

GRE

GRADUATE RECORD EXAMINATIONS

PROCEDURAL MANUAL FOR CONDUCTING CONTENT

REPRESENTATIVENESS STUDIES

Philip K. Oltman

Supplement to
GRE Board Professional Report GREB No. 81-12P

February 1983



EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

Procedural Manual for Conducting Content
Representativeness Studies

Philip K. Oltman

Supplement to
GRE Board Professional Report GREB No. 81-12P

February 1983

Sample

The sample for a given discipline can be drawn from the data base contained in the Higher Education General Information Survey tapes (HEGIS). These tapes are available in the Measurement, Statistics, and Data Analysis Research Division (contact Norma Norris). The program for generating the sample is on file in that division and is described in Appendix F.

The HEGIS survey divides the United States into eight geographical regions, and the program selects institutions to represent each region approximately equally. For each region, the tape is searched for all institutions offering a degree in the particular discipline of interest. Within that list for each discipline, institutions are selected randomly, subject to the following constraints:

1. Approximately 50% public and 50% private control.
2. Approximately 75% with more than 10,000 total enrollment and 25% with between 1,000 and 9,999 enrollment.
3. Approximately 50% offering a doctoral degree in the discipline and 50% offering up to but not including a doctoral degree.

These criteria for selection can be easily modified if desired. In the study described in GRE Report No. 81-12P, we aimed for 8 to 10 institutions per region, and since not every region contained each type of institution, the percentages listed above are approximate.

The resulting sample of institutions is randomly selected for each discipline, subject only to the above stated constraints. While it is not a strictly stratified sample, it does appear to yield a generally representative group of institutions and does not depend upon human judgment other than in setting the constraints. A more detailed description of the procedure is presented in Appendix F. The samples used in the present study are presented in Appendix A.

Chair Letters

Using the list of institutions generated from the HEGIS tapes, a letter inviting participation in the study is addressed to the chairperson of the department at each. In GRE Report No. 81-12P, the addresses were from college directories. The text of the chair letter is presented in Appendix B, and is available on Vydec disk. Addressing the letters and envelopes can be done by Vydec, or by means of copying machine or computer-generated labels.

The letters themselves could be Vydec originals, or printed on ETS letterhead, as was done in GRE 81-12P. When more than one discipline is being studied concurrently, it is advisable to have a code appear in one corner of each page, so that the material can be quickly and accurately identified without searching the text each time.

Telephone follow-up of each letter is an option and would probably increase response rates. It is costly, however, both in toll charges and in personnel time, since chairpersons are almost never at their desks. Several calls and messages are usually necessary to speak to the chairperson. In GRE 81-12P, we initially elected to call when we received no response to the departmental mailing and to a follow-up mailing. These calls did not seem to be worth the added expense, however, and were not continued after the first several dozen. We have concluded that response rates can be adequate without telephone follow-up, but this decision can be made by the study director.

Respondent Letters

The text of the respondent letter is presented in Appendix C. This letter is mailed to each individual designated by the department chair to participate in the study. We requested two names per department, and were usually furnished that number, although in a few cases three or more faculty members agreed to participate.

In GRE 81-12P, the respondent letters were preprinted on ETS letter-head paper. Each name and address was typed on a bond paper sheet and used to generate labels using copying machine label sheets. We produced several identical sets of labels so that we could address the respondent letter, affix another label to the Test Files Order Form (see below) and another to the test itself. In this way all materials received from a respondent were clearly labelled. Each label had a code in the lower right-hand corner that identified the discipline and the geographical region, to facilitate sorting and filing (e.g., the label for a chemistry faculty member in the third region would be coded C3 in the lower-right corner).

Materials Mailed to Respondents

The set of materials mailed to each of the three disciplines are shown in Appendix C. They consist of the following items:

Test Files Order Form. All mailing and receipt of materials is done through Test Files. Test Files is supplied with the Order Form and a set of materials for respondents. They then add a copy of the test form being examined and mail the material to the respondents by certified mail with a return receipt requested. The standard procedures of Test Files are followed, with the exception that the study materials are inserted in the envelope along with the test form. Respondents mail materials back to Test Files, where they are picked up for processing.

