

A Study on Short-answer-question Format in Testing EFL Learners' L2 Reading Ability

Dae-Min Kang
(Seoul National University)

Kang, Dae-Min. (2005). A study on short-answer-question format in testing EFL learners' L2 reading ability. *Language Research* 41(4), 963-983.

This study investigated the substitutability of short-answer-question (SAQ) format for multiple-choice-question (MCQ) format in measuring EFL learners' L2 reading ability and the existence of differences in SAQ format test-taking strategy use among different reading ability groups. A total of 120 Korean high school students were asked to take a 7-question SAQ format reading test and then, immediately after answering each item, write down anything that they thought were going on in their mind when trying to give the answer, i.e., test-taking strategies. The study revealed that SAQ format could be used as an effective alternative to MCQ format with its stronger construct validity in assessing EFL test-takers' reading comprehension ability and learners with poor reading skill could benefit from instruction of SAQ format test-taking strategies, though tentatively and restrictedly.

Key words: short-answer-question (SAQ) format, multiple-choice-question (MCQ) format, test-taking strategy use, EFL learners' L2 reading ability, construct validity

1. Introduction

The testing format for EFL learners' L2 reading ability, often the only avenue through which the learners are readily provided access to English (Celce-Murcia & Olshtain, 2000), in most of large-scale standardized tests is MCQ format (Peirce, 1992; Alderson, 2000). The format, despite its high degree of reliability and practicality, has the danger of assessing test-takers' reading ability inadequately (Alderson, 2000) and its validity could be compromised because test-takers are allowed to rely on guessing rather than tapping their reading ability per se, though. In order to solve this problem, notably an investigation of summary task was

conducted (Joh, 1999) which, notwithstanding some meaningful results, failed to assess all of the three components of reading comprehension ability, i.e., factual, inferential, and evaluative reading ability (Lee, 2003).

The purpose of the present study, undertaken to fill the gap, was to study whether SAQ format could function as a viable alternative to the currently prevalent MCQ format in testing EFL learners' L2 reading comprehension faculty by answering the following research questions.

- (1) What test-taking strategies do EFL learners use when taking SAQ format reading comprehension test?
- (2) Do EFL learners use different test-taking strategies for different types of questions constructed in SAQ format to measure factual, inferential, and evaluative reading ability, which is in relation to construct validation of the testing format?
- (3) Do EFL learners at different L2 reading ability levels use different test-taking strategies for different types of questions constructed in SAQ format to measure factual, inferential, and evaluative reading ability?

2. Literature Review

While MCQ format has been favored over others in testing EFL candidates' reading faculty thanks to its higher degree of reliability, it does not assess the ability a test was originally designed to measure adequately (Henning, 1987).

In order to find solutions to this lack of validity, researchers have undertaken various studies including those that confirmed the validity of summary task (Bean & Steenwyk, 1984; King, Biggs, & Lipsky, 1984; Rinehart, Stahl, & Erickson, 1986; Head, Readence, & Buss, 1989; Hill, 1991; Kirkland & Saunders, 1991; Rinehart & Thomas, 1993; Joh, 1999) whose reliability could be threatened, though, if rated in terms of summary quality (Alderson, 2000). Although the investigations found that summary task ruled out the possibility of test-takers' use of guessing and thus secured more validity than MCQ format, however, it seems that the very nature of summarizing might have posed another threat to appro-

priate measurement of EFL learners' L2 reading ability. That is, since summary task usually requires test-takers' concentration on explicitly written messages of text alone, it runs the risk of assessing only factual reading ability to the exclusion of inferential and evaluative reading ability. According to Lee (2003), in the activity of reading seeking out real meanings of text not expressed on the surface is more important than understanding literal messages. In this respect, it appears that construct validity, which has to do with the ability of a test to measure the psychological construct the test developers designed the test to assess (Ebel & Frisbie, 1991), of summary task limits the format in becoming a proper testing format for EFL test-takers' reading ability.

