Measuring Learners' Interest in Materials via Methodological Triangulation[1)]

Haedong Kim
(Hankuk University of Foreign Studies)

Kim, Haedong. (2007). Measuring learners' interest in materials via methodological triangulation. *Modern English Education, 8*(3), 61-84.

This study aims to illustrate how we measure learners' affective domain, especially their interest in classroom-used materials, throughout methodological triangulation. The used methods of data-collection include learner's daily questionnaire, learner's end-of-semester questionnaire, learners' diaries, semi-structured weekly learner-interviews, and observation of learners' classroom behaviors. The overall results of correlations, checked via Pearson r correlation coefficients, indicate that the learners who took part in the present investigation responded their self-reporting interest in a modestly consistent way. To conclude, it is suggested that, as learners' affective domain belongs to realm inside learners, it should be necessary to employ mixed methods of data collection to enhance the credibility of the findings.

Keywords: 〔methodological triangulation/affective domain/다각적 방법/정의적 영역〕

1. Introduction

As McDonough and McDonough (1997) indicate, employing mixed methods of data collection—so called methodological triangulation

1) This paper is supported by Hankuk University of Foreign Studies in 2007. The earlier version of this paper was reported at the 2007 conference of Applied Linguistics Association of Korea.

(Allwright & Bailey, 1991)—enables a researcher to check whether different data coincide on a number of points, so allows "the opportunity of greater credibility and greater plausibility of interpretation" (p. 71) of the findings. Especially learners' affective domain—to be specific their interest in classroom-used materials—belongs to realm inside learners. Therefore, it should be important for a researcher to employ mixed methods of data collection.

The main aim of the present study is to check whether the learners who took part in the various data-collection methods at the present investigation responded their self-reporting interest and behaviors in a modestly consistent way. The used methods of data-collection—to get self-reporting interest in the classroom used materials—in the present study include i) learner's daily questionnaire, ii) learner's end-of-semester questionnaire, iii) learners' diaries, and iv) semi-structured weekly learner-interviews. Additionally, observation of learners' classroom behaviors during the use of materials is also used for data-collection, since 'self-reporting' responses and classroom 'behaviors' are considered to be different variables (Dörnyei, 1998; Peacock, 1996).

The data utilized in this study is mainly based on the data for answering one of the research questions in the author's unpublished doctoral thesis: "Do the learners consider that 'pedagogic' illustrations in EFL materials are more interesting than 'decorative' ones?" (Kim, 2001). It focuses on the different levels of learners' responses on different types of classroom materials. It can be noted that the focus of the present study is different. The results of the study may identify the levels of agreement among the data from various collection methods and justify the need of using methodological triangulation.

2. Literature Review

Learners' reponses on materials are related to their actual experience of and perceived satisfaction at materials (Deaux, 1993; Hutchinson, 1987; Hutchinson & Klepač, 1982), and so often regarded as valuable information for ELT teachers' coursebooks selection (Breen & Candlin, 1987, Jenks, 1981; O'Sullivan, 1991; Yorio, 1986). Many educational psychologists (Alexander, Jetton, & Kulikowich, 1995; Krapp, Hidi, & Renninger, 1992; Mitchell, 1993; Tobias, 1995) argue that, during the use of materials, a learner's actual 'state' of classroom interest can be affected by external stimuli, such as more interesting type of materials, and, at the same time, it can be dependent upon his or her own personal 'trait'. In a similar vein, Gardner and Tremblay (1994) propose to distinguish state motivation, i.e. actual classroom-situation motivation, from trait motivation, i.e. motivation inside learners that they bring into classroom, in a foreign language learning. Dörnyei (1994) supports that their proposal is "a very useful and vital distinction and ... results using a state/trait paradigm in L2 motivation research are most interesting and promising" (p. 520). Skehan (1989) also considers possible external effect of materials on EFL learners' actual classroom interest. In this sense, learners' interest in materials are multi-dimensional and should be evaluated in various ways.

With regard to measuring learners' responses to ELT materials, collecting learners' 'self-reporting interest' after the use of materials at each class, rather than gathering their expectations, may guarantee a high quality data, i.e. an empirically valid data, so it could be a "legitimate way of conducting an empirical evaluation of teaching materials"(Ellis, 1997, p. 37). McDonough and Shaw (2003) also advocate that the success of materials can be judged after the classroom use. Therefore, it should be better to adopt various data-collection methods, in order to check the reliability of the learners' self-reporting responses to a particular instrument, for example a questionnaire, by comparison with other

self-reported data, such as interviews, learners' diaries and so on. Additionally, observation of learners' classroom behaviors during the use of materials also can be employed for data-collection—in literature, 'self-reporting' responses and classroom 'behaviors' are regarded as different variable (Dörnyei, 1998; Peacock, 1996).

