RETROSPECTIVE DATA COLLECTION: CAN STUDENTS REMEMBER?

Paul Leeming

*School of Economics at Kinki University, Japan*

email: [paulleeming@yahoo.com](mailto:paulleeming@yahoo.com)

Article received: 21 October 2014 Final proof received: 10 December 2014

**Abstract:** It is not always possible or practical to gather data over a long period of time, and researchers have used retrospective data from questionnaire and interview to gain insights into change over time. There has been little discussion within SLA as to the validity of this data, and it is seemingly accepted at face value. This paper examines both qualitative and quantitative retrospective accounts of changes in self-efficacy, by collecting quantitative questionnaire data over a one-year period, and then asking students to quantitatively recollect their feelings of self-efficacy over that time frame, followed by qualitative description. Results showed that students were able to accurately recollect quantitative data, although there was a noticeable degradation in accuracy with time. Interview data suggested that students were reluctant to admit lack of ability to recall events, and sometimes produced conflicting accounts of events. This may in part be due to the dual role of teacher and researcher adopted in this study. The implications for researchers and also for teachers are considered.

**Keywords**: retrospection, self-efficacy, interview

Self-efficacy has been shown to be a strong predictor of performance both in general psychology (Bandura, 1997), and within SLA (Mills, Pajares & Herron, 2007; Hsieh & Kang, 2010). When measuring self-efficacy, as with other individual difference variables, researchers can ask students to report on their current feelings, or for a retrospective account. With the difficulties of collecting longitudinal data, retrospective accounts are an efficient and cheap alternative, and a number of studies within SLA have employed this as a method of data collection (Nikolov, 2001; Ushioda, 2001; Hayashi, 2005; Shoaib & Dörnyei, 2005). Students have been asked to recall feelings from up to 10 years prior to the data collection, and yet this data is accepted without detailed discussion of issues of validity. With the clear benefits of retrospective data collection, it is important for researchers to determine the reliability of retrospective accounts, and this study set out to determine the accuracy with which students recall data gathered over a one-year period. Self-efficacy was measured by questionnaire eight times over a one-year period, at the end of which interviews were conducted and students were asked to recall their feelings over that time period. After introducing self-efficacy, I present a summary of research in SLA that has used retrospection, before describing research in general psychology considering the accuracy of retrospective data. Finally I describe the present study and results are presented and discussed.

*Self-efficacy*

Self-efficacy was first introduced by Bandura (1997) to describe human agency. He defined it as “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (p. 3). Bandura differentiated self-efficacy from more general self-concepts such as self-confidence and self-esteem by claiming that these are general traits with limited predictive power, whereas self-efficacy relates to specific behaviors and is therefore far more powerful in predicting performance.

A number of studies within SLA have investigated the relationship between self-efficacy and performance. There have been issues with measurement, with studies measuring anxiety, motivation, and personal beliefs and not simply self-efficacy (Hsieh & Kang, 2010), leading to problems with interpretation of results. Mills, Parajes, and Herron (2007) found that self-efficacy related to students’ belief in their own ability to regulate their own learning was the strongest predictor of success on a French language course. In a Korean context Hsieh and Kang (2010) examined the relationship between attribution and self-efficacy. Results suggested that students high in self-efficacy were more likely to attribute success to internal factors, and self-efficacy was positively correlated with achievement measures. From this and other research we can conclude that self-efficacy is an important variable within SLA. In the current study I define oral English self-efficacy as a student’s belief in their ability to communicate using English.

*Retrospective studies within SLA*

Although cross-sectional research designs offer information about the current proficiency or motivation of students, researchers are interested in growth or change, and attempting to explain possible reasons for it. Longitudinal data collection presents a number of problems including high cost in terms of time and money, and also attrition from the study reducing the final *n* size (Featherman, 1980). A number of studies within SLA have employed a retrospective methodology in order to gain a profile of language learners over a prolonged period (Nikolov, 2001; Ushioda, 2001; Hayashi, 2005; Shoaib & Dörnyei, 2005).

Ushioda (2001) argued that there is a need for more qualitative data concerning motivational attitudes of students with regards to language learning, and conducted interviews with 20 students of French as a foreign language, first using unstructured interviews to determine the students’ conceptions of motivation, and then one year later using structured interviews to investigate student motivational change over time. One example question from the structured interview is ‘How would you describe your present state of motivation for learning French, and have you experienced any motivational changes over the past year?’ (Ushioda, 2001, p. 99). Students generally cited particular events that had positive or negative influence on their motivation, such as failing a course exam, or having a personal relationship with a native speaker. Ushioda (2001) used these accounts to discuss changes in motivation over time, listing factors that influenced motivation in both positive and negative ways. Although overall the design of the study was longitudinal, interviews were used to track changes over a year, and were therefore retrospective.

In a similar vein, Shoaib and Dörnyei (2005) reported on a study using interview data from 25 non-native learners of English living in England, who were currently studying to improve their English proficiency. Interviews were semi-structured and lasted between 15-20 minutes. After questions related to their English learning histories, interviewees were asked how their motivation had changed over the years they had been studying. The interviews concluded by focusing on motivational changes within the last year. The data was analyzed looking for patterns and attempting to find factors that influenced motivational change throughout the participants’ language learning experiences. Shoaib and Dörnyei (2005, p. 36) explained the factors that had influenced interviewees, and concluded that ‘taking a biographical/autobiographical approach has been a surprisingly positive and fruitful experience.’ The use of the word ‘surprising’ suggests that the researchers were not confident in the research methodology.

Nikolov (2001) was interested in unsuccessful language learners, and interviewed 94 adults in an attempt to discover reasons for their relative failure in learning a foreign language. Interviewees were asked about their lifelong experiences of language learning including activities, exams, teachers, and travel abroad. Nikolov (2001) then attempted to determine the role of these factors in shaping language learning success. One of the limitations provided by Nikolov (2001, p. 167) was that ‘data collection was cross-sectional, and insights into development over time were gained only through retrospective recall.’ This is not listed as a weakness in the two studies mentioned previously, despite the similarity in research design.

In contrast to the studies described above which used interview to gain retrospective accounts of change in students, Hayashi (2005) used a questionnaire to measure the fluctuations in motivation within Japanese ESL learners from a qualitative perspective. University students from years 1 to 4 were asked to quantitatively rate their motivation to learn English from starting junior high school until the present day. This meant a maximum recall time for 4th year students of almost 10 years. From the results Hayashi (2005) grouped the students according to motivational change. Hayashi (2005) did discuss the method of data collection and argued for the validity of retrospective data, claiming that it is cost effective and reliable.

*Retrospective data collection*

The debate concerning the accuracy of retrospective data has continued in social science research methodology for a number of years. Featherman (1980) argued that there was much to be done in understanding how reliable retrospective data was in comparison to data collected concurrently or just after an event, although he did point out that most data is retrospective in nature and that we are often considering differences in the length of time elapsed between the event and the attempted recall.

Retrospective data has several advantages over longitudinal studies that can be costly in terms of time and money, but among the problems given by Featherman (1980) are the contamination of our memories by our own ego, and the influence of the interview or recall situation where social desirability may be a factor. Bernard, Killworth, and Sailer (1984) reviewed a number of studies considering the validity of retrospective accounts and were far more critical stating that ‘on average about half of what informants report is probably incorrect’ (Bernard et al. 1984, p. 503).

Other authors have investigated potential sources of bias that may affect retrospective accounts. Focusing on medical contexts, Schwarz (2007) claimed that for studies of change over time, real-time data capture is preferable, with retrospective accounts subject to bias and driven by inferences based on respondents’ theories of the construct being considered. Schwarz (2007) showed that intensity of feelings are poorly represented in our memory, with peak and end effects being prevalent. Peak effects means that the most intense feeling may be overly influential in shaping our recollections of the past, and end effects means that the feeling at the end of the experience is likely to unduly influence our overall recall of the event. Both peak and end effects cloud our objective recall.