In contrast, SAQ format – composed of “items that are open-ended, where the candidates have to think up the answer for themselves” (Alderson, Clapham, & Wall, 1995, p. 57) and more authentic than any other formats (Alderson, 2000) – promises to bridge the gap opened by summary task and MCQ format in terms of construct validity, since the questions can be constructed in a way to measure inferential and evaluative reading ability besides factual reading faculty. Empirical verification of construct validity of SAQ format, however, would be needed in concluding whether SAQ format has stronger construct validity than summary task and MCQ format. Particularly, investigation of candidates' test-taking strategies (Nevo, 1989; Cohen, 1999; Lee, 2002) seems relevant for construct validation of SAQ format (Bachman, 1990) because it can demonstrate test-takers' reading process (Pritchard, 1990; Weir, 1994; Urquhart & Weir, 1998; Alderson, 2000) most clearly. Although strategies for MCQ format, gap-filling test, and discourse cloze have been studied (Farr, Pritchard, & Smitten, 1990; Allan, 1992; Storey, 1994, 1997), though, very few, if any, inquiries have been conducted for SAQ format, let alone in EFL setting. Examining candidates' test-taking strategies in investigating construct validity of SAQ format test (Bachman, 1990), the present study probed into whether test-takers' strategies for SAQ format test would reflect the operationalization of each of the three reading ability, i.e., construct of reading comprehension ability.

Given this virtual unavailability of information on test-taking strategies for SAQ format test, it seems necessary to conduct a relevant study which can identify for confirmation of presence of strong construct validity in detail what strategies EFL learners at different L2 reading ability levels use in what ways when dealing with items constructed in SAQ

format, which will reveal whether SAQ format can be relied on as an alternative to MCQ format.

3. Method

3.1. Subjects

A total of 120 male high school students in Seoul, Korea participated in the study, out of whom 30 took part in the pretest and 90 in the main test. Both tests took place in June, 2004. The students were all juniors from the same school and assigned to three different reading ability groups. The first criterion for the grouping was quantitative in nature; their scores from both reading sections of English test in a national-level examination administered in the previous year of 2003 and school mid-term examination for the first semester of 2004. Reading section accounted for 34% for the former and 50% for the latter. The second criterion for the assignment was qualitative in nature; the teacher's evaluation of the students' reading faculty collected through interview. The students' self-assessment of their reading ability, which is a valid as well as reliable instrument (Ellis, 2003) and gives useful information on student abilities (Alderson, 2000), was the third criterion and also qualitative in nature.

In determining cut scores for the three reader groups, first candidates' mean scores from the two aforementioned tests were considered since they were norm-referenced (Livingston, 1972). Specifically, the upper and lower limit scores of one standard deviation of the mean in the tests were used as cut scores. That is, participants whose scores fell within one standard deviation of the mean score were assigned to intermediate reader group while those outside that range were classified as either good or poor readers. Given the arbitrary, unrealistic, and illogical nature of setting a cut score (Glass, 1978; Jaeger, 1989), however, it was decided that besides the scores qualitative criteria should be used.

3.2. Materials

For the SAQ format test, the text from Joh (1999) was employed (see Appendix 1). The researcher argued that the text about women's longer life expectancy than men was appropriate for testing reading ability of

Korean high school juniors when considered on the basis of the national curriculum. The two peer reviewers for the current study, a Korean and an American English teachers at the high school, supported her assertion. In addition, it was agreed that the content of the text was culturally neutral, an important consideration since cultural familiarity influences reading comprehension substantially (Kembo, 2001).

From the text, seven English items were prepared in SAQ format for pretesting purpose, a necessary step since predictions of all answers to as well as interpretations of items are much difficult (Alderson, 2000). The questions were designed into three types in order to test subjects' factual (Type I questions), inferential (Type II questions), and evaluative reading comprehension ability (Type III). The three types of items, peer-reviewed so as to minimize subjectivity which usually affects item construction (Pilliner, 1977), were expected to help the author determine whether SAQ format test consisting of the types could sufficiently assess the three components of reading comprehension ability and thus establish its construct validity in measuring subjects' reading faculty. As the language for output, test-takers were allowed to employ Korean. The decision to let them utilize their L1 was made because responding in L1 would be less difficult (Bachman & Palmer, 1996) and construct validity of the reading test in SAQ format could be threatened with L2 use when giving answers in writing. After the pretest, several items were modified and also peer-reviewed for the main test.