Due to the nature of qualitative data collection, it should be difficult for any investigators to invite a large number of learner-subjects for interviews and to collect diaries and essay writings from them. This may be a main reason for lacking a study on the use of methodological triangulation in the ELT field. To this investigator's speculation, there are not many studies on triangulation. Within Korean ELT context, Lee and Oxford (2007) have currently conducted a study with 20 learner-subjects. They investigate reading strategies of L2 learners, by the use of a combination of a questionnaire and think-aloud protocols. They conclude that both instruments are effective to measure students' reading strategies complementarily. Apart from this study, it was difficult to find a relevant study on methodological triangulation. In consideration of its importance as proposed in literature, it should be necessary for us to report a case of measuring learners' affective domain in various ways.

3. Research Method

3.1 Participants and Setting

The learner subjects were limited into the 70 students of 3 sections of a general English course at a university in South Korea. This investigator had taught all 3 sections by using the same course materials in the same order for the whole 16 weeks course period. Class hours of each section were 3 hours per week. The section A had 23, B had 28, and C had 20 registered

students. One student dropped the course in the middle of the course. In all three sections, the majority belonged to age group of 18 or 19. English proficiency, estimated by scores of a nation-wide university entrance examination, was similar to each section.

The commonly used coursebook in the course was *On Target 1* (1991) written by J. Purpura and D. Pinkly and published by ScottForesman in USA. In the blurb, it says that the book aims to develop the communication skills—listening, speaking, reading and writing—within a carefully planned syllabus of functions, notions, and grammar. A total of 50 sets of materials, from the main coursebook and non-coursebook materials, had been used for the data-collection. The average minute used for a set of materials was 35 minutes.

3.2 Instrument

In the present study, the operational definition of 'self-reported interest', i.e. experienced interest, is learners' response of interest directly relating to the used materials at classroom. The referent of self-reported interest is;

A - Learner responds three items representing 'interest' on a day-by-day (daily) questionnaire. The sum of scores on three items indicates a degree of interest in a set of materials.
B - Learner rates interest of each set of materials on an end-of-course questionnaire at the end of the course.
C - Learner writes about the interest of used materials in learner's diary.
D - Learner comments on the interest of used materials during a weekly interview.

The learners' classroom behaviors during the use of materials

were recorded by a video-camera and then scored on observation sheets by this investigator and one independent observer. In the present study, components of observed behaviors are 'emotional expression of interest', and 'distraction.' They are operationally defined.

A - Emotional expression of interest: An individual learner's interest expressed by bodily signals during the use of materials.
B - Distraction, as opposed to 'emotional expression of interest, is an individual learner's distractibility expressed by bodily signals during the use of materials.

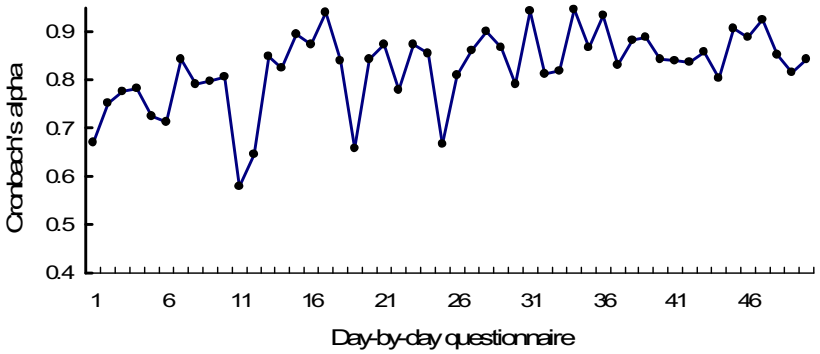
3.2.1 A day-by-day Questionnaire

It was adapted from one in Peacock's study (1996, p. 256: originally adapted from Gliksman, Gardner, & Smythe 1982, p. 646). It was one page of A5 size. It contained 6 items on a 7-point 'semantic differential scale', in which the learner reacted his or her experience of the set of materials by placing a tick along a series of bipolar response continuums (Savignon, 1983, p. 309). A total of 3 items were asking about their experienced 'interest': interesting - boring, unenjoyable - enjoyable, and exciting - dull. We employed a technique of mixed use of positively and negatively worded items in the questionnaire, in order to avoid a possible 'response bias' since some respondents might "tend to answer positively more often than negatively" (Skehan, 1989, p. 11).

