ACT RESEARCH REPORT

No. 65

65

July 1974

CONSIDERATIONS AND PROCEDURES IN NATIONAL NORMING: AN ILLUSTRATION USING THE ACT ASSESSMENT OF CAREER DEVELOPMENT AND ACT CAREER PLANNING PROGRAM, GRADES 8-11

D. L. Bayless
J. W. Bergsten
L. H. Lewis
R. J. Noeth

PUBLISHED BY THE RESEARCH AND DEVELOPMENT DIVISION

THE AMERICAN COLLEGE TESTING PROGRAM

P. O. BOX 168, IOWA CITY, IOWA 52240

CONSIDERATIONS AND PROCEDURES IN NATIONAL NORMING: AN ILLUSTRATION USING THE ACT ASSESSMENT OF CAREER DEVELOPMENT AND ACT CAREER PLANNING PROGRAM, GRADES 8-11

TABLE OF CONTENTS

ABSTRACT

Raw scores on most standardized educational and psychological assessment instruments acquire meaning only when referenced to a set of norms. Test publishers should clearly describe their norming procedures, including the target population and the sample on which the norms are based. The primary purpose of this report is to illustrate some of the major considerations in obtaining normative data, procedures that can be used, and problems encountered in norming studies. Norming procedures used in two of ACT's assessment programs for career guidance, the Assessment of Career Development and the Career Planning Program, Grades 8-11, are described for purposes of illustration. Norms for these programs are based on a three-stage nationwide probability sample of 32,000 8th, 9th, and 11th graders in 197 schools. This report also documents the nature of the sample and the sampling procedures used in ACT's *Nationwide Study of Student Career Development.*

CONSIDERATIONS AND PROCEDURES IN NATIONAL NORMING: AN ILLUSTRATION USING THE ACT ASSESSMENT OF CAREER DEVELOPMENT AND ACT CAREER PLANNING PROGRAM, GRADES 8-11

David L. Bayless
Jane W. Bergsten
Louise H. Lewis
Richard J. Noeth[1]

I. The Need for Norms

The raw scores on most standardized educational and psychological assessment instruments have little, if any, meaning either to the individuals who have taken these measures or to the educators who must interpret them. These raw scores acquire meaning only when they are referenced to a set of norms. For example, reference to norms may show that a student's raw score of 23 on a mechanical reasoning test falls at the 93rd percentile among a nationwide sample of 9th grade boys, or that the mean score of a school's 10th graders on a citizenship exam ranks 103rd from the top among 167 high schools in a state. In both of these cases, test performance is described in terms of relative standing in a norm group (i.e., a nationwide sample of 9th grade boys, 10th graders in a state's high schools).

These normative standings must then be interpreted by professional educators familiar with the individual student and/or school. The evaluation of the test standings (e.g., the implication of ranking at the 93rd percentile) must take into account the circumstances and conditions relevant to the individual or group. Only then can informed decisions be made. Thus, norms may facilitate the interpretation of test scores, but it is the users of tests who must determine the implications of a normative standing.

The characteristics of norming samples obtained by test publishers should be described in test manuals and handbooks. Furthermore, the *population* which the normative sample represents must be clearly defined and must constitute a group with whom users of the test will ordinarily wish to compare the persons tested (APA, 1966). For educational and psychological assessment instruments intended for use with students or schools in general, norms based on a representative sample of a nationwide target population are usually required.

---

[1]David Bayless is a Senior Educational Sampling Statistician, Jane Bergsten is a Senior Statistician, and Louise Lewis is a Junior Statistician, all at Research Triangle Institute. Richard Noeth is a Research Psychologist in the Research and Development Division of ACT. The authors would like to thank Ralph Folsom, Senior Statistician at RTI, and Leo Munday, Vice President of the Research and Development Division of ACT, for their helpful comments on an earlier draft of the manuscript. We are particularly grateful to Dale Prediger, Director of the Developmental Research Department at ACT, for his continued assistance and suggestions as well as his thorough reviews of earlier drafts of this report.

Purpose of Study

The primary purpose of this report is to illustrate some of the major considerations in obtaining normative data, procedures that can be used, and problems encountered in norming studies. Norming procedures followed in the recent development of two of ACT's assessment programs for career guidance, the Assessment of Career Development (ACD) and the Career Planning Program, Grades 8-11 (CPP 8-11), are described for purposes of illustration. This report also discusses the nature of the sample and the sampling procedures used in ACT's *Nationwide Study of Student Career Development* (Prediger, Roth, & Noeth, 1973).