As a way of securing reliability of the test, answer keys – words, phrases, or a sentence involving core words or concepts (Wang, 2001) – were devised in Korean (see Appendix 2 for their English translations). Based on the answer keys, criteria for giving scores of 0 to 4 were developed. The two raters, English teachers from another high school, double-marked candidates' responses. For both pre- and main tests, processes of constructing answer keys, developing scoring criteria, and double-marking were the same.

3.3. Procedure

For pretest, students were given verbal instruction in Korean on how to answer the SAQ format questions (Alderson, Clapham, & Wall, 1995) for the purpose of enhancing their test performance (Bachman & Palmer, 1996). The subjects were also briefed on how to report in writing in

Korean their test-taking strategies in an open-ended way (Alderson, 2000) because writing in L2 could blur evidence of comprehension and interfere with recall (Lee, 1986). More specifically, the students were requested to write down anything that they thought went on inside their mind while trying to give responses to items right after answering each question, in consideration of recency for effective introspection. This method of introspection (Feldman & Stemmer, 1987), classified as self-observation according to Cohen's taxonomy of introspection (1987), was argued to be indispensable in validating a test for its ability to probe into candidates' test-taking processes and thus deepen our understanding of the construct validity of the test (Storey, 1994). The briefing was held right before the test that ran for 45 minutes during an after-hour class. Furthermore, the test-takers were asked to practice much sincerity since research subjects tend to choose test-taking strategies different from the ones they would normally employ in actual school tests (Cohen, 1999).

After the pretest, the students' responses were, while being scored, reviewed to see if they actually responded in ways they had been expected to. Following the review, it was determined that several questions needed modifying and thus changes were made. Additionally, the subjects' test-taking strategies were analyzed and as a result some modifications were made in the content of strategy reporting instruction for main test.

The 90 students for main test were also instructed and briefed on test-taking in SAQ format and reporting of test-taking strategies. Each of them was interviewed on the following day in a semi-structured way for self-assessment of L2 reading ability, about the SAQ format test, and their strategies for and attitudes towards the test. The teacher was also interviewed a day before both pre- and main tests for his evaluation of the students' individual English reading ability.

3.4. Data Analysis

Analysis of data from the test was made in the following sequence.

- (1) Subjects' test-taking strategies reported in writing were reviewed and coded.
- (2) The coded strategies were analyzed for their respective frequency of use across question type.

- (3) The data from the written reports were analyzed to identify differences in strategy use of different reading ability groups.
- (4) Interviews with the teacher and students were reviewed to complement the quantitatively analyzed data.

4. Results

4.1. Test-taking Strategies

It was found that strategies used by the subjects were rather broad and not so informative in content as well as restricted in number. Despite the request through briefing to record anything flowing through their mind while trying to answer each item, it was revealed by the post-test interview that the participants' unfamiliarity with SAQ format test caused them to apply only the limited number of strategies, many of which they would have utilized in MCQ-format tests as well. Meanwhile, the broadness of strategies in terms of content was attributed to the candidates' lack of reporting ability in their LI, the interview also showed.

The subjects' scores mostly corresponded with their L2 reading ability evaluated by their teacher and themselves. However, several members of intermediate and poor reader groups earned scores similar to those found in either the higher ability group, for the poor-level readers, or lower ability group, for the intermediate-level readers. To be particularly noted is that those intermediate-level candidates who scored lower than even a few of poor-level readers were found to have used the strategy of 'guessing with no specific ideas,' which was Korean EFL university students' third most frequently employed strategy in MCQ format test (Lee, 2002). It seems that the use of the strategy, a case of strategy transfer from one testing format to another, affected these test-takers negatively since 'guessing with no specific ideas' could hardly help them give correct answers in SAQ format test, unlike in MCQ format test.