The internal consistency reliability, i.e. the item reliability, of the 3 items representing 'interest', in each of the 50 sets of questionnaires was estimated. Cronbach's alpha was chosen as a measure. The following Figure 1 displays visual results.

The total of 50 reliability coefficients of 'interest' had a mean of .82 (SD = .08), and never dropped below .55 during the data-collection

FIGURE 1
Item Reliabilities of the 50 Sets of Daily Questionnaires



period. These results indicated that the learner-subjects had filled out the 3 items measuring self-reporting 'interest' with acceptable degree of care during the whole course period. Therefore, we defined 'self-reporting interest' as the summed score of the 3 items. Each item score resulted in a score of 1 to 7, so the summed score ranged from 3 and 21. For the calculation, the score on the negatively worded item of 'unenjoyable - enjoyable' was recoded.

3.2.2 An End-of-course Questionnaire

The learners were asked to rate 'interest' of each set of classroom used materials on an end-of-course questionnaire. The absentees on the class were excluded in the survey. A total of 56 learners took part in the end-of-course questionnaire.

The end-of-course questionnaire included the titles of 48 sets of materials—the last 2 sets, among the total of 50 sets of materials, were excluded in asking, since the learners had just used them at the class before the last class. Under the each title, the learners were asked to rate experienced interest. Ratings were on a 7-point

Likert scale on which 7 = *very interesting* and 1 = *not at all interesting*. It took about 25 minutes to administer the questionnaire. Since answering the questionnaire gave the learners an opportunity to reflect what they had studied through the materials, it also served a pedagogical aspect of the last class of the course.

3.2.3 Learners' Diaries

Keeping a diary was a course requirement for the learner-subjects. At the first class, the learners were asked to write about their experienced interest (or disinterest) in a set of materials used in the classroom. It was recommended to write 100 to 300 words for one set of materials. This teacher-investigator collected diaries every two weeks. The collected diaries returned to the learners at the next class. To avoid a possible sensitization in 'interest', the learners were also encouraged to write other inner experience, such as experienced usefulness, and/or other material-relevant classroom experience, such as reactions of teacher, pair-work partner, or other learners in using materials.

A total of 54 learners (77 %) gave their diaries for this investigator at the end of the course. However, not all the learners had kept diaries until the end of the course. 7 learners' diaries were omitted due to the reason, so 47 diaries (67 %) were utilized for the present analysis.

For the quantification, the strength of the expression, i.e. "saliency" (Allwright & Baily, 1991, p. 193) was focused. The following scale was employed for scoring: 4 = comment of 'interesting' with positive adverb (e.g. very, extremely and so), 3 = comment of 'interesting', 2 = negative comment of 'interesting', 1 = negative comment of 'interesting' with adverb. Synonyms and antonyms of 'interesting' were referred to for the identification. Synonyms were 'enjoyable', 'exciting', 'stimulating', 'involving',

'engaging', 'appealing', 'absorbing', and 'funny', and antonyms were 'boring' and 'indifferent' (adapted from Murphy, 1993; Peacock, 1996; Schiefele, 1992). Throughout reading diaries during the course, it was found out that this group of learners used other words as synonyms and antonyms of 'interesting'. Therefore, those words were referred to as well. Synonyms of 'interesting' were 'pleasing', 'pleasurable', 'humorous', 'laughable', 'amusing', 'novel', 'impressive' and 'merry', and antonyms were 'dull', 'monotonous' and 'tedious'.

After the data-quantification, we checked reliability of quantification. Reliability was calculated by correlating two sets of ratings for a single learner's diary via Pearson r. The mean of 47 intra-rater Pearson r correlation coefficient reliabilities between this investigator's first ratings and the second ratings on each of 47 learners' diaries was .94 (SD = .12). The mean of 47 inter-rater Pearson r correlation coefficients between this investigator's first ratings and the ratings of an independent analyst was .91 (SD = .12). The mean of 47 inter-rater Pearson r correlation coefficients between this investigator's second ratings and the ratings of an independent analyst was .91 (SD = .10). The results indicated a high level of agreement among three occasions of quantification across all of the 47 learners' diaries.