The Test Files Order Forms are preprinted with a code in the upper-left corner to indicate the discipline, along with additional information including the program, name of test, form, P/J, etc. Test Files personnel affix a respondent label to each test as it is sent out. The Test Files Order Form is followed by:

1. Respondent Letter. This letter is on Vydec disk. A respondent label is attached to each letter.
2. Background Information Questionnaire. This form is on Vydec disk and can easily be edited to meet particular requirements. It asks respondents to supply several items of descriptive information so the sample can be described more fully.
3. Introduction to the Specifications. Describes the specifications and requests respondents to review them.
4. Detailed Specifications. The level of detail of the working specifications varies with the discipline. The fineness of detail of the outline presented to respondents should be decided with the advice of the ETS test specialist. We included finer levels of detail than were allocated by respondents as a way of explicating the meaning of the higher level categories.
5. Introduction to the Specification Distribution. This introduction requests respondents to recommend how much of the test should fall into the main content categories of the specifications, both to represent their current curriculum and an ideal graduate-school-preparation curriculum. The main categories should number no more than about eight for this task. If this high a level of generality would be uninformative, additional judgments can be obtained (see chemistry, below).
6. Specification Distribution Sheet. The layout should lend itself to encoding the data for later analysis.
7. Special Specification Distribution. A more detailed specification distribution judgment may be included when the main categories are not sufficiently informative. For example, in GRE 81-12P, the main categories dividing the domain of chemistry were too general to be of much value. Therefore, respondents were asked to distribute percentages of coverage within their field of specialization. Fields were identified in advance from a professional directory. Percentages were assigned both at the subheading level (A, B, C...) and at the topic level within subheadings (1, 2, 3...). See Appendix C for more detail. A page of instructions for this task

was included; respondents indicated their percentage assignments on the specifications themselves. (Specifications should be printed so this is possible.)

8. Introduction to Item Classification. Respondents are asked to classify a subset of the test items into the categories of the specifications. Classification of an item into more than one category is allowed. In GRE 81-12P, respondents were asked to classify no more than 50 items. For disciplines for which main category classification was satisfactory, the tests were spiraled so that each succeeding respondent classified a different subset of items. The tests were premarked with labels to indicate which set of items should be classified.

In GRE 81-12P, the item classification task for chemistry was done within fields of specialization, rather than by arbitrary subsets of items. Each respondent received a test with item numbers circled in red to correspond to his or her specialty in chemistry. The respondent then classified the items by subheading (A, B, C...) and, within subheading, by topic (1, 2, 3...).

9. Test Form. The form to be used should be selected by the test specialist. In GRE 81-12P, it was a current form (3DGR).

Mailing Procedure

Test Files should be contacted well in advance of the study so that a sufficient supply of the form to be used can be on hand. These forms should be premarked with the item set to be classified and kept in a separate storage area in Test Files. If tests are to be premarked in specialty areas, as was done for chemistry, it may be more convenient for the supply of tests to be kept by the test specialist and items marked as needed for mailing.

Letters to the department chairpersons are mailed out next, with preaddressed postage-paid return envelopes enclosed. A return form may also be enclosed for the chairperson to indicate names of participants or to indicate that they decline to participate. Such a form might increase response rate, although it was not used in the present study. Return mailings from departments should be logged in on a master roster of departments organized by region. Such an organization by region facilitates location of an institution and highlights regions that are slow to respond.

As returns from departments arrive, the names and mailing addresses of respondents are typed onto respondent rosters. For the present study, these were typed using a template to achieve proper spacing for running

off copying machine labels. As pages of the roster were completed (21 names and addresses per page in this study), the labels were run off and sets of materials were assembled.

Items 1 through 8 on the above list of materials mailed to respondents were stapled together (i.e., all material except the Test Files Order Form and the test itself; see Figure 1). If possible, these materials could all be printed, collated, and stapled by the Printing Department rather than being assembled later.

After assembling the materials and affixing the respondent labels to the Test Files Order Form and the respondent letter, the Order Form and materials packet are clipped together. The set of material is then taken to Test Files, along with a sheet of labels for Test Files personnel to affix to the tests. Mailing and receipt is then handled by Test Files.