Instruments Used in Illustration

The ACD (ACT, 1974a) was developed for use with students in grades 8 through 11 as part of a school's developmental career guidance program. Its primary purpose is to help counselors and administrators obtain information needed to develop effective guidance programs tailored to student needs, and to assess the outcomes of career guidance programs initiated through special projects and studies. Group summaries are provided for 11 ACD scales as well as for responses to 42 individual items. List reports containing each student's scale scores, career plans, self-evaluations of career planning, and requests for help are also provided. Concurrent with the norming of the ACD and CPP 8-11, ACT conducted a nationwide study of student career development (Prediger, Roth, & Noeth, 1973) to assess and summarize core aspects of the career development of our nation's youth. The data for this study were based on scales and items in the ACD.

The CPP 8-11 (ACT, 1974b) was developed to assist junior and senior high school students with self/career exploration, career planning, and career-related decision making. This program provides a broad base of interrelated information about a student's personal characteristics, including scores for six interest, six ability, and six experience scales. Self-reported information on career plans (including occupational preference), job values, working condition preferences, and career-related abilities is also provided. A student's interests, abilities, and occupational preferences are used to suggest "job families" for exploration.

The results for the ACD and CPP 8-11 were intended for interpretation mainly through their comparison to a reference or norm group (although the ACD does lend itself to criterion-referenced interpretation). Because of the nature of the variables assessed by these instruments and their intended application in schools throughout the nation, it was important to select a nationwide, representative sample of junior and senior high school students as the norm group. Furthermore, to insure that this nationwide norm group (normative *sample*) would provide unbiased *population* estimates of student scores and responses for ACD and CPP 8-11 scales and items, probability sampling procedures were used.

Probability Samples

Norms based on national samples are appropriate for most tests used in education and guidance. However, norms based on national samples which are not scientifically designed may be biased. In addition such norms may contain large amounts of random error that cannot be estimated statistically. Angoff (1971) cites a number of sampling procedures that can produce a poor set of norms. Included are samples of convenience (e.g., students attending schools that have purchased and used a test prior to the development of national norms); a sample which is based on outdated or incomplete information for the target population from which the sample is to be drawn; and a "representative-area" sample in which groups of individuals in each of several locations are judged to be "typical" and thus selected as the "representative" sample.

ACT's goal, however, in norming the ACD and CPP 8-11 was to obtain a truly representative nationwide sample and hence, to avoid these types of sampling procedures. A probability sample was selected to achieve this goal. In a probability sample, each individual (or primary element) in the sample has some known probability of being selected into the sample. Each step in the selection of individuals or elements into the sample is defined, and weights derived from the probabilities are used in the estimation of population parameters (Cornell, 1960). Probability sampling also allows for the objective evaluation of sampling error (Angoff, 1971).

In order to insure that a representative and scientifically drawn sample would be obtained, ACT contracted with Research Triangle Institute (RTI) of North Carolina to design the normative sample for the ACD and CPP 8-11. RTI has considerable experience and expertise in designing nationwide samples; for example, RTI selected the sample used

in the National Assessment of Educational Progress (NAEP). Because of this experience, RTI had data describing crucial characteristics of the target population for the ACD/CPP 8-11 norming study (e.g., identity of all schools in the nation, size of school, size of community, socioeconomic status of community).

Generally test publishers have neither the capability nor the necessary information about the target population to scientifically select samples for the norming of tests. Contracting for professional sampling services is one way to insure representative normative data for educational and psychological assessment instruments.

II. The Sample Design: Considerations Illustrated

The general principles to be considered in designing a nationwide probability sample for guidance assessment instruments are described below. Also discussed are the general sample design requirements of the ACD/CPP 8-11 norming study. Detailed procedures used in sample selection and weighting are covered in Parts III and IV.

General Considerations

For reasons noted above, only probability sampling procedures were considered in designing the ACD/CPP 8-11 norming study. Stated briefly, a probability sample is one in which each student in the defined or target population has a chance of being selected into the sample. Furthermore, in selecting a probability sample, stratification (the partitioning of a population into homogeneous groups or strata and sampling independently from each) generally can be utilized to increase the precision of the population estimates based on the sample data. For measurement scales similar to those on the ACD and CPP 8-11, the ideal way to stratify the target population is to place schools of like normative distributions for these scales in the same group or stratum and then to select a sample of schools from each homogeneous group. However, since data relevant to the normative distributions of each of the ACD/CPP 8-11 scales were not available for the target population, other criteria were used for stratification.