For data analyses, three scores for each set of materials, obtained from three occasions of quantification, were summed, yielding a possible score ranging from 3 to 12. It, then, was found out that no scores were 7 or 8 across all the summed scores, mainly due to high levels of intra-rater and inter-rater reliabilities. Therefore, the summed scores ranging from 3 to 6 were categorized as 'comment of not-interesting' and the summed scores ranging from 9 to 12 as 'comment of interesting'. These two categories of comments were utilized for further present analyses. A total of 556 comments, from 47 learners' diaries, were identified, in relating to their experienced interest in classroom used materials. The average

number of individual's interest-comments was 11.8 (SD = 5.49, with maximum comments 24 and minimum comments 4).

3.2.3 Weekly Learner Interviews

Since the learners took part in interviews on a voluntary basis, the number of interviewees were different in each week. As a group, the average number of weekly interviewees was 50.3 (72%), with minimum 40 and maximum 61 learners.

This teacher-investigator had carried out all the interviews. The type of interview was semi-structured and individual, in order to get "more personalized responses" from the learners (McDonough & McDonough, 1997, p. 184). The prepared questions for the interviews were as follows: "Which materials was interesting to you at the last classes?" "Why was it interesting to you?"

Interviews were held at this investigator's office. On the desk, the sets of classroom used materials were placed to let the interviewee remind them of their experienced interest. The language used for the interviews were English only, because the course did not permit to use other than English as a medium of instruction or conversation. The learners preferred to use English for their language practice, especially with their course teacher—and it seemed to encourage them to attend voluntary interviews.

The method of note-taking was used for recording the interview data. All the interviews were also recorded with permission of the learners. The data of the learners' self-reporting interest in classroom used materials were quantified by this investigator twice. For quantification, this investigator scored a learner's comment on each set of materials with 1 as 'a negative comment of experienced interest', 2 as 'no comment in relating to experienced interest' and 3 as 'a positive comment of experienced interest.' This investigator referred to synonyms and antonyms of 'interesting' for identification, as he did for quantification of the

data from learners' diaries.

Intra-rater reliability was calculated by correlating the two sets of scores for each section in each case of weekly-interview via Pearson r. A total of 17 intra-rater reliabilities were obtained, as 6 weekly- interviews for the learners of section B and section C and 5 weekly- interviews for section A. The mean of 17 intra-rater reliabilities was .86 (SD = .10), with minimum r = .61 and maximum r = .97. These results indicated an acceptable level of reliability.

A total of 1,251 individual data were involved in the scoring of the data collected from the interviewee-learners. For data analyses, two scores on a set of materials were summed, yielding a possible score ranging from 2 to 6. The summed scores ranging from 2 and 3 were categorized as 'comment of not-interesting' and the summed scores ranging from 5 to 6 as 'comment of interesting.' The summed score of 4 means 'no-comment.'

3.2.4 Interest Behavior and Distraction Behavior

The degree of the learners' material-relevant classroom interest was also estimated by their classroom behaviors during the use of materials, to be specific their expression of 'interest' and 'distraction'. These classroom behaviors were recorded by a video-camera, and then scored on observation sheets by this investigator and one independent observer.

The actual video observation had been carried out by this investigator and an independent observer simultaneously for over 2 months. A coding scheme was developed and an observation sheet was used for marking occurrences of individual learners' behaviors of emotional expression of interest and distraction. The markings were carried out only to the period of using a set of materials at classroom. The total observed number of learners were 792 (section A = 264, section B = 275 and section C = 253). An

average of 5.8 learners was observed in one observation case. Across the total observation cases, the average observation to one individual was 11.5 times (SD = 5.69, maximum observation cases=28 times and minimum observation case = 1 time).

The category of 'interest behavior' was regarded as "a high inference descriptor"—that requires observers to infer the learners' feelings from observable behaviors (Nunan, 1992, p. 60). Thus, various bodily signals, that could convey relevant information, were specifically identified in order to secure the precision of marking. For the marking, the bodily signals were assessed altogether to get a total impression of an individual learner's behavior. The following details of bodily signals (adapted from Argyle, 1988) were references for the marking: facial expression: laughing, more smiling (not false or miserable smile), eyebrows down, circular wrinkles, upward movements of mouth, gaze: smiling eyes, more gaze, glances, fixation, dilation and less blink, eyes track and look, gesture: lively movements of hands and shoulders, and head-nods, posture: forward lean, draws back legs, (bodily) orientation: more direct, but side by side for some situations of group or pair work, non-verbal vocalization (tone of voice): higher pitch, upward pitch contours, orotund. In a similar way, the following details (adapted from Argyle, 1988) were references of marking the occurrence of each individual learner's 'distraction': facial expression: blank face, gaze: less gaze, glances, fixation, dilation and more blink, look downwards, irrelevant objects, e.g. a watch or a bag, or outside the window, gesture: more self-touching, e.g. picking and scratching, posture: lowered head, supports head on hand, leans back, stretches legs, turns head away, (bodily) orientation: reduced proximity in pair work, non-verbal vocalization (tone of voice): silence, a lot of long pauses.