As respondents return their materials to Test Files, the materials can be picked up and brought to the project office for coding.

A secure storage cabinet should be available for maintaining the files, since they include the current test forms. When all information has been coded from the test forms, they can be sent to Confidential Scrap for shredding.

Coding the Data for Analysis

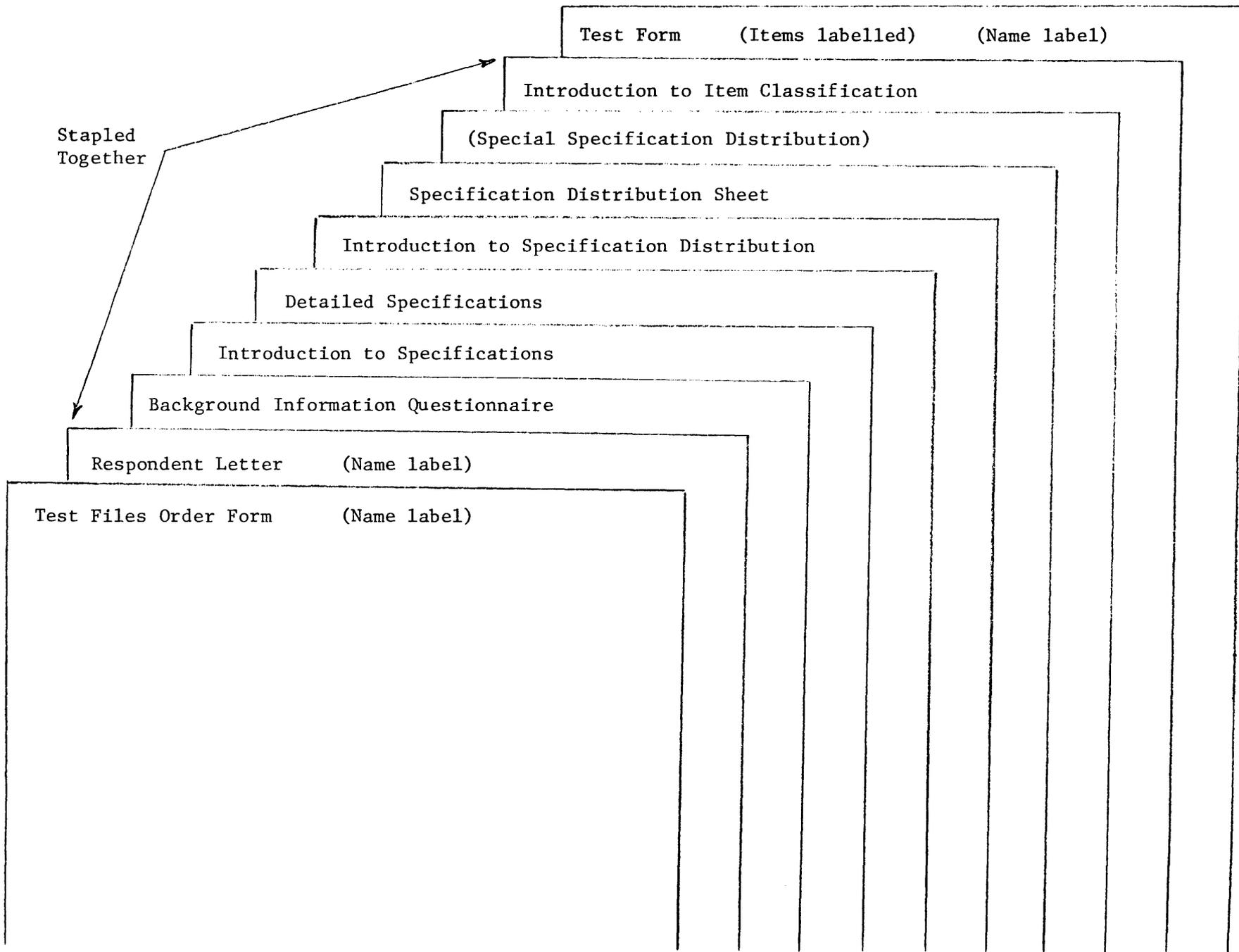
The data should be coded from the materials into keypunchable format. Even if it is decided to do the analyses by hand, this type of coding is neat, orderly, and convenient. If hand analysis proves too time-consuming, cards can readily be punched and the analyses run on the computer. Standard statistical packages (BMDP, SPSS) can readily compute most of what is needed.

