

CHAPTER 5

INSTRUMENTS USED: THEIR CREATION AND VALIDATION

In conducting the experimental portion of this thesis, it was necessary to assess the subjects' knowledge of gerunds and infinitives used with the four matrix verbs 'stop', 'remember', 'forget', and 'regret' both prior to and after receiving the experimental treatments. In addition, it was necessary to measure students' attitudes toward their experience with learning on computers. Toward the former purpose, two 40 item exams were created, validated, and shown to be approximately equal to one another. Toward the latter aim, a questionnaire was developed. In this chapter, the process of creating and evaluating the pre and post tests will be described, following which the questionnaire will be discussed.

5.1 Creating and Validating the Pre and Post Tests

5.1.1 Balancing Items Within Linguistic Domains

The first step in creating the pre and post tests was to isolate the linguistic domains to be tested. Since the principle taught in these lessons were reliably applicable in English only to the four matrix verbs used in the lessons, it was decided that all items on the pre and post tests would contain one of these four verbs in the matrix. This principle was, however, applicable to all possible complement

verbs in the English language. Of this set of all possible verbs, only the verbs 'talk', 'thank', 'introduce', and 'drink' were mentioned specifically in the problem section of the lessons. It was therefore felt that sixteen domains would have to be included in the pre and post tests, these being the product of combining each of the four matrix verbs with gerund and infinitive complements composed of (1) complements whose verbs appeared in the CALL lessons, and (2) complements whose verbs were not found in the lessons.

Potentially useful items were systematically developed for each of these categories. The researcher started with a chart which had eight headings: STOP TO DO, STOP DOING, FORGET TO DO, FORGET DOING, REMEMBER TO DO, etc. Under each heading, fourteen sample test items were listed. Four of these items contained one each of the complement verbs used in the computerized lessons, so that under the heading STOP TO DO, for example, there was an item containing 'stop to talk', another containing 'stop to thank', one with 'stop to introduce', and one with 'stop to drink'. This pattern was repeated for each of the other major headings. Then, under each heading, ten more items were developed containing complement verbs other than those that the students were going to encounter on the computer.

5.1.2 Adjusting Items Within the Domains

Following the procedure outlined above, fourteen sample

test items were listed under each of the eight major headings (i.e., headings for the four matrix verbs, each with separate listings for gerund and infinitive complements). These 14 items were in turn divided into items whose complements appeared in the lessons and those whose complements would not previously have been encountered by the students. The result was a list of 112 items, distributed throughout the 16 domains. These sample test items were then shown to four native speakers of English who offered suggestions for clarity and for reduction of ambiguity. Numerous changes were made in the sample items, and seven of the items were deleted outright as being unsuitable for foreign students (because they relied on idioms or presuppositions unfamiliar to foreigners), or because they were particularly subject to ambiguous interpretation.

Five of the deletions were in the REGRET TO DO category. Except in certain speech acts (e.g. "I regret to inform you that . . . "), the infinitive complement with regret is in practically free variation with the gerund when the action in the complement follows the act of regretting. For example, one of the test items, "We regret to leave so soon, but we really must be getting back," alternates with "We regret leaving so soon, " etc. Whereas a native speaker would never use an infinitive complement in a situation in which the action in the complement precedes that of regretting (as

in "I'll never regret buying gold last year"), such a speaker is not so constrained when the order of actions is reversed. Items were developed for the latter category anyway, on the theory that one might demonstrate a grasp of the paradigm involved by choosing the infinitive in cases where the action in the complement followed the action of regretting. As it proved difficult to find items in the category REGRET TO DO for which the order of actions was clearly distinct, many of these items were found to be unsuitable. There resulted an inequity in the final distribution of items throughout the 16 domains which reflected these considerations but which was not deemed to be detrimental to the study; on the contrary, it seemed desirable to reduce the number of items representing this particular gray area in an otherwise dependable paradigm.

5.1.3 Developing and Administering the Trial Exam

The next step was to draw up a trial exam containing the 105 intuitively valid items. This exam would then be tested on ESL students and the validity of each item scrutinized by item analysis. Items were randomized for the trial exam in the following manner: all 105 items were printed on separate slips of paper, and the slips drawn from a container. These were then printed on the trial exam in the order drawn.

The exam was administered to 44 students in two separate ELI 70 classes at the University of Hawaii. ELI 70 is the lowest level listening comprehension class offered to foreign students at the University of Hawaii. Students from these classes were chosen to help with the item analysis because, of all ESL students accessible to the researcher, students in these classes were nearest the level of the students who would subsequently be used in the actual experiment at Hawaii Loa College (and probably a little above; foreign students must score 450 or above on the TOEFL test to gain admission to UHM, whereas no such requirement exists at HLC).

The exams were given during regular class periods in the two different classes. No set time period was enforced for the exams, and students finished the 105 item tests in times ranging from fifteen to about 35 minutes. Students were free to review work on their papers until the last person was finished, but many chose to hand in their exams shortly after completion.

5.1.4 Pairing Items According to Item Analysis

After the exams were scored, difficulty and discrimination indices were computed for each of the 105 items. Discrimination indices were used in the analysis only to eliminate from the pool those items with indices of less than $-.20$ and to provide a means of fine measurement of

item parity in cases where more than one choice was possible from comparison of the primary means of parity, the difficulty indices.