Ross (1989) suggests that there are two theories guiding our recall of personal attitudes and beliefs. ‘Stability’ means that a respondent considers their current attitude and believes that the attitude being considered is not susceptible to change, and therefore prior attitudes must have been similar if not exactly the same. This leads respondents to overestimate the similarity between current and past feelings and attitudes. An example given by Conway and Ross (1986) is that of racism, where we believe that our current view represents our lifelong attitude. ‘Change’ means that a respondent considers the attitude to be susceptible to change and that therefore their prior attitude must have been different to their current one. An example of this would be an adventurous attitude, which people assume is a trait associated with youth, and therefore over-estimate their own attitude when recalling past feelings of adventurousness.

The issue of retrospective accounts is considered by Tourangeau, Rips, and Rasinski in their book, *The Psychology of Survey Response* (2000). The authors state that ‘the best attested fact about autobiographical memory is that the longer the interval between the time of the event and the time of the interview, the less likely that a person will remember it’(Tourangeau et al. 2000, p. 82). They use data from a variety of studies to show that the ability to accurately recall is dependent on the kind of data being considered. The authors claim that names are commonly used and therefore salient, but we are not required to use street names so regularly and therefore this kind of data is more readily forgotten. They also argue that infrequent or atypical events are remembered more easily, and that important events are reported more accurately. Tourangeau et al. (2000, p. 97) conclude by stating that ‘memory for experience is intertwined with general knowledge; similarly the process of remembering what actually happened is inextricably bound up with the process of inferring what probably happened.’ This suggests that any retrospective account is influenced by the current knowledge base, and inference on the part of the respondent.

**METHOD**

*Participants and methodology*

The participants in this study were first year students in a private university in Western Japan (9 male, 2 female). The students were enrolled in the school of science and technology, and took compulsory English classes in reading, writing and oral communication. Students were pre-intermediate in terms of ability (Average TOEIC score 390), and generally motivation was quite low. The data for this study was gathered in the oral communication class which met once a week for 90 minutes, with a total of 14 classes in both Semester 1 and Semester 2. I was the teacher for the oral and writing classes.

*Questionnaire*

As part of a doctoral research project the 11 students in this study were asked to complete a questionnaire measuring self-efficacy with regards English language communication over the course of an academic year (see Appendix A for English version). The self-efficacy measure was designed to measure individual students’ self-efficacy with regard to the oral communication course. The items were adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) originally developed by Pintrich and de Groot (1990) to measure general academic efficacy within specific academic subjects. The measure was piloted before the study, and a Rasch analysis showed no problematic items. In the first semester items one to seven were used, but following a further Rasch analysis at the end of semester 1, item 4 was found to misfit the Rasch model (Bond & Fox, 2007), and was replaced with items eight and nine in the second semester. An example item is *I can understand what is taught in English Communication Class*, and students were required to respond on a six point Likert scale.

Table 1. Administration of SE measure

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| SE1 | SE2 | SE3 | SE4 | SE5 | SE6 | SE7 | SE8 |
| April 10  Week 1 | May 8  Week 6 | June 12  Week 10 | July 10  Week 14 | September 18  Week 1 | October 16  Week 6 | November 27  Week 10 | December 18  Week 13 |

Note: SE1-4 given in Semester 1; SE5-8 given in Semester 2

Students completed the questionnaire eight times in total over the academic year. Details of the timing of administration of the questionnaire are provided in Table 1. Dates are approximate as the students were in two different classes and therefore took the questionnaire on different dates. Each semester was 14 weeks in duration. The questionnaire was in written form and was administered during class time, either at the start or end of each class. Students were given approximately five minutes to complete the survey and all were able to complete it within the allotted time. Three different classes of students completed the survey with a total of 78 students split almost equally between classes. If a student was absent they were asked to complete the questionnaire in the following class, and all students completed all eight of the questionnaires. The survey was explained to the students in Japanese, and they were given a chance to ask questions. The survey was administered through the online site, Survey Monkey (www.surveymonkey.com), guaranteeing that there were no missing responses in the data set.

*Interviews*

At the end of the second semester the students were interviewed (January of the following year as the academic year runs from April to March in Japan) (see Appendix B for interview procedure). Students were assured that the interview was voluntary and would have no impact on their grade for the course. All interviews were scheduled following the final class, and were conducted either in empty classrooms, or my own office depending on available space. Interviews were video recorded and an IC recorder was used to provide additional clarity. Students all signed a consent form and were told of the general purpose of the interview. Although I planned to interview 12 students, one student was absent and did not contact me to reschedule the interview.

The interview was conducted in the students’ first language (Japanese). I was the interviewer and although a non-native speaker of Japanese, I am proficient in Japanese and therefore was comfortable conducting interviews in the students’ first language. The interview began with an explanation of what I wanted the students to do with the retrospection, and then I left students alone to complete the retrospective data chart (Appendix C). Appendix C shows the chart for semester 1 only although students actually completed charts for both semesters. Students were required to think back over the year and to recall their feelings of self-efficacy at the eight times that they completed the actual questionnaire presented above. The students were given approximately five minutes to complete the form. Subsequent interviews lasted approximately 20 minutes. I presented the retrospective sheet completed by students and asked students to explain it, before presenting the longitudinal data sheet which I had filled in prior to the interview based on their actual data. I then asked students for a subsequent explanation of any differences. Figure 1 shows the charts completed by the researcher on the left using data gathered through questionnaires, and by the student (Kuni) on the right, completed at the start of the interview. Students completed the charts retrospectively and only after explaining the changes in self-efficacy were they shown the actual data which I had completed on the chart prior to the interview.