The general considerations for the ACD/CPP 8-11 norming study naturally centered around the requirements established for the design of the sample. The major requirements were as follows:

1. The target population of students would consist of all full-time 8th, 9th, and 11th grade students enrolled in public or Catholic schools in the United States in the spring of 1973. (The target population originally included grade 10 as well. The sample for grade 10 was actually selected but could not be used because of unexpected costs which resulted from both an underestimate of per grade enrollment and the decision to allow schools to test *all* students in a grade, rather than the required subsample, if they so desired.)

2. Each student in the sample would respond to *all* ACD and CPP 8-11 instrument items. Hence matrix sampling (Wilks, 1962) was not possible.

3. The sample of schools and/or students should be representative of the target population with respect to the following variables:
   a. Geographical divisions or regions of the United States
   b. Areas of residence, referred to in this report as size of community (SOC)
   c. Socioeconomic status (SES).
   These three variables were used as stratification variables in the sample design.

4. Normative distributions for each ACD/CPP 8-11 scale were to be estimated for males and females in grades 8, 9, and 11. Thus, the six grade-by-sex subpopulations which are the reporting groups should be adequately represented in the sample in order to provide normative distributions with the desired statistical precision.

5. The desired statistical precision for each grade-by-sex normative distribution was as follows: The 95% confidence interval for the scale value corresponding to any percentile rank, P, which exceeded 74, should cover at most 6 percentile rank units. That is, if repeated samples were selected according to the ACD/CPP 8-11 norming study sample design, for 95% of the samples, the scale values corresponding to $P \pm 3$ (when P exceeds

74) should span the population scale value. Appendix A describes the technical details of this requirement from a statistical viewpoint. This precision requirement was determined by the manner in which scores were to be reported to students, i.e., in stanine bands. With the precision requirement specified above, a 3-stanine band would be highly likely to take into account *both* sampling error and measurement error when it is used to represent standing in the national population.

Overview of Sample Design

The target population of schools for the ACD/CPP 8-11 norming study included all public and Catholic schools in the United States having enrollment in grades 8, 9, or 11 in the spring of 1973.

A two-stage probability sample of schools was selected. Initially, primary sampling units (PSUs), each consisting of the geographical area making up a county or a group of counties, were selected. Then, within each sample PSU, three samples of schools, one sample each for grades 8, 9, and 11, were selected. Since each selected school was given the option of testing all students at the given grade level or of subsampling 60 students in that grade, the final sample of students became a three-stage sample design. Thus, instead of subsampling 60 students, the number required by the sample design, a few schools tested as many as 700 students in a grade. The figure of 60 students was arrived at after investigating the effect that different cluster sizes had on cost, administrative feasibility, and sampling error. The special option of subsampling students is discussed in Part IV.

The first-stage sampling frame of PSUs consisted of a list of counties or groups of contiguous counties stratified by region, size of community (SOC), and socioeconomic status (SES). Each separate entry in the list, whether one county or a group of contiguous counties, is called a primary sampling unit (PSU). The list of PSUs was identical to the first-stage or primary sampling frame for the NAEP sample for the 1971-72 school year. A NAEP tape file containing these PSUs was the source for the first-stage sampling frame. The PSU sample selection procedure involved selecting the primary sampling units without replacement, with probability strictly proportional to a PSU enrollment size measure (Moore, Chromy, & Rogers, 1974), using a rejective sampling method (Sampford, 1967). The number of 17-year-olds in a PSU was used as the measure of size. The procedure used is discussed in detail in Appendix B.

For each sample PSU selected, a second-stage sampling frame, i.e., a school sampling frame, was constructed. The frame included all public and Catholic schools with grades 8, 9, or 11. Three independent samples of schools, one for each of the three grade levels, were selected with probability proportional to size, i.e., to grade enrollment. The sampling procedures used to select the schools for each of the grade level samples in each sample PSU are discussed in Appendix C. To meet the required statistical precision, a total sample size of 107 schools per grade was needed. The technical justification for this sample size is presented in Appendix A.

In addition to the selection of a "basic sample" of 107 schools per grade level, a "back-up sample" of 107 schools, one for each "basic sample" school, was selected from within the sample PSUs. In the event that one of the "basic sample" schools could not participate in the norming study, the "back-up sample" school was used as a replacement.

Stratification Variables

The requirements that the sample of students be representative of the target population with respect to geographical regions of the United States, size of community (SOC), and socioeconomic status (SES) were met through stratification. Therefore, the first-stage sampling frame was composed of PSUs stratified by region, SOC, and SES.