Among the other test-taking strategies used by subjects, presented in Table 1, 'making careful inferences based on the text and the question' (57.5%) and 'taking into consideration the writer's motives and attitudes for passages' (11.7%) were the two most popular strategies. They ac-

Table 1. Classification of SAQ Format Test-taking Strategies (%)

1	guessing with no specific ideas	1.6
2	taking note of the key words in the question and reviewing the passages preceding or following the words in the text	6.4
3	searching for possible clues from answers to other questions	3.5
4	pre-reading all the questions	3.0
5	rereading the question	0.9
6	reading other questions	3.3
7	rereading the text	3.7
8	making careful inferences based on the text and the question	57.5
9	taking into consideration the writer's motives and attitudes for passages	11.7
10	modifying the already written answer	0.9
11	relying on background knowledge	7.5

counted for 69.2% of all strategy uses, when combined. Their percentile superiority indicates that candidates favored higher-level strategies than revealed in Lee's 2002 study in which Korean university students reported 'selecting a response because it is stated in the text' (61.5%) and 'selecting a response based on understanding the material read' (16.9%) as their two most favored strategies for MCQ format test. Also interesting is the difference in percentile terms between two similar strategies in Lee's inquiry and current investigation; 'selecting a response because it is stated in the text' (61.5%) in Lee's probe and 'taking note of the key words in the question and reviewing the passages preceding or following the words in the text' (6.4%) in the present study. It seems that the difference was caused primarily by the dissimilarity in testing formats of the two studies. In other words, while MCQ format test allowed the test-takers to guess correct answers through linking options with the text content, the use of the strategy was greatly discouraged for the SAQ format test because it had no options. In fact, candidates in MCQ format test were revealed to do a surface-structure reading of items (Cohen, 1984) in rather sharp contrast to subjects' preference for in-depth reading and inferring found here.

The third most often used strategy was 'relying on background knowledge' (7.5%). It was decided that use of background knowledge, which fa-

ilitates reading (Alderson, 2000), was not a source of test bias for the current study since the SAQ format test had been constructed to assess generalized reading ability and the candidates did not have different content specializations (Bachman, 1990). In the post-test interview, all of the students reported that the text content of the test, i.e., women's relative longevity in comparison with men, enabled them to employ their background knowledge. The third place in ranking as it was, the strategy was identified as not having had any major impact on the subjects since its percentage amounted to just over one-tenths of the top two strategies combined.

4.2. Use of Strategy Depending on Question Type and Construct Validation of SAQ Format Test

From the analysis of the data, it was revealed that there existed different patterns of strategy use according to question type as presented in Table 2. The most popular strategy of 'making careful inferences based on the text and the question' (362 reports or 57.5% out of the total) was used more frequently for Type II questions (160 reports), constructed to measure subjects' inferential reading ability, than for Type I questions (85 reports), written to test candidate's factual reading faculty, or Type III questions (117 reports), designed to assess test-takers' evaluative reading ability. 'Taking into consideration the writer's motives and attitudes for passages,' the second in ranking, was applied more often for Type III questions (37 reports) than for Type II questions (29 reports) and Type I questions (8 reports). The third- and fourth-ranked strategies, 'relying on background knowledge' and 'taking note of the key words in the question and reviewing the passages preceding or following the words in the text,' were also used repeatedly for Type II and Type I questions, respectively.

These results demonstrate that the three types of questions in the SAQ format test prompted the test-takers to employ the very set of faculties the questions had been constructed to measure and thus construct validity of SAQ format test in assessing EFL learners' L2 reading ability was established, though much tentatively and restrictedly. The findings further indicate that SAQ format with its greater construct validity could function as an effective alternative to MCQ format test, provided reliability is secured adequately.

Table 2. Frequencies of Reported Strategy Use Depending on Question Type

Strategies	Question Type			
	Type* I (k*=2)	Type II (k=3)	Type III (k=2)	Total
guessing with no specific ideas	10			10
taking note of the key words in the question and reviewing the passages preceding or following the words in the text	22	18		40
searching for possible clues from answers to other questions	10	12		22
pre-reading all the questions	6	5	8	19
rereading the question	1	5		6
reading other questions	7	11	3	21
rereading the text	10	3	10	23
making careful inferences based on the text and the question	85	160	117	362 (57.5%)
taking into consideration the writer's motives and attitudes for passages	8	29	37	74 (11.7%)
modifying the already written answer	3	3		6
relying on background knowledge	18	24	5	47
Total	180 (28.6%)	270 (42.8%)	180 (28.6%)	630