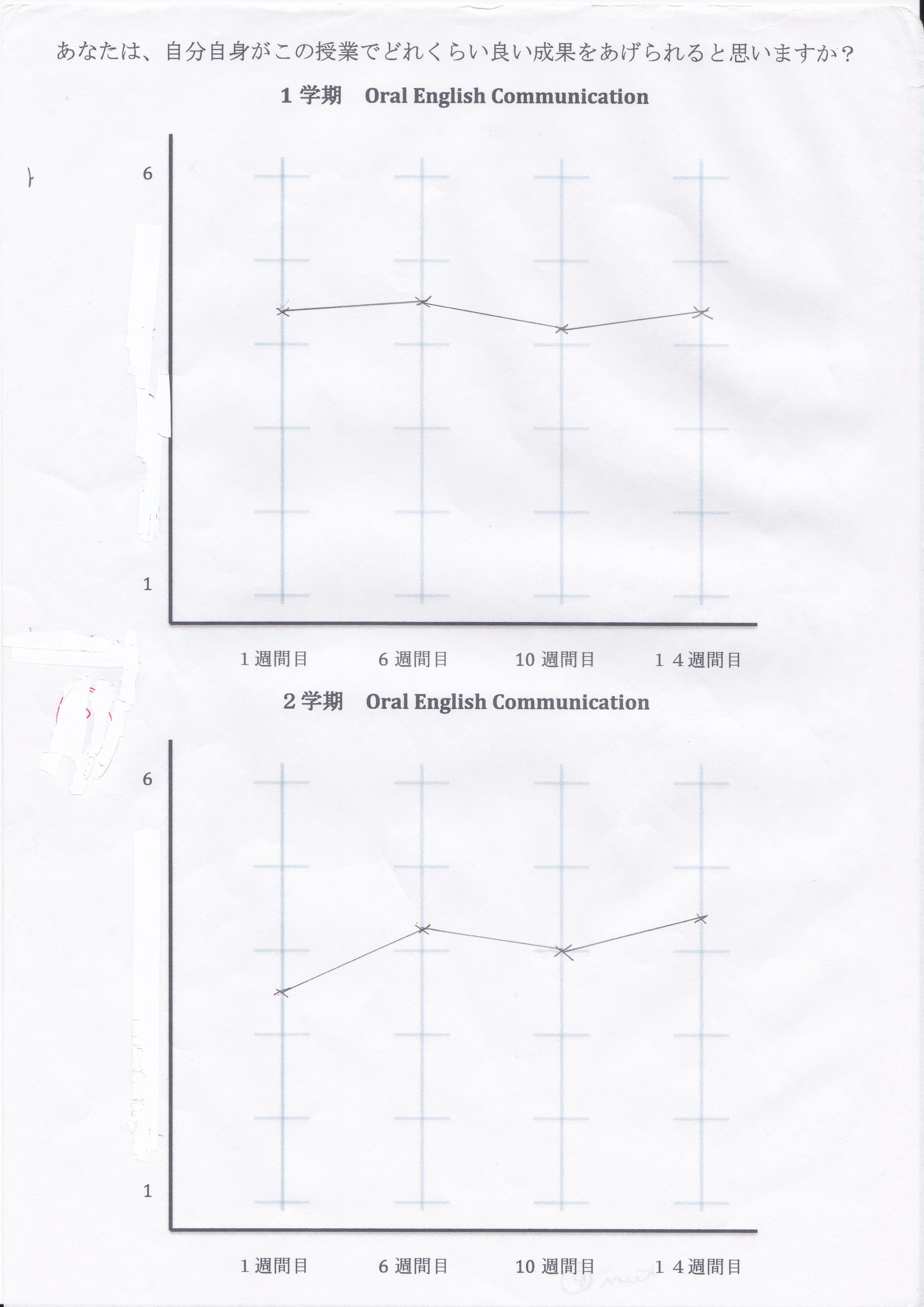
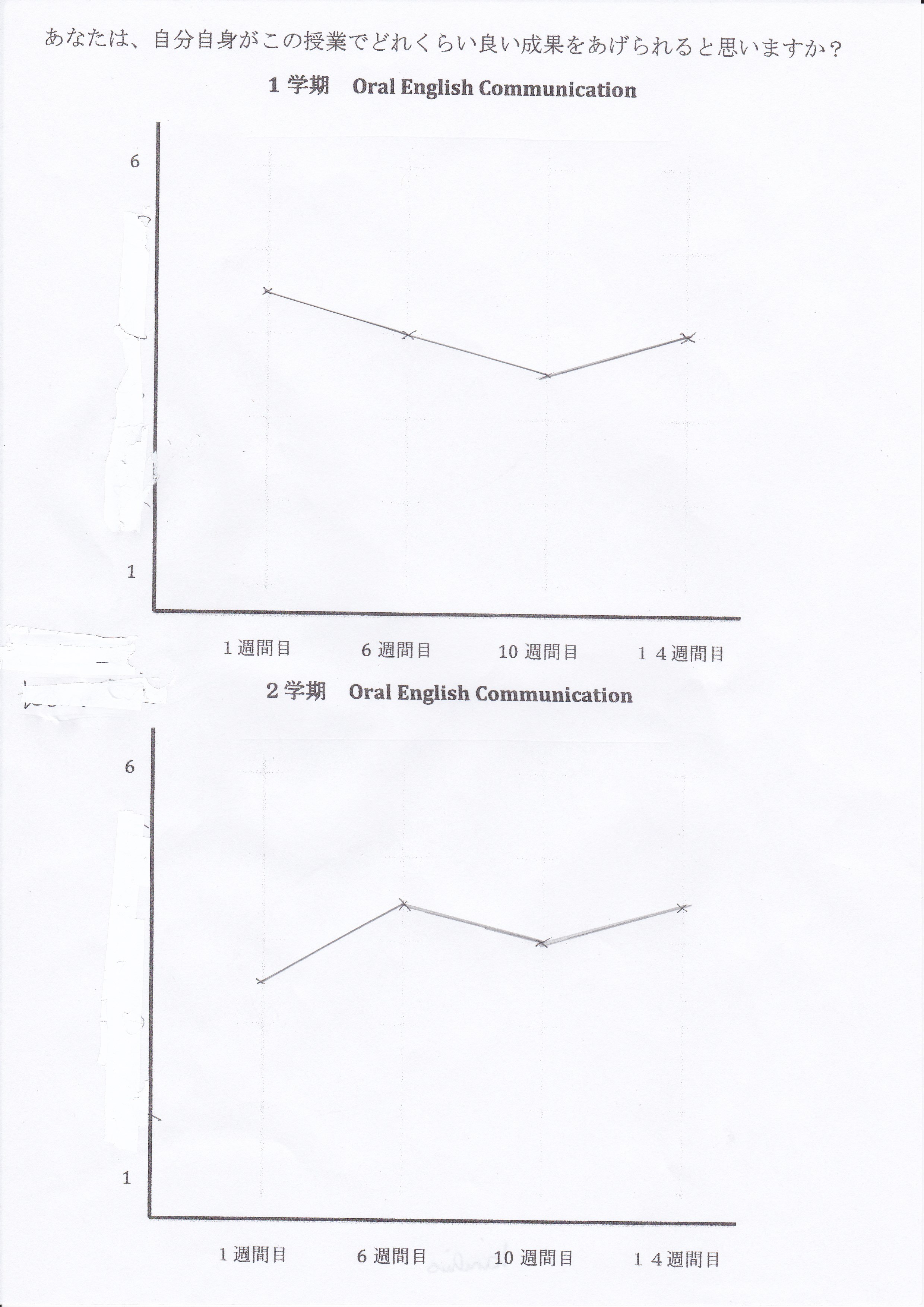
 

Figure 1. Data from eight questionnaires (left) and retrospective data from interview (right) for Kuni. Note: Top picture is for semester 1 and bottom picture for semeseter 2.

Interviews were watched repeatedly and transcribed. The level of detail in the transcription was a compromise between clarity and detail, with the priority placed on legibility for those not familiar with the conventions of conversation analysis (see Appendix C for transcription key).

**FINDINGS AND DISCUSSION**

*Retrospective data*

The numerical data for retrospection was extrapolated from the graphs completed by students. Table 2 shows the longitudinal and retrospective data from each of the 11 students in the current study. Before interpreting the results it is important to consider the fact that although students answered seven or eight questions in total for each administration of the questionnaire, for the retrospection they summarized their feelings at a given time to produce one value.

Examining the patterns of change showed that there were only two students who were more accurate in the first semester than the second semester in recalling the pattern of change. Makoto and Hiro (all names used have been changed) both matched the patterns of change more accurately for the first semester. Interestingly both of these students said that they had been able to vaguely recall their feelings. Five students (Chiemi, Kuni, Tomi, Kyo, and Ryo) claimed to have not been able to recall their feelings, and yet these were the students (with the exception of Ryo) who had been able to most accurately recall the patterns of change in the second semester. It seems that the students who claimed to remember were actually worse at remembering and seem to lack awareness. Perhaps the students who said they couldn’t remember were more aware of their lack of ability and therefore tried harder to recollect their feelings of self-efficacy, ultimately resulting in more success.

*Correlation analysis*

I performed a bivariate correlation with values obtained from the retrospective and longitudinal data. The results can be seen in Table 3 below.