The independent observer got an MA degree on descriptive and applied linguistics, so was familiar with the basic notion of

classroom observation. After the pilot viewing, the followings were further agreed. First, a 20-seconds was decided as a unit of occurrence of 'emotional expression of interest' or 'distraction', since those behaviors did not last long at one occasion. Second, learners' behaviour of consulting a dictionary or of asking a question to other students during the pair work was not considered as a cue of distraction, because both observers agreed that those could be parts of normal classroom activities. Third, due to the same reason, the teacher's interruption or intrusion during the pair work or the small group work was not regarded as a cue of distraction of an individual learner.

After finishing the video-observation, the reliability of the marking was evaluated. The inter-rater reliability was assessed via Pearson r. A total of 136 correlations [3 sections x (50 sets of materials—14 sets of materials, which were not recorded)] were obtained for one category. Taken as a group, the 136 reliability coefficients for 'emotional expression of interest' had a mean of .81 (SD = .12). All the reliability coefficients were always higher than .5. The 136 reliability coefficients for 'distraction' had a mean of .82 (SD = .14). All the reliability coefficients never dropped below .4. These results indicated that the inter-rater reliabilities for the observed 'emotional expression of interest' and 'distraction' were satisfactory across all the observation cases.

As the results of the intra-rater reliabilities were satisfactory, it was decided to use both observers' markings. For data analyses, we utilized the 'percentage-score' of an individual learner's 'interest' and 'distraction' behaviors: i) each of average score for individual's behavioral 'interest' and 'distraction' was obtained by calculating the mean of the two observers' counts to each category, ii) to eliminate a possible time- length effect (i.e. the more time a learner was observed, the more frequency might be counted), each average score was divided by the total observed minutes of each observation case, and iii) to meet the equal variance requirement

for further analysis, we transformed every minute-average score of 'interest' and 'distraction' into a percentage score of 'interest' and 'distraction'. A total of 792 percentage-scores of 'interest', and a total of 792 percentage scores of 'distractions' as well were obtained. Finally, the obtained percentage-scores of 'interest' and 'distraction' were used for the analysis.

3.3 Method of Analysis

To check the correlations, the collected qualitative data were transformed into quantitative data. It means that ′a counting approach′ (McDonough & McDonough, 1997, p. 125) was employed to analyse the present qualitative data—related to the data obtained from the quantitative data-collection instruments.

4. Results

4.1 Correlation Between Questionnaires

To calculate the correlation between the scores of the quantitative data-collection instruments, the ratings (ranging 3 to 18) of the daily questionnaires and the ratings (ranging 1 to 7) of the end-of-course questionnaires were utilized.

First, the correlation was calculated between daily questionnaires ratings and end-of-course questionnaires ratings, by treating the same learner-subjects as the cases. The average Pearson r correlation coefficient was computed as .58, indicating a moderate level of correlation of the ratings between the two questionnaire instruments.

The reasons for a moderate level of correlation seem twofold. First, the end-of-course questionnaire simply elicited the learners'

self-reported interest using one Likert-type item only, whereas the daily questionnaire included one negatively worded item among the three items representing interest. Therefore, the end-of-course questionnaire scores might have been affected by some learners' tendency to positive responses, i.e. "response bias" (Skehan, 1989, p. 11), on each set of materials—especially because it was intended to ask the learners to respond to the 48 sets of materials at one time. Second, the individual learners' memory of experienced interest in the materials used in the classroom might have decreased as the course went by. The mean of 12 Pearson r for first set to 12th sets of materials was .33 (SD = .11), the mean of 12 Pearson r for 13th set to 24th sets of materials .37 (SD = .15), the mean of 12 Pearson r for 25th set to 36th sets of materials .39 (SD = .15), and the mean of 12 Pearson r for 37th set to 48th sets of materials .35 (SD = .12). As the average figures of the four periods indicated, the mean correlation of the first period showed the lowest.

4.2 Agreement Between Qualitative Data

The reliability of the learners' self-reported responses between the diaries and the interviews was checked by 'percentage agreement'. We counted the agreement of comments of 'interesting' and comments of 'not-interesting', including 'no-comments', between the two sets of data, and calculated the percentage agreement.