* Type I, II, III for measuring factual, inferential, and evaluative reading comprehension ability, respectively

k = number of items for each type

4.3. Difference in Use of Strategy by Different Reader Groups Depending on Question Type

In addition to construct validation of the testing format, for pedagogical applicability the data were analyzed to examine whether there existed different patterns of strategy use by different reading ability groups according to question type. More specifically, analysis was made to see if the need to give explicit instruction in relevant SAQ format test-taking strategies to EFL students at lower reading ability level could be justified. Notwithstanding ongoing pros and cons, results from the current investigation show that poor readers stand to gain benefit from the strategy teaching, if it will be carried out with such care as to make

it clear to them that learning the strategies should be considered only as a means for ultimately achieving the goal of enhanced reading ability.

Table 3 shows how different reader groups used different test-taking strategies for different types of questions. The first-ranked strategy of 'making careful inferences based on the text and the question,' related to Type II questions, was found to have been favored by good readers for most of Type II items (90.9%). This percentage is remarkable when compared to 33.3% for poor reader group and 69.5% for intermediate reader group. The good readers also noticeably and correctly preferred the second-ranked strategy of 'taking into consideration the writer's motives and attitudes for passages' to other strategies when trying to give answers to Type III items (81.8%). Meanwhile, intermediate and poor reader groups lagged far behind, 20.2% and 0% respectively. The strategy of 'taking note of the key words in the question and reviewing the passages preceding or following the words in the text,' associated with Type I questions, was again used more often for Type I items by good readers (45.5%) than by either intermediate (12.8%) or poor readers (0%).

These differences in strategy use by three reader groups demonstrate that the lower the test-takers' reading faculty, the less appropriate strategies they employed. In fact, students with good reading ability had not been taught explicitly in class what strategies to use for different types of questions in the SAQ format test, yet they proved themselves excellent at using relevant strategies depending on item type unlike intermediate or poor readers.

In terms of pedagogical target group, poor reader group was found to be in the acutest need of learning how to use different strategies properly as evidenced by the percentile differences between the group and good reader group – 45.5%, 57.6%, and 81.8% for Type I, II, and III questions respectively, though the gaps between good and intermediate reader groups were also wide – 32.7%, 21.4%, and 61.6% for Type I, II, and III items respectively – and thus the group also appears to have much to learn in terms of appropriate use of test-taking strategy.

Table 3. Frequencies of Reported Strategy Use by Different Reader Groups

Strategies	Reader Group									Total
	Group I* (k=11)			Group II (k=47)			Group III (k=32)			
	T I* (k*=2)	T II (k=3)	T III (k=2)	T I (k=2)	T II (k=3)	T III (k=2)	T I (k=2)	T II (k=3)	T III (k=2)	
guessing with no specific ideas	0	0	0	8	0	0	2	0	0	10
taking note of the key words in the question and reviewing the passages preceding or following the words in the text	10 (45.5%)	0	0	12 (12.8%)	0	0	0 (0%)	18	0	40
searching for possible clues from answers to other questions	2	0	0	6	7	0	2	5	0	22
pre-reading all the questions	2	3	2	2	2	4	2	0	2	19
rereading the question	1	0	0	0	0	0	0	5	0	6
reading other questions	0	0	0	1	5	3	6	6	0	21
rereading the text	0	0	0	10	0	8	0	3	2	23
making careful inferences based on the text and the question	7	30 (90.9%)	2	34	98 (69.5%)	60	44	32 (33.3%)	55	362
taking into consideration the writer's motives and attitudes for passages	0	0	18 (81.8%)	0	29	19 (20.2%)	8	0	0 (0%)	74
modifying the already written answer	0	0	0	3	0	0	0	3	0	6
relying on background knowledge	0	0	0	18	0	0	0	24	5	47
Total	22 (100%)	33 (100%)	22 (100%)	94 (100%)	141 (100%)	94 (100%)	64 (100%)	96 (100%)	64 (100%)	630