1. Respondent Identification. In GRE 81-12P, respondents were identified by a code consisting of the first letter(s) of the discipline name, the number corresponding to the geographical region, and the first five letters of the respondent's last name (e.g., a chemist from the fourth region named Gordon would be coded C4GORDO).

2. Institution Characteristics. Public vs. private, up to doctorate vs. doctorate, and small vs. large can be coded 1 or 2 respectively. This information is printed for each institution in the output of the sampling program.

3. Item Set. A code is entered to indicate the set of items the respondent classified (e.g., for a 200-item test, with sets of 50 items, codes could be 1, 2, 3, and 4).

4. Background Information Questionnaire. In GRE 81-12P, the following items were coded as indicated:



Assembly of Materials

Figure 1

Academic rank: 1 = Professor, 2 = Associate Professor,
3 = Assistant Professor, 4 = other

Administrative responsibilities: 1 = any administrative
responsibilities, 2 = none indicated

Years teaching: The number given

Year of highest degree: The last two digits of the year given

Average teaching load: The hours given

Number of doctoral dissertations directed: The number given

Number of master's theses directed: The number given

The indicated areas of specialization were not coded for quantitative analysis. They can be recorded and categorized to help describe the sample.

5. Specification Distribution Sheet. The percentages assigned are coded in each category. Leave sufficient room on the data sheet to record 100 in the unlikely event that a respondent assigns all content to one category.

Check that percentages assigned add to 100. If they do not, usually because of a computational error by the respondent, then add the entries, and divide each by the sum to convert to a 0 to 100 scale. Then record the converted values.

(5a). Special Specification Distribution. If this detailed breakdown is included, it is coded at two levels. First, the A, B, C... levels are recorded; the entries should total 100. Then, within each of those subheadings, the topics labelled 1, 2, 3... are recorded; those should each total 100 within subheadings.

6. Item Classification. These data are taken from the test form, on which respondents have indicated beside each item the specification category or categories to which they judged it to belong. The classification(s) of each item is entered on an intermediate form with columns labelled with the Roman numerals (or subheadings and topics for detailed classification) corresponding to each category. After these are entered, they are totalled for each category and converted to percentages by adding the total number of classifications and dividing each category entry by that total. Thus, if an item were classified into two different categories, it would be counted once in each category. If a respondent sorted 50 items, using only single categorization, the total of the entries would be 50; if multiple categorizations were used, the total of the entries would exceed 50. In any case, percentages are arrived at by dividing the entries in each category by the total number of entries.

Note that the above method will reflect the distribution of categorizations made by the respondent but will not preserve the categorizations

of individual items. If agreement on individual items is to be assessed, an items vs. categories matrix must be constructed for each respondent. In GRE 81-12P, individual item agreements were not coded, since the focus was on the distribution of content that each respondent saw in the items.

Appendix F

Memorandum for: PHILIP K. OLTMAN

cc: John Barone
Norma Norris

Subject: Documentation for
Hegis Sampling Procedure

Date: September 11, 1981

From: David Freund

The five HEGIS tapes have been copied to standard ETS T-tapes and are cataloged as follows:

NAN6600.HEGIS.OFE11.Y7677 - T33524
NAN6600.HEGIS.ERD12.Y7677 - T37469
NAN6600.HEGIS.OFE13.Y7879 - T29867
NAN6600.HEGIS.ERD14.Y7879 - T37231
NAN6600.HEGIS.ACE. Y7879 - T32240

CREATION OF HEGIS SAMPLE:

There were five major steps and several miscellaneous ones performed in creating the sample. The programs are cataloged under:DSF6600.HEGIS.SAMPLE.FORT.