Contingency Table 1 is the source table used for the calculation. Among 2,350 comments in diaries (47 learners x 50 sets of materials) and 1,251 comments in interviewees (average 50.3 learners x average 24.9 sets of materials), a total of 1,016 paired comments were utilized for the calculation. As the table illustrates, among the total of 1,016 paired comments, the percentage of agreement, i.e. matching comments, was 66 % (n =

676). On further analysis, the average percentage of agreement from 47 individual learners was 69 % (percentage SD = 11.35, maximum = 100 %, minimum = 40 %).

TABLE 1
Contingency Table: Diaries and Weekly Interviews

		Comments from the learners' Diaries			
		Interesting	No comment	Not interesting	*Total*
Comments from the weekly Interviews	Interesting	**73*** **(7.2 %)**	162 (15.9 %)	5 (.5 %)	240 (23.6 %)
	No comment	124 (12.2 %)	**599*** **(59.0 %)**	37 (3.6 %)	760 (74.8 %)
	Not-interesting	2 (.2 %)	10 (1.0 %)	**4*** **(.4 %)**	16 (1.6 %)
	Total	199 (19.6 %)	771 (75.9 %)	46 (4.5 %)	1016 (100.0 %)

Note. '*' = agreement of comments between the two instruments

It may be suggested that the figure of 66 (or 69) percentage agreement indicates a moderate level of consistency in the learners' self-reported responses in both qualitative data-collection instruments.

4.3 Agreement Between Qualitative and Quantitative Data

To check the level of agreement in the learners' self-reported responses between the qualitative and the quantitative data-collection instruments, the method of percentage agreement was used. For the calculation, we categorized each individual's rating from the daily questionnaires into 'positive' or 'negative', by checking whether it was above or below an individual learner's median score—estimated from the individual's ratings on the possible 50 sets of questionnaires used during the whole course period. In the same way, we also categorized each individual's scores from the end-of-course questionnaires. When utilizing the qualitative data

of the learners' diaries and the weekly interviews, we excluded the category of 'no-comment', because it could mean either 'marginally interesting - positive' or 'marginally uninteresting - negative'.

Based on the categorization, we counted the number of agreements (the cases of 'positive and interesting' and 'negative and not-interesting') and disagreements (the cases of 'positive and not-interesting' and 'negative and interesting') between the two quantitative and qualitative data-collection instruments. Table 2 shows the agreement and the disagreement between the four data-collection instruments.

The results show that, in all cases of the two data-collection instruments, the percentage of agreement was over 60. It might be pointed that this percentage showed a moderate level of agreement, in comparison with the average correlation (average r = .58) between the questionnaire data, which showed 62 percent agreement. In Peacock's study (1997), 69 percent agreement between the data of interviews and questionnaires was obtained from Korean adult learners of English, and he proposes that the figure was some indication of a reasonably high level of learner truthfulness in interviews.

TABLE 2
Percentage of Agreement between the Data-collection Instruments

		Agreement (n =)	Disagreement(n =)	*total*
Day-by-Day Questionnaire with	Diaries	68.2 % (379)	31.8 % (177)	556
	Interviews	65.0 % (201)	35.0 % (108)	309
End-of-course Questionnaires with	Diaries	64.2 % (326)	35.8 % (182)	508
	Interviews	61.4 % (156)	38.6 % (98)	254
Between questionnaires(r = .5863)		61.9 % (1466)	38.0 % (902)	2368
Between diaries and interviews *		68.5 % (858)	31.5 % (393)	1251

Note. '*n*'= number of matching or not-matching cases, '*' = the comparison between the diaries and the interviews including 'no comment' data.

Among the results, the end-of-course questionnaire instrument showed the lowest level of agreement with all the other instruments, below 65 percent agreement. The low level of agreement seemed to be caused by the relatively less reliable procedure and design of the questionnaire: the end-of-course questionnaire was administered at one time and employed one item only to measure self-reported interest. Of the qualitative data-collection methods, the diary instrument showed a slightly higher level of agreement with both questionnaires than did the interview instrument. It seems that the learners in the present study were slightly more comfortable in reporting their experienced interest in 'writing' rather than by 'talking'. Another possible reason might be that not all the 50 sets of course materials were asked about in the interviews (about 25 sets of materials were covered), whereas the learners were requested to write in their diaries for the whole course period and so covered more materials.

The overall results of percentage agreement suggest that the learners who took part in the four data-collection instruments in the present investigation recorded their self-reported interest in a moderately consistent way.