Four geographical divisions of the nation were used for regional strata. These strata and the states located within each are presented in Table 1.

The SOC strata were defined as follows:

SOC 1 Big Cities—All counties containing a central city whose population is 200,000 or greater.

SOC 2 Fringes around Big Cities—The remaining counties within the same Standard Metropolitan Statistical Area (SMSA) as the Big City.

SOC 3 Medium Sized Cities—All SMSA counties not included in SOC 1 or 2 plus all counties containing at least one city with a population of 25,000 or more. If such a city is located in more than one county, the county containing a majority of the city population was classified in this category.

SOC 4 All counties and combinations of counties not included in the other three SOC strata.

TABLE 1

States in the Four Geographical Regions Used in the Sample Design

Region 1 Northeast	Region 2 Southeast	Region 3 Central	Region 4 West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
			Utah
			Washington
			Wyoming

The criteria for defining region and SOC are identical to those used in the NAEP sample design for the 1971-72 school year.

Four large PSUs in SOC 1 were considered separately because of their size. Los Angeles County (Los Angeles, California), Cook County (Chicago, Illinois), Wayne County (Detroit, Michigan), and metropolitan New York City each constituted a separate "self-representing" SOC substratum in its respective region. These substrata, designated "SOC 1SR," were "self-representing" in that each of these four large PSUs would occur in the first-stage probability sample with certainty, i.e., with probability 1. The remaining PSUs in SOC 1 were used to construct the SOC 1 substrata that were sampled. These latter sampled substrata were designated "SOC 1S." Table 2 shows the number of PSUs in the primary sample frame contained in each of the region by SOC strata.

The SES index used to construct substrata within the region by SOC cells of Table 2 was based on 1970 county census data such as housing value, rent, and the proportion of housing units lacking some plumbing facilities. These county data were entered into a multiple regression equation to estimate the proportion of families in the PSU with incomes less than $3,000, the SES index. When the ACD/CPP 8-11 norming sample was selected, income data from the 1970 census were not available. Therefore, regression coefficients were obtained using 1960 census data. These regression coefficients were then used with the 1970 housing variables to estimate the 1970 SES index. The PSUs within a region by SOC cell were ranked on the basis of this SES index and the SES substrata were formed by grouping the PSUs. The number of PSUs in the primary sample frame distributed by region, SOC, and SES substrata are presented in Table 3.

The difference between stratification variables and reporting variables is an important consideration in the norming of guidance assessment instruments. While stratification variables are used to subdivide the population into homogeneous groupings or strata prior to the actual sample selection, reporting variables are used in the analysis plan of the survey. Therefore, it is possible for a variable to be used both as a stratification variable for sample design purposes and as a reporting variable for analysis purposes. For example, schools in the population could be classified according to selected grades for stratification purposes and then these same grades could be used

TABLE 2

Number of PSUs by Region and Size of Community (SOC) Strata

Region	Size of community					Total
	1SR[a]	*1S*[b]	*2*	*3*	*4*	
Northeast	1	7	34	57	67	166
Southeast	N[c]	6	23	110	241	380
Central	2	10	34	113	224	383
West	1	14	29	82	150	276
Total	4	37	120	362	682	1,205

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no PSUs were allocated to this substratum.

TABLE 3

Number of PSUs by Region, SOC, and SES Substrata

Region	Size of community													Total
			2 SES substrata		*3* SES substrata				*4* SES substrata					
	1SR[a]	1S[b]	1	2	1	2	3	4	1	2	3	4	5	
Northeast	1	7	17	17	17	18	22	N	67	N	N	N	N	166
Southeast	N[c]	6	23	N	33	24	20	33	68	44	44	41	44	380
Central	2	10	34	N	30	30	53	N	86	69	69	N	N	383
West	1	14	29	N	17	29	36	N	90	60	N	N	N	276
Total	4	37	103	17	97	101	131	33	311	173	113	41	44	1,205

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no PSUs were allocated to these substrata.

for reporting normative distributions. This was done in the ACD/CPP 8-11 norming study using grades 8, 9, and 11.