* Group I, II, and III denote good-, immediate-, and poor-reader groups, while T I, T II, and T III denote question types I, II, and III, respectively, as shown in Table 2

k = number of test-takers in each group

5. Conclusion

With the ultimate view to investigating into whether SAQ format test could be effectively utilized as an alternative to MCQ format in measuring EFL learners' L2 reading ability, the current study inquired into SAQ format test-taking strategies of 90 Korean high school students who were assigned to good, intermediate, or poor reader groups. Results of the study show that the three types of questions in the test caused subjects to use what the questions had been prepared to assess, which means that construct validation of SAQ format test in the students' L2 reading ability was made, though limitedly and tentatively. Analysis of data from the investigation also demonstrates that there existed inter-group differences in terms of use of relevant strategy in which good reader group applied strategies the most appropriate to each question type in contrast to lower level readers, more specifically poor readers.

Implications from the current study can be considered from two areas: application of SAQ format to EFL reading ability testing in classroom or beyond and pedagogical promises of instruction in SAQ format test-taking strategies.

First, with respect to introduction of SAQ format into national-level testing of L2 reading faculty, it seems that practicality and reliability should be taken into account. Without strong reliability achieved through examiner training and inter- and intra-rater reliability monitoring (Alderson, Clapham, & Wall, 1995) and the resulting balance between validity and reliability (Pilliner, 1977), SAQ format could not be employed for high-stake, large-scale examination such as College Scholastic Ability Test in Korea. Concerning application of SAQ format to classroom testing setting, washback effect from national-level test in which SAQ format is employed should trigger its use at classroom level. Irrespective of washback effect, however, teachers in classroom context need to employ SAQ format for their EFL students' enhanced reading ability which will be discussed below.

Second, regarding instruction of SAQ format test-taking strategies for EFL learners, an approach different from the case of MCQ format should be taken because unlike the MCQ format the present investigation found that there was a correspondence between the very set of ability the SAQ format test items had been constructed to measure and the test-taking

strategies used by good readers, particularly in the areas of inferential and evaluative reading ability. This finding indicates that lower level readers, particularly poor readers, could be taught relevant SAQ format test-taking strategies to their advantage. In the long run, the result implies, instruction of SAQ format test-taking strategies will build EFL learners' L2 reading ability. Additionally, teaching of the strategies will not reduce learners to test-wise candidates because it is highly expected to provide them with opportunities to develop the essential, underlying ability in the area of L2 reading.

Regardless of the interesting findings and implications, the current study needs to be interpreted and treated carefully in that not only was it conducted with a restricted number of subjects and test items but also there will exist limitations in observation and quantification (Bachman, 1990) when reading is assessed. Despite this and other potential weaknesses, it is highly hoped that the present investigation will lead to more sophisticated and meaningful researches on SAQ format including those on possible gender differences in terms of SAQ format test-taking strategy use. Such studies will contribute greatly to firm confirmation of usefulness of SAQ format test in diagnosing EFL learners' L2 reading ability effectively and adequately.

References

- Alderson, J. C. (2000) *Assessing Reading*. Cambridge, England: Cambridge University Press.
- Alderson, J. C., Clapham, C., and D. Wall (1995) *Language Test Construction and Evaluation*. Cambridge, England: Cambridge University Press.
- Allan, A. I. C. G. (1992) *EFL Reading Comprehension Test Validation: Investigating Aspects of Process Approaches*. Unpublished doctoral dissertation. Lancaster University, England.
- Bachman, L. F. (1990) *Fundamental Considerations in Language Testing*. Oxford, England: Oxford University Press.
- Bachman, L. F. and A. S. Palmer (1996) *Language Testing in Practice*. Oxford, England: Oxford University Press.
- Bean, I. W. and F. L. Steenwyk (1984) The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior* 16, 297-306.