Table 2. Longitudinal and retrospective data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SE1 | SE2 | SE3 | SE4 | SE5 | SE6 | SE7 | SE8 |
| Dai | **3.85** (3.47) | **4.57** (4) | **5.42**  (4) | **5.57** (4) | **2.12** (4) | **3.37** (4.47) | **3.75** (5) | **4.62** (5.53) |
| Hiro | **3.57** (3) | **3.42** (3) | **4.28** (4) | **4.42** (4) | **2.75** (4) | **3.37** (5) | **3.62** (5) | **4.12**  (5) |
| Kyo | **1.85** (1) | **2**  (1.21) | **1.42** (1.26) | **1.57** (1.68) | **2** (1.47) | **1.62** (1.74) | **2.12** (2.16) | **2.5** (2.26) |
| Yusuke | **3.42** (2) | **3.28**  (2) | **3.85**  (3) | **3.85**  (3) | **3.5** (3) | **3.25**  (4) | **3.62** (4) | **3.62** (4) |
| Makoto | **1.7** (2) | **1.42** (2) | **1.42**  (4) | **5.57** (4) | **3.5**  (3) | **2.37** (3.53) | **3.25** (4) | **3.12** (4) |
| Shun | **2.14** (4) | **2.42** (4) | **5.42**  (4) | **4.57**  (4) | **6**  (4) | **5** (5) | **5** (4) | **5.25** (4) |
| Shota | **3**  (5.53) | **4**  (5.53) | **4.57** (5.47) | **4.71** (5.47) | **6** (5.53) | **2.12** (5.53) | **4.87** (5.53) | **5.25** (5.53) |
| Ryo | **3.28** (3.37) | **2.71** (3.74) | **2.85** (4.26) | **3.71** (4.79) | **4** (4.26) | **2.75** (4.32) | **3.62** (4.79) | **3.5** (5) |
| Tomi | **2.28** (2) | **1.71** (2) | **1.71** (3) | **2.57** (3.63) | **1.87** (3) | **1.15** (2.53) | **2.12** (2.53) | **2.25** (3) |
| Kuni | **4.42** (4.53) | **4.57** (4) | **4.14** (3.53) | **4.28** (4) | **3.62** (3.53) | **4.25** (4.42) | **4** (4) | **4.37** (4.42) |
| Chiemi | **3.85** (3) | **3** (4) | **2** (3) | **3.57** (3) | **2.5** (3) | **3.25** (4) | **4.87** (5) | **4.37** (4) |

Note: Numbers in bold represent data gathered throughout the year from questionnaires. Numbers in brackets represent retrospective data gathered in final interview.

Table 3. Correlations for the longitudinal and retrospective data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SE1 | SE2 | SE3 | SE4 | SE5 | SE6 | SE7 | SE8 |
| RSE1 | **.48** | .68\* | .70\* | .53 | .70\* | .47 | .78\* | .85\*\* |
| RSE2 | .52 | **.66\*** | .64\* | .51 | .61\* | .44 | .84\*\* | .87\*\* |
| RSE3 | .22 | .39 | **.56** | .77\*\* | .67\* | .25 | .62\* | .66\* |
| RSE4 | .22 | .35 | .48 | **.65\*** | .63\* | .16 | .50 | .54 |
| RSE5 | .35 | .52 | .68\* | .62\* | **.67\*** | .28 | .65\* | .73\* |
| RSE6 | .50 | .63\* | .80\*\* | .73\* | .68\* | **.63\*** | .84\* | .89\*\* |
| RSE7 | .58 | .61\* | .55 | .71\* | .41 | .41 | **.78\*\*** | .77\*\* |
| RSE8 | .58 | .71\* | .69\* | .77\*\* | .40 | .38 | .62\* | **.71\*** |

RSE= Retrospective Self-efficacy; SE= Self-efficacy;\* p < .05; \*\* p < .01

The accuracy of recall was high in most cases, with strong, significant correlations in the second semester. Students’ accuracy of recall seems to worsen with time with the most significant correlation for the most recently completed questionnaires (SE7, SE8). Students had completed SE8 approximately two weeks before the interviews, and the correlation with retrospective data is strong and significant. Correlations remain significant throughout the second semester and it seems that students struggled more to accurately recall first semester scores. Recall of SE1 was insignificantly correlated with the actual data, as was the recall of SE3. From data collection to retrospective account there was a time interval of between seven to nine months for these two measures and this seemed to prove problematic. Although there is a drop in recall accuracy going back in time, the students seemed able to recall data from the most recent semester with reasonable accuracy.

*Ability to recall feelings of self-efficacy*

Upon re-entering the room where the interview was conducted I asked students how easy it had been to recall the data. At that point five of the 11 students stated that they had been unable to remember their feelings, with one of the students stating ‘*mattaku oboeteinai*’ (I don’t remember at all). Four of the students claimed to approximately remember the data, with expressions ‘*nanto naku*’ (roughly), ‘*daitai*’ (mostly), and ‘*kanaa*’ (maybe) all used as expressions suggesting that recall had been difficult. Despite the large number of students who admitted having problems completing the form, all of the students completed it, and when asked, gave a description of their self-efficacy during the first and second semesters that matched the quantitative retrospective data that they had provided.

While students were filling in the form I left the room but the video camera was recording. One student, Shun, looked directly at the camera and stated ‘*wasureta*’ (I have forgotten). He marked his retrospective data as 4 for the entire first semester, and for three of the four ratings in the second semester. This was different from his longitudinal data, and when I asked him if his self-efficacy had changed in the first semester, he said that it may have changed but he was unable to remember. His longitudinal data showed that his self-efficacy began at approximately 2, and rose rapidly through the first semester. The fact that Shun was unable to recall and yet filled in the form highlights a potential problem with quantitative data. This suggests that even if a student is unable to recall, and willing to admit this in interview, they may still complete the quantitative measure, adding error to any measurement, and suggesting that explicitly providing students with an ‘*I can’t remember*’ option may be advisable.

I asked the students to explain their changes in self-efficacy based on the retrospective accounts that they had completed. The majority of students explained in detail how their self-efficacy had changed over the academic year, even if they had initially admitted limited ability to recall. However when shown the actual quantitative data which in many cases conflicted with their retrospection, students admitted they had been unable to recall how they had felt, and struggled to explain the changes. It is possible that they had initially believed that they could recall their feelings and the longitudinal data showed them they could not, or that they had used inference to compensate for gaps in memory but had been reluctant to admit this earlier in the interview.

Hiro initially claimed he had been able to recall his feelings of self-efficacy, using the expression *nanto naku (roughly)*, suggesting that he was not entirely confident. He was able to give a reasonably detailed account of the changes in self-efficacy over the course of the first and second semesters with few hesitations. Following on from his account I showed him the longitudinal data and he became far less confident as shown in the excerpt below.

Excerpt 1 Hiro interview

38 H： =ichi gakki ha niteimasu

(Semester one looks the same)

39 P： ==un ichi gakki ha ni gakki ha..dandan a ni ichi gakki kara ni gakki ha sagatta kedo/

(Yeah semester one is in semester 2 gradually ah it went down from semester 1 to semester 2 though)

40 H： hai..nani shiteta..ah summer vacation [laughs]

(Yeah, what did I do)

41 P： un unde hikuku natte.. un ni gakki ha nanka/

(Yeah, it got lower and then in semester 2 kind of)

42 H： agattemasune//

(It goes up doesn’t it)

43 P： nande kannaa//

(Wonder why)

44 H： [(41) hand to mouth, breaths heavily, moves hand to head and strokes head, laughs]　chotto wakarania//

(don’t really know)

Hiro’s retrospective quantitative data is accurate for the first semester, but the second semester is different, and this causes him problems when asked to explain the changes. In Line 40 Hiro asks himself what he did that made his self-efficacy decrease, and finishes this comment with a nervous laugh. Hiro is struggling to recall and there is a silence of 41 seconds, and he seems uncomfortable during this time, avoiding eye-contact, breathing deeply and scratching his head. He breaks the uncomfortable silence with a nervous laugh. Following this Hiro attempts to explain what he did during the summer in order to explain the possible discrepancy, before laughing and finishing with a question indicating a final lack of ability to recall. Laughter is used to ease the awkward situation where he is unable to provide an explanation. It seems that Hiro is unable to remember the reasons for change, and yet he is very reticent in admitting this. It is possible that as I was his teacher he felt pressure to provide an explanation, although he was unable to do so.

When initially talking about his retrospective data Hiro gave a confident, fluent explanation and seemed to recall the feelings of self-efficacy through the year until presented with the actual data, which provided a different view. He then became hesitant, asking questions and seeming to be unable to explain what had happened. Interestingly in his case the retrospection for the first semester was more accurate, despite the relative temporal proximity of the second semester.