4.4 Agreement Between Behavior and Self-Reporting Data

We also check correlations between the observed behaviors and the learners' self-reporting responses. From quantitative self-reporting instruments, the interest-score of day-by-day questionnaires and the end-of-course questionnaires are utilized for estimating correlations. Since the observation score is standardized by minute and percentized by referring to the observation scores of the group, we calculate a group correlation.

The Pearson r correlation coefficients between observed 'interest' and the interest-score of day-by-day questionnaire is computed as $r = .15$ ($n = 751$), and between observed 'interest'and the

interest-score of end-of-course questionnaire as r = .12 (n = 591). In case of observed behaviour of 'distraction', Pearson r correlation coefficients with the interest-score of day-by-day questionnaire is computed as r = - .08 (n = 751), and with the interest-score of end-of-course questionnaire as r = - .17 (n = 591).

On the surface, the levels of estimated correlations indicate there are no or virtually no relationship between observation scores and scores of self-reporting questionnaires. In other words, the learners' behaviors and their self-reporting interest in questionnaires are different variables. However, the cross comparison of correlations shows that the scores of observed 'interest' are positively correlated with the scores of each of questionnaires, whereas the scores of observed 'distraction' are negatively correlated. This pattern of results does not contradict with our common-sense notion that 'interest' and 'distraction' are opposed behaviors of learners. Therefore, we suggest that the learners' behaviors of 'interest' and 'distraction' behaviors during the use of materials should not be uncorrelated with their self-reporting interest in the questionnaires after the use of materials.

In case of qualitative data-collection instruments, the 'percentage agreement' between learners' diaries and observed behaviors is checked. The data of weekly learner-interviews are excluded in checking the level of agreement, because of the small number of valid cases for the estimation: There are only 71 valid cases, i.e. the cases of the same learners' observation-data and interview-data. For the estimation of the percentage agreement, we categorize an individual percentage score of observed 'interest' into 'positive' or 'negative', by checking whether it is above or below the group's median percentage-score. The same method of median-split is used for categorizing the percentage scores of observed 'distraction', but as 'distraction' is opposed to 'interest', the percentage score above median is categorzied as 'negative', the below median as 'positive'.

The results of counting the number of agreements (the cases of 'positive and interesting' and 'negative and not-interesting') and disagreements (the cases of 'positive and not-interesting' and 'negative and interesting') between the observed behaviours and diary-data are as follows:

Between observed 'interest' and learners' diaries: Agreement = 52 % (n = 75), Disagreement = 47 % (n = 67)—total 142 cases
Between observed 'distraction' and learners' diaries: Agreement = 54 % (n = 77), Disagreement = 45 % (n = 65)—total 142 cases

In considering that the levels of agreement between diary-data and other learners' self-reporting-data are over 60 percentage, the results of below 60 percentage agreement indicate that there is no or virtually no relationship between observed 'interest/ distraction' and self-reporting diaries. However, we still can find that the converted data of 'distraction' show a positive percentage of agreement with dairy-data. Overall the results, therefore, imply that 'interest' and 'distraction' are opposed behaviors of the classroom learners, and probably they are not essentially uncorrelated with the learners' self-reporting interest expressed via diaries.

5. Conclusion

In the present study, we checked the methodological triangulation on learners′ interest in the classroom used materials via learner's daily questionnaire, learner's end-of-semester questionnaire, learners' diaries, semi-structured weekly learner-interviews, and observation of learners' classroom behaviors of interest and distraction during

the use of materials.

A classroom-based research was suitable for the present research. First, it has been often argued that the collected learner-data on the effect of certain type of materials should be more reliable, if materials would be once used to learners at classroom setting (Rea-Dickins, 1994; van Els, Bongaerts, Extra, van Os, & Janssen-van Dieten, 1984). Second, learners' interest generated by materials might be often likely to disappear (Mitchell, 1993) and so become difficult to detect, especially if they would be located in non-classroom settings. We planned to utilize real classroom situations for the investigation, but did not attempt any random assignment of subjects for grouping learners, so research was a quasi-experimental type (Nunan, 1992). McDonough and McDonough (1997) mentions that a quasi-experiment design is more usual in real-world ELT situations outside a psycholinguistic laboratory.

The overall results of the present study suggest that the learners who took part in the four data-collection instruments responded their self-reporting interest in a modestly consistent way. The modest level of agreement between the data collection instruments explains that the same pattern of result has been consistently found from the data of each instrument. It propose that various data collection instruments should be used to measure leaners' affective domain in relating to ELT materials.