Three items were eliminated outright due to grossly negative discrimination indices. The remaining 102 items were paired according to linguistic domain and degree of difficulty. The eighty items which lent themselves most readily to pairing (so that an item in one of the 16 linguistic domains could be matched with another item of similar domain and difficulty) were selected for use on the pre and post tests.

5.1.5 Assigning Items to Pre and Post Tests

Whenever possible, for items containing a given matrix verb, the author attempted to match complement verbs occurring in the CALL lessons with items containing identical complement verbs, but matches with items containing one of the other three complement verbs used in the computer lessons were also permitted (given the same matrix verb). In case a given complement verb not used in the computer lessons occurred in a second item with the same matrix verb, these recurrences were also paired (whenever practical).

Once similar items were paired as to difficulty and linguistic domain, items were assigned to the two different tests, designated Test A and Test B. This was done by coin flip, except where assignment of an item in one linguistic

domain to one test would force assignment of a similar item to the other test. This is better explained by example. 'Remember thanking' and 'remember introducing' were paired since they both had a difficulty index of .38 and both consisted of an identical matrix plus a complement appearing in the CALL lesson. By coin flip, 'remember thanking' was assigned to Test A. 'Stop thanking' and 'stop introducing', another pair each having a single matrix plus complements found in the computer lessons, and each with a difficulty index of .58, were therefore assigned to tests B and A respectively. In this way, each instrument received an item with the verb 'thanking' in complement position plus an item with a difficulty index nearly equal to the corresponding item with the same matrix verb, 'thanking'. This kind of chaining frequently dictated assignment of items to one or the other of the two tests, but all other assignments were made by coin toss.

Once items had been assigned to the two tests, they were randomized in the manner used for the 105-item test; that is, the items were printed each on a slip of paper, drawn out of a container, and placed on the exams in the order drawn. Finally, Test A was determined to be the pre test by coin toss, and Test B was therefore the post test by default. Copies of Tests A and B are presented in Appendixes A and B, respectively.

As can be seen from the Chart of Matrix/Complement Combinations Showing the Linguistic Content of Tests A and B (App. C), not all linguistic domains were covered on each of the two instruments. For example, domain #13 ('regret to' + complement verb appearing in the computer lessons) was not covered at all. In such cases, the choice was either to not have an item representative of a given domain or to pair two items which differed from each other by more than .10 in index of difficulty. It did not seem so necessary to have items representative of every single domain as it did to have items of similar indices of difficulty; therefore, minor omissions occurred in favor of the latter criterion. These omissions, where they did occur, compromised only negligibly the equivalence of the two instruments, if at all.

5.1.6 Assessing the Equivalence of the Pre and Post Tests

The indices of difficulty of all the items settled on for each of the two instruments were tallied. Mean difficulty for Test A was found to be 63.45 with a standard deviation of 17.31, and for Test B to be 64.48 with a standard deviation of 16.59. Comparing these two means by applying a t-test, the value of 0.27 is obtained (d.f. = 78; $p > .10$). As this is not a significant difference, the two tests were found to be essentially equal in difficulty. In addition, a control group of ten students who took the tests during the thesis experiment at Leeward Community College scored

essentially the same on both tests ($t = .322$; d.f. = 16; $p > .10$; see Table 6). Because the control group received no instruction pertaining to gerunds and infinitives between pre and post testing, there was every reason to expect that they should have done equally well on both tests. As this was the case, the contention that the two tests were equal in difficulty is further supported.

5.2 The Questionnaire

A questionnaire was formulated after observation of the students working at the computers at HLC. Thus, questions on the questionnaire were those that had occurred to the researcher while watching the subjects undergo treatment. These questions concerned the students' familiarity with computers prior to undergoing treatment, their attitudes toward their learning by computer, their understanding of the instructions and vocabulary (both in the CALL lessons and on the pre and post tests), and their perceptions of their degree of control over their own learning. There were in addition three questions on the questionnaire from which it was hoped to shed some light on the degree of retention of the grammatical items taught.

The questionnaire was formatted in three parts. The first part asked nine questions, each with either four or five answers that allowed respondents to register degrees of positive or negative feeling. For example, Question #2

asked what the students thought before they had worked the lesson. (This question was followed up by another asking what the students thought after they had worked the lesson.)

Possible answers were:

1. I did NOT want to use the computer.
2. I wanted to try the computer, but I was a little worried about it.
3. I didn't worry about using the computer, but I wasn't excited about it either.
4. I wanted to use the computer very much.

Other questions asked students' opinions about the time allowed to work the CALL lessons, about how they liked the game paddles, about their comprehension of the CALL lessons and instruments, and about what they thought they learned from the lessons.

Questions on the second part of the questionnaire (Question #10, a-g) covered similar areas as did the first, but instead of questions, students were presented with seven sets of three statements. Statements in each set were either positive, negative, or neutral on a given issue. For example, one of the these triads (Question #10a) was:

1. I wanted to work alone, but the experimenter kept telling me what to do.
2. I felt I was in control of my learning, and that I could work at my own speed.
3. Neither of the above applies to me.

The final part of the questionnaire (Question #11, a-c) simply consisted of three questions designed to gauge students' retention of the linguistic features taught. For example, Question #11a read: "Max stopped knocking on the door. Which did he do first? 1. stop? or knock?"

A copy of the complete questionnaire is given in Appendix D.