- Celce-Murcia, M. and E. Olshtain (2000) *Discourse and Context in Language Teaching: A Guide for Language Teachers*. Cambridge, England: Cambridge University Press.
- Cohen, A. D. (1984) On taking language test: What the students report. *Language Testing* 1, 70-81.
- Cohen, A. D. (1987) Studying learner strategies: How we get the information. In A. Wenden and J. Rubin, eds., *Learner Strategies in Language Learning* (pp. 31-40). Englewood Cliffs, NJ: Prentice-Hall.
- Cohen, A. D. (1999) Strategies and processes in test taking and SLA. In L. F. Bachman and A. D. Cohen, eds., *Interfaces Between Second Language Acquisition and Language Testing Research* (pp. 90-111). Cambridge, England: Cambridge University Press.
- Ebel, R. L. and D. A. Frisbie (1991) *Essentials of Educational Measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Ellis, R. (2003) *Task-based Language Learning and Teaching*. Oxford, England: Oxford University Press.
- Farr, R., Pritchard, R., and B. Smitten (1990) A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement* 27, 209-226.
- Feldmann, U. and B. Stemmer (1987) Thin_aloud a_retrospective da_in C-Te_taking: Diffe_languages diff__learners - sa_approaches? In C. Faerch and G. Kasper, eds., *Introspection in Second Language Research* (pp. 251-267). Clevedon, UK: Multilingual Matters.
- Glass, G. V. (1978) Standards and criteria. *Journal of Educational Measurement*, 15, 237-261.
- Head, M. H., Readence, J. E., and R. R. Buss (1989) An examination of summary writing as a measure of reading comprehension. *Reading Research and Instruction* 28, 1-11.
- Henning, G. (1987) *A Guide to Language Testing*. Cambridge, MA: Newbury House.
- Hill, M. (1991) Writing summaries promotes thinking and learning across the curriculum - but why are they so difficult to write? *Journal of Reading* 34, 536-539.
- Jaeger, R. M. (1989) Certification of student competence. In R. L. Linn, ed., *Educational Measurement* (pp. 485-514). New York: American Council on Education & Macmillan.
- Joh, J. (1999) An analysis of performance in two different measures of reading comprehension: A preliminary study. *English Teaching* 54,

239-258.

- Kembo, J. A. (2001) Testing of inferencing behaviour in a second language. *International Journal of Bilingual Education and Bilingualism* 4, 77-96.
- King, J. R., Biggs, S., and S. Lipsky (1984) Students' self-questioning and summarizing as reading study strategies. *Journal of Reading Behavior* 16, 297-306.
- Kirkland, M. R. and M. A. P. Saunders (1991) Maximizing student performance in summary writing: managing cognitive load. *TESOL Quarterly* 25, 105-121.
- Lee, J. (2002) A study on multiple-choice test-taking strategies in English reading comprehension. *English Language Teaching* 14, 45-72.
- Lee, J. F. (1986) On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition* 8, 201-212.
- Lee, W. (2003) *A Guide to English Language Testing*. Seoul, Korea: Moonjin Media.
- Livingston, S. A. (1972) Criterion-referenced applications of classical test theory. *Journal of Educational Measurement* 9, 13-26.
- Nevo, N. (1989) Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing* 6, 199-215.
- Peirce, B. (1992) Demystifying the TOEFL reading test. *TESOL Quarterly* 26, 665-691.
- Pilliner, A. (1977) Testing English as a foreign language: A general study. *PASAA* 7, 315-348.
- Pritchard, R. (1990) The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly* 25, 273-295.
- Rinehart, S. D., Stahl, S. A., and L. G. Erickson (1986) Some effects of summarization training on reading and studying. *Reading Research Quarterly* 21, 422-438.
- Rinehart, S. D. and K. F. Thomas (1993) Summarization ability and text recall by novice students. *Reading Research and Instruction* 32, 24-32.
- Storey, P. (1994) *Investigating Construct Validity through Test-taker Introspection*. Unpublished doctoral dissertation. University of Reading, England.
- Storey, P. (1997) Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing* 14, 214-231.
- Urquhart, S. and C. Weir (1998) *Reading in a Second Language: Process, Product and Practice*. London: Longman.

- Wang, H. (2001) *Designing Short Answer Reading Comprehension Questions for College English Test in China*. Unpublished doctoral dissertation. Columbia University, N.Y.
- Weir, C. J. (1994) *Understanding and Developing Language Tests*. London: Prentice Hall Regents.