspond with any recognised measure or component of writing ability (see (2.1)). It also emerged that they were not always confident of their rating, sometimes wondering if it was appropriate (see (2.2)); while their criteria for some levels, especially the highest and the lowest, were not always clearly established (see (2.3)).

As for the FCE scoring, the diaries revealed that the teachers found some assessment categories and descriptors inappropriate for the specific test-takers and context. They found one of the assessment categories in the scale unnecessary (Register), since intermediate level test-takers such as Korean high school students do not command a variety of registers depending on language use context, but focus on making grammatical sentences *per se* (see (2.4)). The teachers' diaries also revealed that they thought that some additional assessment categories should be included in the scale, such as 'length' and 'development of idea' (see (2.5)). With regard to descriptors, the teachers found that the use of vague quantifiers such as "all", "some", "little" and "limited", and ambiguous words such as "effectively", "clearly" and "inadequately" made it difficult to grasp the differences between descriptors (see (2.6)). They also pointed out that the scale lacked descriptors that could deal with frequently observed phenomena in specific test-takers' scripts, and that this made it inconvenient to use (see (2.7)). For example, many test-takers tried to use a range of vocabulary and idioms, but did so inappropriately and with awkward results, and the descriptors only included the range itself but not this observed phenomenon.

(2.1)

Band 3

I hesitated between Bands 2 and 3 for this script. It could be Band 2 in that it is only half as long as the other scripts, is poorly organised and shows poor sentence construction and grammar. There are few well-constructed sentences without grammatical errors. This notwithstanding, I assigned it Band 3, because it looks as though the writer has made an effort.

(2.2)

Band 4

If I assign Band 4 to this script, I think it would be very harsh. This is because this script is fairly good in terms of length, organisation and

The Use of Diaries as a Qualitative Research Method

content. However, its weak point is that the writer tried to use a variety of expressions and this resulted in some awkwardness. While I read through the first half of the script I thought these awkward expressions would not negatively affect the grade, but in the second half, they negatively affected my impression of it and in turn the grade, because they sometimes hindered communication and intelligibility, and the former is one of the most important assessment categories for me. However, I wonder if it is fair to mark it down merely because of a few sentences that are not intelligible, even though the script communicates well on the whole. I am not sure whether the grade I assigned is fair.

(2.3)

Band 2

This script is extremely short – just one paragraph long. And not only is it short, but it contains many grammatical errors, so I can't help but assign it to Band 2. But if I mark it as Band 2, I'm not sure which scripts deserve to be put in Band 1.

Band 6

This script is excellent - the best so far. It is fantastic in organization, genre format and paragraphing, and each point is equally developed and written very clearly. There are just a few grammatical errors. But even though it is written well, I hesitated in assigning Band 6 to this script because I'm not sure whether it is really good enough to deserve Band 6, given that Band 6 is the highest band and would mean "perfect". Even though I am unsure about this, I assigned Band 6 to this script because it is the best bit of writing I have seen so far.

(2.4)

Band 5

... Looking back on the three previous scripts, I have to admit that I assessed the category of Register without really understanding it. I wonder if there are differences between Korean students' scripts in terms of Register. I suppose they just pay attention to "making sentences" that are accurate and grammatical because of their intermediate or low level of English, so they won't have much variety of register depending on the situation given in the prompts.

So I don't think there is much difference between them in terms of Register. If this is the case, I don't think that this category is necessary in the Korean situation....

(2.5)

Band 4

.... This script is paragraphed, but it is not done appropriately. Each paragraph is too short and not developed enough. However, it looks as though there is no descriptor to deal with this kind of situation....

(2.6)

Band 4

This script introduces Ha-hoe town in Ahndong in detail. Unfortunately, it has quite a number of grammatical errors. The errors are not local but global errors, which affects my understanding of what the sentences mean. For assessment of this, I had a look at the descriptors in Accuracy, hoping to find the most appropriate band for this case. However, the words in the descriptors, such as “a number of errors” and “frequent errors” look very ambiguous to me. I cannot see the difference between them and I am not sure which would be more appropriate for this situation. As neither of them is clear to me, I just chose Band 3, according to my intuition.

(2.7)

Band 4

... As for Range, it looks like this writer tried to use a variety of words, but their uses are awkward or inappropriate. In this case, I am unsure what band to assign to this script. The rating scheme does not address this situation. Having trouble with this point, I just chose Band 4...