The behavior displayed by Hiro was also observed with other students, who provided detailed descriptions of self-efficacy over the academic year, and then when presented with longitudinal data showing a different pattern became hesitant, and unable to explain what had happened. Japanese uses a number of expressions to express uncertainty including *kana*, *kamo,* and *darou (yarou* is Kansai dialect)*,* and there was an increase in these expressions when students attempted to explain longitudinal data which contradicted their retrospective account. This uncertainty implies that students are using inference as well as recall to aid their own reconstruction of events. Although this does not mean that the explanations are not valid, it should be considered when interpreting comments.

Despite the fact that second semester quantitative retrospection was more accurate as shown by the correlation data, there was no noticeable difference in fluency as students gave their initial accounts explaining the data. Students moved smoothly from semester 1 to semester 2, explaining changes with little hesitation. Although the start of semester 1 was almost 10 months prior to the interview, students were fluent in their accounts and there was no discernable difference in language or behavior of respondents between first and second semester, despite the relative proximity of the second semester. Some students had more accurately reproduced their first semester data, while others were more accurate with the second semester, but there was no noticeable pattern in terms of fluency of accounts when describing the data. This suggests that at least from the perspective of the students, they felt able to recall the data for the entire year.

*Contrasting accounts*

Although some students were able to accurately recall, in general there were a number of discrepancies between the retrospective data and the longitudinal data. Students produced reasonably detailed verbal accounts as to why their self-efficacy had changed, and these accounts matched the retrospective quantitative data that they had produced. Following on from their explanation, the longitudinal data was presented, allowing students to see what they had originally produced. I asked why the changes had occurred in their longitudinal data, focusing on differences between retrospective data and longitudinal data.

When presented with the contrasting data some students admitted that they were unable to explain the changes in the data, but a number of students attempted to provide explanations of change in the longitudinal data, as illustrated in the following excerpt.

Excerpt 2. Kuni interview

Kuni explaining retrospective data:

9 P： =ah ok jaa jishinn atte dou shitann {[laugh] desuka?}

(ah ok so you were confident then what happened)

10 K： [laugh] eto omotta yori ka sono guruupu de katsudou suru koto ga ookatta node sono guruupu no minna o toiuka encourage to iuka dekinakatta no o

(erm there was a lot more group work than I had thought and I couldn't encourage or or erm them couldn’t encourage them so)

11 P： =aaa un un un

(ah ok yeah)

12 K： =omoinagara de guruupu de kangaeru to jibun de umaku seika ageru ( )=

(thinking of that and the group I can do well in class ( )

13 P： =un un un

(okay yeah)

14 K：=to iu node sagatte kitete/

(so it started to go down)

Kuni presented with longitudinal data:

40 K： kore ha chotto ichi gakki ha oboetenakattannde

(this one first semester I couldn't really remember so)

41 P： ==un chotto agatta saisho roku shuukan me demo hotondo hobo kawaran (4) nande agatta kana//

(yeah it went up a little at week six but mostly it hardly changed wonder why it went up)

42 K：{[sits back and strokes chin] nani ga attan kana/..ichi gakki (5) ichi gakki a conversation ha saisho no gofun juppun toka} no conversation ga gofun juppun no conversation ga tabbun saisho no ichiban saisho no kai gurai de sogoi kaiwa ga tsuzukanakute=

(what happened I wonder first semester first semester ah the initial five or ten minutes conversation was maybe the fist time around the very first time the conversation didn't last at all)

43 P： =aaa un

(Oh okay)

44 K： =de jojun ni follow up question toka ( )=

(then there were more and more follow up questions)

45 P： =aaa un

(Oh okay)

46 K： ­=ano mukou no kiki kata ga wakatan de jibun mo kiki kata ga wakatan de

(I understood the way they asked questions and also figured out how to ask questions so)

47 P： =un

(okay)

48 K： =sore de tabun jakkan agattanjanai kana/=

(so maybe because of that it went up a little maybe)

Kuni initially displays confidence in his account, explaining how he began with high self-efficacy in his English due to prior experience interacting in English and also with foreigners, but that once he realized the large amount of group work he lowered his expectations of his grade. He explained that in the final week his self-efficacy managed to make somewhat of a recovery as the group finally made an effort around the end of the semester. This account was delivered with confidence and a lack of hesitation by Kuni.

When presented with the retrospective data in line 40 Kuni states that he was not able to accurately recall the first semester. His self-efficacy was quite stable in the first semester, in contrast to his retrospective account where he claimed to have lost self-efficacy as a result of considering his group. (See Figure 1 for Kuni data). Line 42 shows Kuni attempting to recall what happened and he strokes his chin, implying thought, leaves a short pause of five seconds, and then uses ‘oh’ as if to indicate that he has remembered what happened. After this he explains that his self-efficacy rose as the group began to understand how to use follow-up questions and to engage in basic conversation that occurred at the start of each lesson. This account is in contrast to his previous explanation, but is more hesitant with more use of expressions of doubt as exemplified by line 48 where he states ‘so maybe because of that it went up a little maybe.’ This extensive use of particles to express uncertainty suggests that he is not confident in the account and inference has been used.

This account is similar to that provided by Chiemi, who again gave contrasting accounts of her changes in self-efficacy during the first semester (see Appendix E for retrospective and longitudinal data). When explaining her own data she stated that although she began with a lack of confidence due to her dislike of English, she compared herself to others and grew in confidence through the first semester. When presented with the longitudinal data she acknowledged that she was not able to recall the first semester. After some time, she provided a new account of the data which was in stark contrast to her previous account. When introducing her new account she stated ‘*aa jaa ima kochi o mite omou to ganbarou to omotte ganbarou to omottan desukedo*’ (Okay, so looking at this now, I thought I must try harder, I must try harder). This phrase very much resembles prompted recall, where students are shown video and asked to recollect their thoughts or feelings at a given time, and it is possible that showing students their actual quantitative data did in some way allow them to recall what had happened. She did qualify her new explanation at the end by reiterating that she was unable to recall the first semester.

In the interviews both Chiemi and Kuni acknowledged that the longitudinal data is different from the retrospective data, and then constructed a new and contrasting narrative to explain the longitudinal data. There are two possibilities to explain how students come to their second account. It is possible that, in the same way as prompted recall, seeing the actual data helps the students to recall their feelings and therefore they are able to give a more accurate explanation of what occurred. A second possibility is that students are unable to recall but feel pressure to compose an accurate explanation of their data and therefore construct a narrative to match the data, based on their knowledge of what occurred during the year including major events such as tests. As Tourangeau et al. (2000) state, it is likely that recall is a mixture of memory and inference, and therefore there is likely to be a high degree of inference in both accounts delivered by students.

*Self-efficacy: change and stability*

In all cases the students’ retrospective quantitative data differed to some degree from the longitudinal data and this prompted two different kinds of response. A number of students seemed to accept that the longitudinal data was an accurate representation of their self-efficacy, as evidenced by attempts to explain the data. Even though most students attempted to explain the data at some level, four of the students clearly stated that they believed that their retrospective account was more accurate than their longitudinal data. Two of the students explained that the longitudinal data was subject to weekly fluctuations due to factors such as the previous lesson, and believed that the retrospective account was therefore more accurate. Ryo addressed this, explaining that the longitudinal data may be subject to temporary factors such as emotions, and experiences in the previous class.

Excerpt 3 Ryo Interview

40 P： ahh ja kimochi ga kono eigo ni tai suru jishin ha kotei shitenai..kekkou kouu na kawaru to iu koto/

(ahh so the feeling this confidence in English is not stable you mean it changes a lot)

41 R： sou desukane sono jugyou mae no jugyou toka de anma shaberenkattari shitara kawaru kamo shirenai (desu ne)/

(I guess so if I couldn't talk in the previous class then it might change maybe)

In line 41 Ryo states that if he was unable to speak in a prior lesson then his subsequent self-efficacy may have been affected, suggesting that self-efficacy is susceptible to change. When asked which account is the more reliable, he states that his retrospective account is more accurate. He explains that his feeling was consistently around four for the entire semester, as represented by his retrospective account. It seems that Ryo believes that although self-efficacy changes on a weekly basis depending on classroom experiences, the underlying construct itself is far less susceptible to change and remains constant through a semester. He therefore believes his retrospective account to be the more accurate as it is not swayed by weekly experiences but a more general underlying feeling. Another student, Yuichi, also felt that his actual data was influenced by his feelings at the time and claimed that his retrospective account was a more accurate representation of his general feeling of self-efficacy. Both of these students’ accounts suggest that they believe that self-efficacy operates at two different levels. Underlying self-efficacy is relatively stable and can be recalled clearly, but at a more superficial level self-efficacy is temporal, affected by events at a given time and can be quite susceptible to change.