1. CRDOMAIN - The first step done was copying the records of the disciplines (Cols. 87-89) from which samples were to be drawn to disk, using TAPE.T37231. By copying the records to disk, only one pass through the approximately 160,000 records was necessary.
2. SORTDOM - This program assigns a number from 1 to 64 (writes to Cols. 176, 178) for each discipline based on the following strata:

8 geographical regions (Col. 24, OBE region) of the U.S. (yield 8-10 schools in each).

Within each region:

- A) 75% of the schools offer a doctoral program and 25% offer a bachelor's degree up to a doctoral degree (Cols. 32, 33).
- B) 50% of the schools are large (>10,000 students) and 50% are small (1,000 - 10,000) (Col. 39).
- C) 50% are public and 50% are private (Col. 26).

It was determined the best way to fit the strata was as follows:

For any region j , $j=1,\dots,8$:

	Small		Large		
	Public	Private	Public	Private	
Doctoral	2	1	Doctoral	2	1
Nondoctoral	1	1	Nondoctoral	1	1

3. SAMPACTS - Computes the total N for each cell of the domain.

4. SAMPLE - actually creates the sample. The programming technique is quite simple; for each cell there exists an N as computed in step #3. The program assigns n random integers from 1 to N ; n is the desired sample size for a given cell as illustrated above. A counter then matches the random counts with the actual record counts. When a match occurs, the record is put into the sample.

5. LISTSAMP - lists the sample broken down into the various strata.

Note: In some instances, it was not possible to get a sample size of 8-10, because some of the cells within a region had no schools that met the particular criteria for that cell.

Attached is a copy of the program.

FORTRAN IV G1 RELEASE 2.0

MAIN

DATE = 81245

08/10/45

```

0001      REAL*8 REGION(8)/'REGION 1','REGION 2','REGION 3','REGION 4',
+REGION 5','REGION 6','REGION 7','REGION 8'/,
+SCHTYP(2)/'PUBLIC ','PRIVATE '/, HIGHDF(2)/'BAC-PH.D',
+PH.D '/, ENROLL(2)/'K - 10K','> 10,000'/,
+DISP(3,3)/'GENERAL ','CHEMISTR','Y ','GENERAL ','COMPUTER',
+SCIENCE','EDUCATIO','N ','/'/, SCHOOL(4)/4*' '/,
+HED(8)/'DISCIPLI','NE IS: ','3*' ',' - ',' '*,'*'/
0002      REAL REC(45)
0003      CALL START('HEGIS STUDY: RANDOM SAMPLE OF SCHOOLS THAT FIT DEFINED
+ STRATA*')
0004      CALL OPEN(10)
0005      DO 10 J=1,3
0006          JREG=1
0007          CALL MOVSTR(24,1,DISP(1,J),16,HED)
0008          DO 20 K=1,8
0009              IOLD=JREG
0010              CALL MOVSTR(8,1,REGION(K),45,HED)
0011              IFIRST=1
0012              CALL NEWPG(HED)
0013              IF (J.EQ.1 .AND. K.EQ.1 )GO TO 5
0014              GO TO 50
0015          5      CALL GET(10,REC,8999)
0016              CALL MOVSTR(24,46,REC,1,SCHOOL)
0017              JREG=ICOL(24)
0018              JPUB=ICOL(26)
0019              JDEG=1
0020              IF (ICOLS(32,33).EQ.8) JDEG=2
0021              JK=1
0022              IF (ICOL(39).GT.6) JK=2
0023              IF (IOLD.NE.JREG) GO TO 20
0024              IOLD=JREG
0025          50      IF (IFIRST.EQ.1) WRITE(6,6000)
+                SCHOOL,SCHTYP(JPUB),HIGHDF(JDEG),ENROLL(JK)
0026          +      IF (IFIRST.NE.1) WRITE(6,7000)
+                SCHOOL,SCHTYP(JPUB),HIGHDF(JDEG),ENROLL(JK)
0027          IFIRST=2
0028          GO TO 5
0029          20      CONTINUE
0030          10      CONTINUE
0031          6000  FORMAT(/,/,1X,4A8,'- ',A8,' - ',A8,' - ',A8)
0032          7000  FORMAT(1X,4A8,'- ',A8,' - ',A8,' - ',A8)
0033          999  CALL STEXIT
0034          CALL EXIT
0035          END

```