2.3.3 Teachers’ understanding of the rating schemes

The diaries were also helpful in revealing how the teachers understood the provided rating scheme, the FCE scale. The issue of how teachers understand the rating scale they are using needs to be investigated in terms of validity. If their understanding does not coincide with the use suggested or intended by the developers of the scale, or if some assessment categories or descriptors are not clear to the teachers, the validity of their rating and use of the ratings are likely to be questionable. The diaries were found to be helpful in investigating this point. For example, the teachers for this study found that some assessment categories were unclear, such as Appropriacy of Register and Format, Range and Target Reader, as can be seen in (3.1).

(3.1)

Band 4

...The most peculiar error in this script is that it is not well-paragraphed and is written in a very colloquial style. Since this is the case, I suppose that marks will have to be cut in terms of both Accuracy and Register, but I am not sure. Still, I am unsure about the category of Register.

3 Conclusions

This study has shown that diaries can be used to reveal various aspects with regard to assessment, specifically rating behaviour and raters’ perception. Through recording teachers’ internal thought processes during assessment, it revealed how

The Use of Diaries as a Qualitative Research Method

they understood the rating schemes and found them to use, and what kind of problems exist in their rating and the rating schemes themselves.

Given these findings, it can be said that diaries can be adopted for assessment purposes, in addition to the purposes mentioned in the literature: for pedagogical purposes, for course evaluation and for basic research. The investigation of teachers' perceptions and use of rating schemes could highlight any problems rating schemes may have, either before they are to be introduced, or when existing rating schemes in use need to be examined. This is desirable in light of two aspects: in investigating the validity of assessment, which has become a main concern in assessment since the 1980s; and in identifying the aid and guidance that teachers may need when using a rating scheme in question.

The former is possible, given that diaries could reveal whether a rating scheme in question is appropriate and valid for the specific context, teachers and learners. Therefore, it seems reasonable to suppose that diaries could be employed as a possible validation method, along with the other methods suggested in other studies (e.g. Fulcher, 2003), such as think-aloud protocols, interviews and questionnaires.

As for the latter, teachers' diaries enable rating scale developers to develop or revise manuals, descriptors and assessment categories for a rating scheme so as to aid teachers. They also enable teachers to realize the problems with their subjective holistic scoring in terms of concerns about the validity of assessments.

It can be concluded, therefore, that diary study could be one of useful research methods with regard to assessment. Further, given such applicability to assessment field, diary study could be extended to various purposes which need qualitative data, as well as language learning, teaching and assessment.

References

- Bailey, K. (1983). 'Competitiveness and anxiety in adult second language learning: looking at and through the diary studies' in H. W. Seliger, & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp.67-102). Rowley, Mass.: Newbury House Publishers.
- FCE Handbook. (2001). Cambridge: University of Cambridge ESOL Examinations.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow: Pearson Education Ltd.
- Gray, J. (1998). 'The language learner as teacher: the use of interactive diaries in teacher training'. *ELT Journal*, 52(1), 29-37.
- Halbach, A. (1999). 'Using trainee diaries to evaluate a teacher training course'. *ELT Journal*, 53(3), 183-190.
- Hamp-Lyons, L. (1991). 'Scoring procedures for ESL contexts' in L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp.241-278). Westport, Conn.: Ablex Publishing.

Jyi-yeon Yi

- Howell-Richardson, C., & Parkinson, B. (1988). 'Learner diaries: possibilities and pitfalls' in P. Grunwell (Ed.), *Applied linguistics in society* (pp.74-79). London: CILT/BAAL.
- Jarvis, J. (1992). 'Using diaries for teacher reflection on in-service courses'. *ELT Journal*, 46(2), 133-143.
- Lakshmy, A., & Lee, H. (2002). 'Diaries: listening to 'voices' from the multicultural classroom'. *ELT Journal*, 56(3), 227-239.
- Parkinson, B., Benson, C., & Jenkins, M. (2003). 'Learner diary research with 'Cambridge' examination candidates'. *Edinburgh Working Papers in Applied Linguistics*, 12, 45-63.
- Woodfield, H. & Lazarus, E. (1998). 'Diaries: a reflective tool on an INSET language course'. *ELT Journal*, 52(4), 315-322.

Jyi-yeon Yi
Department of English Education
Chongshin University
31-3 Sadang-dong, Dongjak-gu, Seoul, 156-763, Korea
Tel: 82-2-3479-0355
E-mail: jyyi@chongshin.ac.kr

Received: February 27, 2008
Revised: May 21, 2008
Accepted: June 30, 2008