Another student who claimed that his retrospective data was more accurate was Shota. He felt that his self-efficacy was stable over the course of the academic year, as shown by his retrospective data in Appendix E. His retrospective data showed no change with self-efficacy at 5.5 throughout the entire year, although his actual data showed a gradual increase in self-efficacy in both the first and second semesters. When I re-entered the room I asked Shota how difficult it had been to recall his feelings, and he claimed that it was difficult but he remembered marking his self-efficacy as hovering between five and six for the whole year. When presented with the longitudinal data, Shota changed his account, claiming that he had deliberately manipulated the longitudinal data by giving false low ratings of his own self-efficacy in order to avoid appearing too confident in his English. He felt that if he had rated his self-efficacy high, I may have selected him for extra responsibility. Shota was adamant that his retrospective account was the more accurate description of his feelings, and seemed to believe that his self-efficacy was stable and had not changed over the year.

The research highlights the danger of over-reliance on a single method of data collection, be it qualitative or quantitative, interview or questionnaire. Shota claimed to have deliberately manipulated his own responses to the self-efficacy questionnaire, in order to avoid drawing attention from the teacher. He had little to gain by admitting fabricating his data, other than arguing that his retrospective account was more accurate. This suggests that quantitative data should be subject to the same examination as qualitative data, where issues such as the context, and the relationship between researcher and participant need to be clearly discussed.

**CONCLUSION**

The study showed that students have a reasonable ability to recall data over short periods of time, but that the quality and accuracy of retrospection generally weaken with the passage of time, as suggested by Tourangeau et al (2000). Despite all of the data being collected less than a year prior to the interview, and although students completed the forms and explained their changes in self-efficacy, when presented with the actual data many struggled to recall their feeling. Students seemed reluctant to admit that they were unable to remember events, and when presented with longitudinal data that differed from their retrospective data, some attempted to explain reasons for change, even if accounts conflicted with previously elicited explanations.

There are several implications for researchers which can be taken from this study. While interviews provide a useful insight into the thoughts and feelings of students, it must be appreciated that we are accessing their current views, and even when students are asked about prior events the researcher cannot be entirely confident that a great deal of inference is not being used, rather than actual recall of the events. Presenting students with the quantitative data may have led to prompted recall, and aided students in accessing memories of self-efficacy, or simply led students to use inference to create an explanation matching the data. Both accounts of change highlight the multitude of feelings that students have over the course of an academic year, and it is important that researchers consider this in their interpretation.

From a pedagogical perspective, it is clear that students have a somewhat limited ability to recall their own feelings of self-efficacy from one year ago, and therefore it is likely that students are unaware of their own gains made in self-efficacy. Self-efficacy has been shown to have a direct link with language learning gains, and in both retrospective and longitudinal accounts of the data students were able to make gains in their self-efficacy. During the class I provided students with many simple tasks and opportunities for interaction, and it seemed that in most cases self-efficacy did increase. As teachers we must strive to aid students in increasing self-efficacy, which in turn will have a positive impact on performance. The results of this study suggest that self-efficacy is quite a malleable construct, and teachers must therefore work hard to protect and foster increased feelings of efficacy in their students.

Retrospective data collection may have some advantages in terms of cost, but it is important that researchers are aware of the limitations. The accounts provided by students above show various influences on self-efficacy during the course of the year, but also show that students struggle to recall their own feelings, even within a relatively short timeframe. I believe that in SLA more qualitative accounts of change over time will help us to understand the factors influencing students, and how variables such as motivation may change, but also that studies should be very explicit in recognizing that we are accessing students’ memories of events rather than the events themselves. More than anything else I think this study suggests that it is essential to have data from multiple sources to provide a clear description of change.

**REFERENCES**

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Bernard, H., Killworth, P., & Sailer, L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology*, *13*, pp. 495–517.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd Edition). New Jersey: Routledge.

Conway, M., & Ross, M. (1986). Remembering one’s own past: The construction of personal histories. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition. Foundations of social behavior* (pp. 123–144). New York: The Guildford Press.

Featherman, D. (1980). Retrospective longitudinal research: Methodological considerations. *Journal of Economics and Business*, *32*(2), pp. 152–169.

Hayashi, H. (2005). Identifying different motivational transitions of Japanese ESL learners using cluster analysis: self-determination perspectives. *JACET Bulletin*, *41*, pp. 1–17.

Hsieh, P., & Kang, H. S. (2010). Attribution and Self-Efficacy and their interrelationship in the Korean EFL context. *Language Learning*, *60*(3), pp. 606–627.

Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning, 57*(3), pp. 417–442.

Nikolov, M. (2001). A study of unsuccessful language learners. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and Second Language Acquisition* (pp. 149-170). University of Hawaii, Second Language Teaching and Curriculum Center.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), pp. 33–40.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*(2), pp. 341–357.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11–26). New York: Oxford University Press.

Shoaib, A., & Dörnyei, Z. (2005). Affect in lifelong learning: Exploring L2 motivation as a dynamic process. In P. Benson & D. Nunan (Eds.), *Learners’ stories: Difference and diversity in language learning* (pp. 22–41). Cambridge: Cambridge University Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press.

Ushioda E. (2001). Language learning at university: exploring the role of motivational thinking. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and Second Language Acquisition* (pp. 93- 125). University of Hawaii, Second Language Teaching and Curriculum Center.

**APPENDIX A**

**ENGLISH COMMUNICATION CLASS SELF-EFFICACY (ENGLISH VERSION)**

How well do you believe you will perform in this class? Rank yourself on the following statements. 1 (not at all true), 2 (slightly true), 3 (somewhat true), 4 (quite true), 5 (true), 6 (very true):

1. I can understand what is taught in English Communication Class.
2. I can do very well in English Communication Class.
3. I can learn the material for English Communication Class.
4. I can receive a good grade in English Communication Class.
5. I can speak English fluently when giving a presentation in front of the class.
6. I can speak English fluently when taking part in a group discussion.
7. I can give my opinion in English.
8. I can enjoy conversation in English.

**APPENDIX B**

**INTERVIEW PROCEDURE**

1. Introduce purpose-consent form. Confidential. No right or wrong answer.
2. Have students fill out retrospective chart of self-efficacy (Appendix C).
3. Changes in self-efficacy. Show students retrospective data (completed at the start of the interview) and ask for explanation.
4. Show actual data. Compare accounts and ask for explanation.
5. Questions/comments.

**APPENDIX C**

**CHART USED IN INTERVIEW FOR RETROSPECTIVE DATA COLLECTION (SEMESTER 1 ONLY)-ENGLISH TRANSLATION OF JAPANESE ORIGINAL**

How well do you think you can do in this oral English class?