References

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology, 87,* 559-575.

Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom.* Cambridge: Cambridge University Press.

Argyle, M. (1988). *Bodily communication.* New York: Methuen & Co. Ltd.

Breen, M. P., & Candlin, C. N. (1987). Which materials?: A consumer's and designer's guide. In L. E. Sheldon (Ed.), *ELT textbooks and materials: Problems in evaluation and development* (ELT Documents 126) (pp. 13-28), London: Modern English Publications/The British Council.

Deaux, M. K. (1993). *User evaluation of beginning-level Japanese-language textbooks.* Unpublished Ph.D. Dissertation. Temple University, USA.

Dörnyei, Z (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*, 273-284.

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching, 31*, 117-135.

Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal, 51*, 36-42.

Gardner, R. C., & Tremblay, P. F. (1994). On motivation, research agendas, and theoretical frameworks. *The Modern Language Journal, 78*, 359-369.

Gliksman, L., Gardner, R. C., & Smythe, P. C. (1982). The role of the integrative motive on students' participation in the French classroom. *Canadian Modern Language Review, 38*, 625-647.

Hutchinson, T. (1987). What's underneath?: An interactive view of materials evaluation. In L. E. Sheldon (Ed.), *ELT textbooks and materials: Problems in evaluation and development* (ELT Documents 126) (pp. 37-44). London: Modern English Publications/The British Council.

Hutchinson, T., & Klepač, M. (1982). The communicative approach: A question of materials or attitudes? *System, 10*, 135-143.

Jenks, F. L. (1981). Learners' needs and the selection of compatible materials. In J. E. Alatis, H. B. Altman & P. M. Alatis (Eds.), *The second language classroom: Directions for the 1980's* (pp. 221-226). New York: Oxford University Press.

Kim, H. (2001). *The opinions, classroom interest and motivational orientation of students on a general English course at South Korean universities with regard to the materials employed to teach them.* Unpublished Ph.D. Thesis. Essex University, UK.

Krapp, A., Hidi, S. & Renninger, K. A. (1992). Interest, learning and

development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3-26). New Jersey: Lawrence Erlbaum Associates, Publishers.

Lee, K-R. & Oxford, R. (2007). Triangulating think-alouds and questionnaires in reading strategy assessment, *English Teaching, 62*(3), 331-356.

McDonough, J., & Shaw, C. (2003). *Materials and method in ELT: A teacher's guide* (2nd ed). Oxford: Blackwell.

McDonough, J., & McDonough, S. (1997). *Research methods for English language teachers*. London: Edward Arnold.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational psychology, 85*, 424-436.

Murphy, D. (1993). Evaluating language learning tasks in the classroom. In G. Crookes. & S. Gass (Eds.), *Tasks in a pedagogical context: Integrating theory and practice* (pp. 139-161). Clevedon: Multilingual Matters.

Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.

O'Sullivan, T. (1991). Foreign language coursebooks: Ask your pupils! *Language Learning Journal, 3*, 10-13.

Peacock, M. (1996). *The motivation of adult EFL learners with authentic materials and artificial materials*. Unpublished Ph.D. Thesis. Essex University, UK.

Peacock, M. (1997) Truthfulness in learner interviews: An empirical investigation. *Essex Graduate Student Papers in Language and Linguistics, 1*, 70-77.

Purpura, J., & Pinkly, D. (1991). *On target 1.* New York: ScottForesman.

Rea-Dickins, P. (1994). Evaluation and English language teaching. *Language Teaching, 27*, 71-91.

Savignon, S. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA.: Addison-Welsey.

Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151-182). New Jersey: Lawrence Erlbaum Associates, Publishers.

Skehan, P. (1989). *Individual differences in second-language learning*.

London: Edward Arnold.
Tobias, S. (1995). Interest and metacognitive word knowledge. *Journal of Educational Psychology, 87*, 399-405.
van Els, T., Bongaerts, T., Extra, G., van Os, C., & Janssen-van Dieten, A-M. (1984). *Applied linguistics and the learning and teaching of foreign languages*, London: Edward Arnold.
Yorio, C. A. (1986). Consumerism in second language learning and teaching. *Canadian Modern Language Review, 42*, 668-687.

Haedong Kim
Hankuk University of Foreign Studies
270 Imun-dong, Dongdaemun-gu
Seoul, 130-791 Korea
02) 2173-3017, 019-844-3715
khd@hufs.ac.kr

논문접수일자: 2007.10.15
게재확정일자: 2007.12.1