**Semester 1 Oral English Communication**

6

1

　 　Week 1 　 Week 6 Week 10 Week 14

**APPENDIX D**

**TRANSCRIPTION KEY**

|  |
| --- |
| Transcription key (from Roberts 2007)  // Final fall  / Slight fall (indicating more could be said)  ? Final rise  , Slight rise (more is expected)  - Truncation  .. Pause of less than .5 of a second  … Pause of more than .5 of a second  = Overlap  = = Latching on to previous utterance  { [ ] } Non-lexical phenomena which overlays the lexical stretch  [ ] Non-lexical phenomena which interrupts the lexical stretch  ( ) Unclear word  (chair) Guess at unclear word |

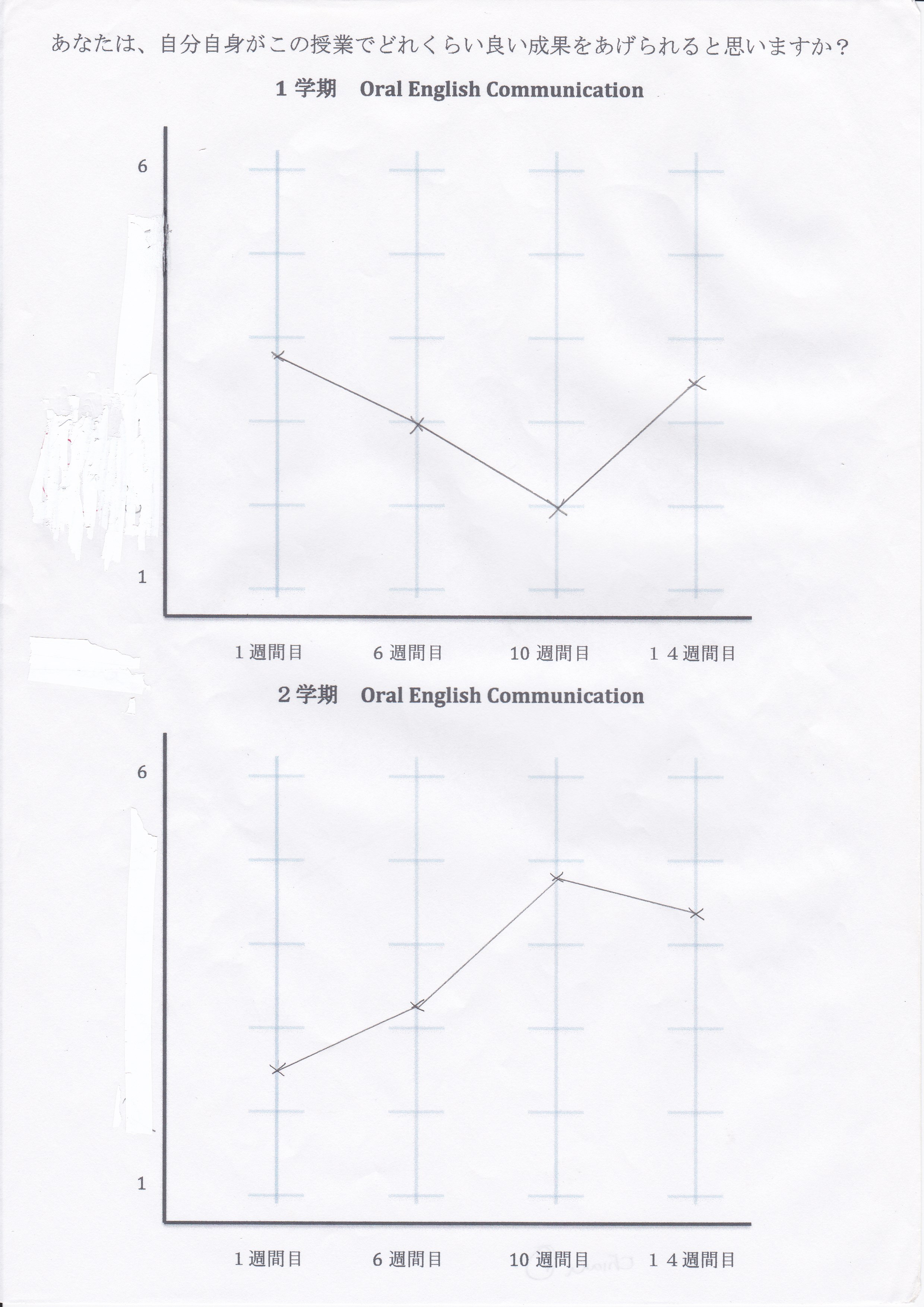
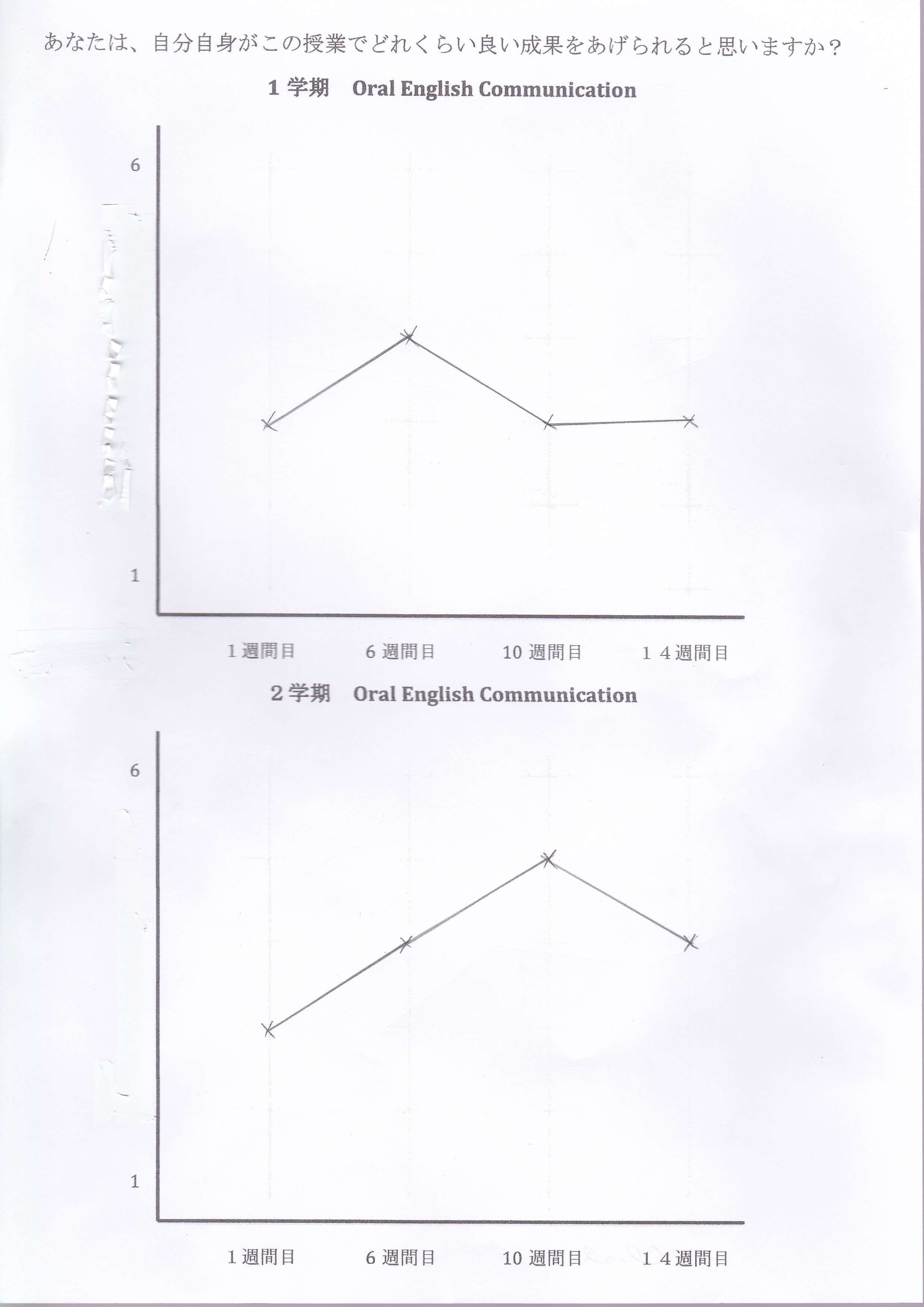
**Romanization of Japanese script followed the Hepburn system**

**[http://en.wikipedia.org/wiki/Hepburn\_system]**

**APPENDIX E**

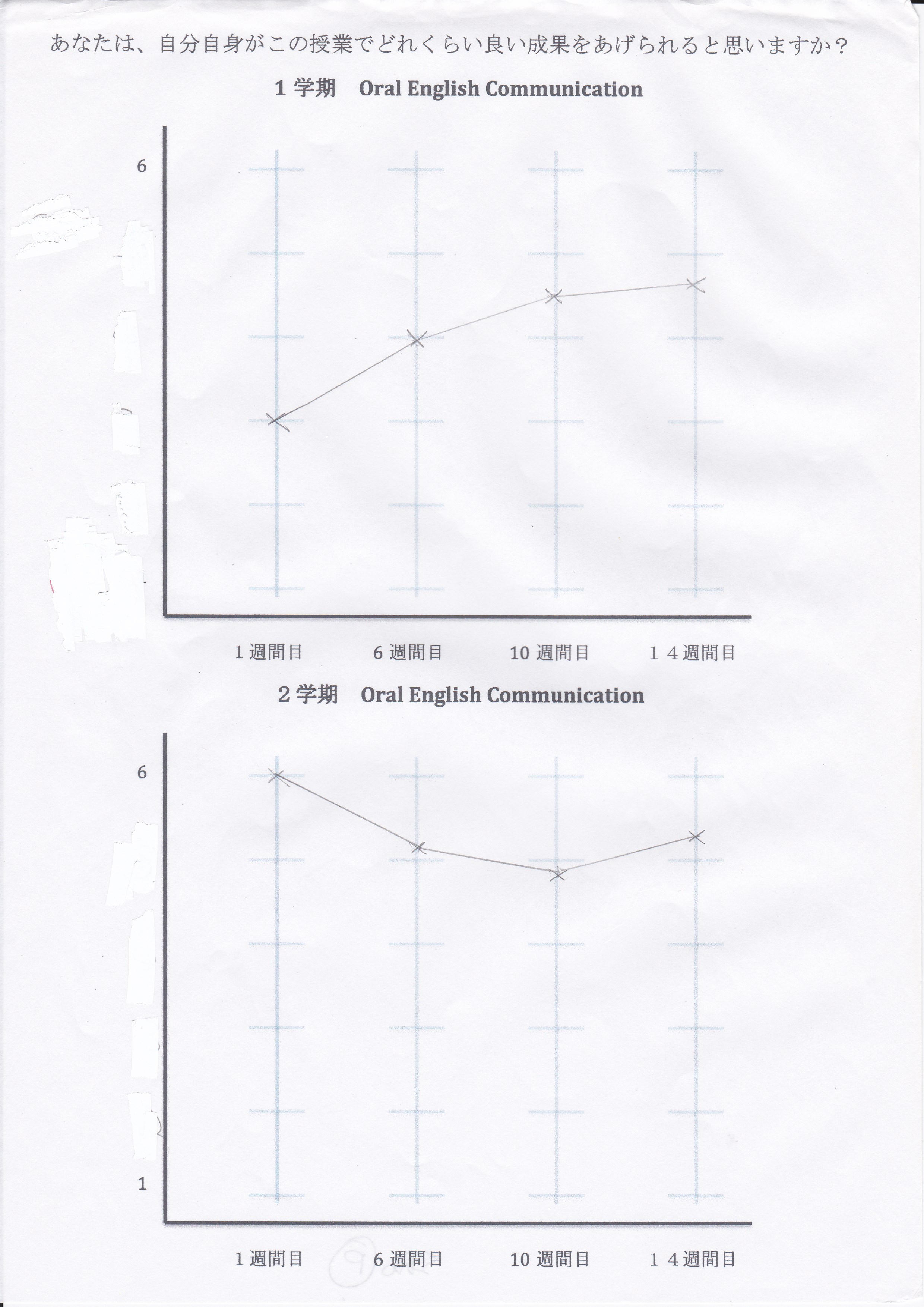
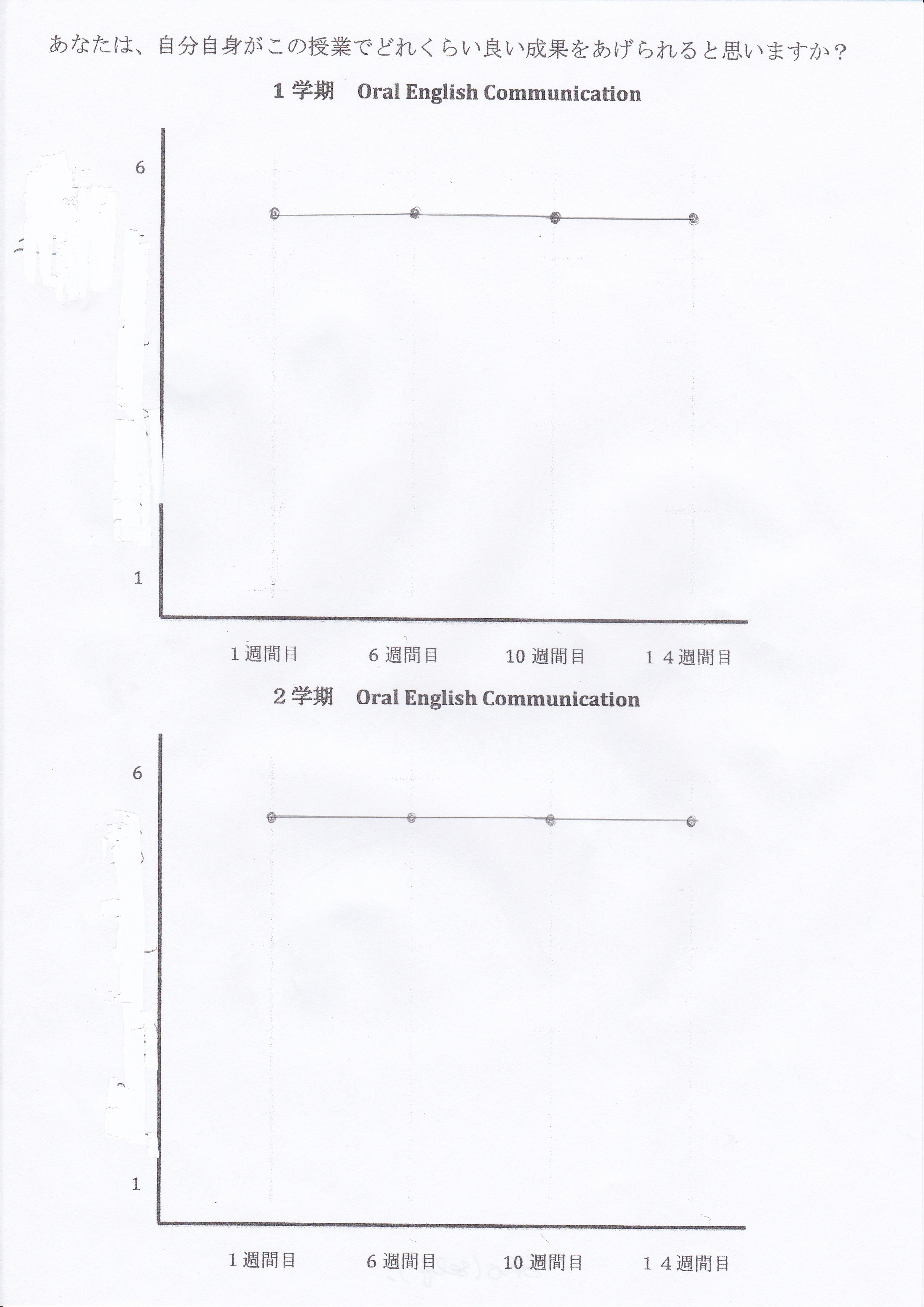
**RETROSPECTIVE DATA**

Chiemi

Actual data (from eight questionnaires) Retrospective data (from final interview)

Sho

Actual data (from eight questionnaires) Retrospective data (from final interview)