It is also possible for a variable to be used as a stratification variable but not as a reporting variable, and vice versa. For example, it was decided that ACD/CPP 8-11 normative distributions would not be reported for regions of the U.S. even though the population was stratified by region prior to sample selection. Region was thus used as a stratification variable in the sample design and not as a reporting variable. Independent of the statistical gains derived, stratification of the schools by region before sampling serves as a guarantee that the sample will be geographically spread across the nation and will have "representative credibility." An example of a reporting variable not used as a stratification variable for the ACD/CPP 8-11 norming study

sample design is the sex of the student. While none of the schools of the sampled population had been asked to distinguish, for stratification purposes, their numbers of male or female students prior to sample selection, ACD/CPP 8-11 norming study results have been reported by sex.

III. Allocation and Selection of Sample

Number of Schools Required

The first step in allocating the sample to the various strata was to determine the number of schools to be selected from each stratum. Using this allocation of schools, the number of PSUs to be selected was then determined. To meet the planned statistical precision of the sample, it was estimated (see Appendix A) that each grade level sample should contain 104 schools. It was further decided that an equal allocation of the schools to the four regions should be used. Thus, 26 sample schools per grade level were to be selected from each of the four regions. The allocation of sample schools to the SOC strata within each region was roughly proportional to the size of the 17-year-old population of the strata. Table 4 shows the number of 17-year-olds by region and SOC strata. The unadjusted and unrounded number of sample schools per grade level to be allocated to each region and SOC stratum was calculated from Table 4 by dividing the number of 17-year-olds for each SOC stratum by the region total and multiplying the results by 26. The unrounded allocation is presented in Table 5. The number of schools per grade level, shown in Table 5, was rounded to the nearest integer and is given in Table 6. The school sample size per grade level increased from 104 to 106 as two additional schools were added to the sample when the unrounded proportional allocation of 104 schools was rounded.

An analysis of the "between PSUs" and "between schools within PSUs" variance-component estimates using NAEP data revealed greater variation in achievement within SOC 1 and 2 PSUs and less variation within SOC 3 and 4 PSUs. It was therefore decided to select *two* schools per grade level from each PSU sampled from both SOC 1 and SOC 2 strata and *one* school per grade per PSU in SOC 3

TABLE 4

Number of 17-Year-Olds by Region and SOC Strata

Region	Size of community					Total
	1SR[a]	*1S*[b]	*2*	*3*	*4*	
Northeast	122,746	143,717	391,376	268,028	119,683	1,045,550
Southeast	N[c]	73,601	164,237	295,469	378,143	911,450
Central	159,461	167,134	259,703	268,870	310,847	1,166,015
West	129,971	261,599	200,049	222,597	189,441	1,003,657
Total	412,178	646,051	1,015,365	1,054,964	998,114	4,126,672

Note.—Information taken from RTI National Assessment of Educational Progress data bank.

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no 17-year-olds were allocated to this substratum.

TABLE 5

Proportional Allocation of 104 Sample Schools (26 per Region) to the Region and SOC Strata per Grade Level

Region	Size of community					Total
	1SR[a]	*1S*[b]	*2*	*3*	*4*	
Northeast	3.1	3.6	9.7	6.7	3.0	26.1
Southeast	N[c]	2.1	4.7	8.4	10.8	26.0
Central	3.6	3.7	5.8	6.0	6.9	26.0
West	3.4	6.8	5.2	5.7	4.9	25.9
Total	10.1	16.2	25.4	26.8	25.6	104.0

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no schools were allocated to this substratum.

TABLE 6

Rounded Proportional Allocation of Sample Schools to the Region and SOC Strata per Grade Level

Region	Size of community					Total
	1SR[a]	*1S*[b]	*2*	*3*	*4*	
Northeast	3	4	10	7	3	27
Southeast	N[c]	2	5	8	11	26
Central	4[d]	4	6	6	7	27
West	3	7	5	6	5	26
Total	10	17	26	27	26	106

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no schools were allocated to this substratum.

[d]Two schools were selected from Chicago and two schools from Detroit.

and SOC 4. The selection of two schools per PSU allows for SES stratification within the Big City (SOC 1) and Suburban Fringe (SOC 2) PSUs at the school sample selection stage. (See Figure 1.)

For the "self-representing" SOC substratum (SOC 1SR) three schools per grade were selected from metropolitan New York (Region 1, SOC 1SR), two schools each from Cook County (Chicago, Illinois) and from Wayne County (Detroit, Michigan; Region 3, SOC 1SR), and three schools from Los Angeles County (Los Angeles, California, Region 4, SOC 1SR). This allocation allows for SES stratification at the school level within New York City, Cook, Wayne, and Los Angeles counties.

After adjusting for all of the above factors, a total planned sample size of 107 schools per grade level was selected. Table 7 gives the final allocation of the school sample to the region by SOC strata.

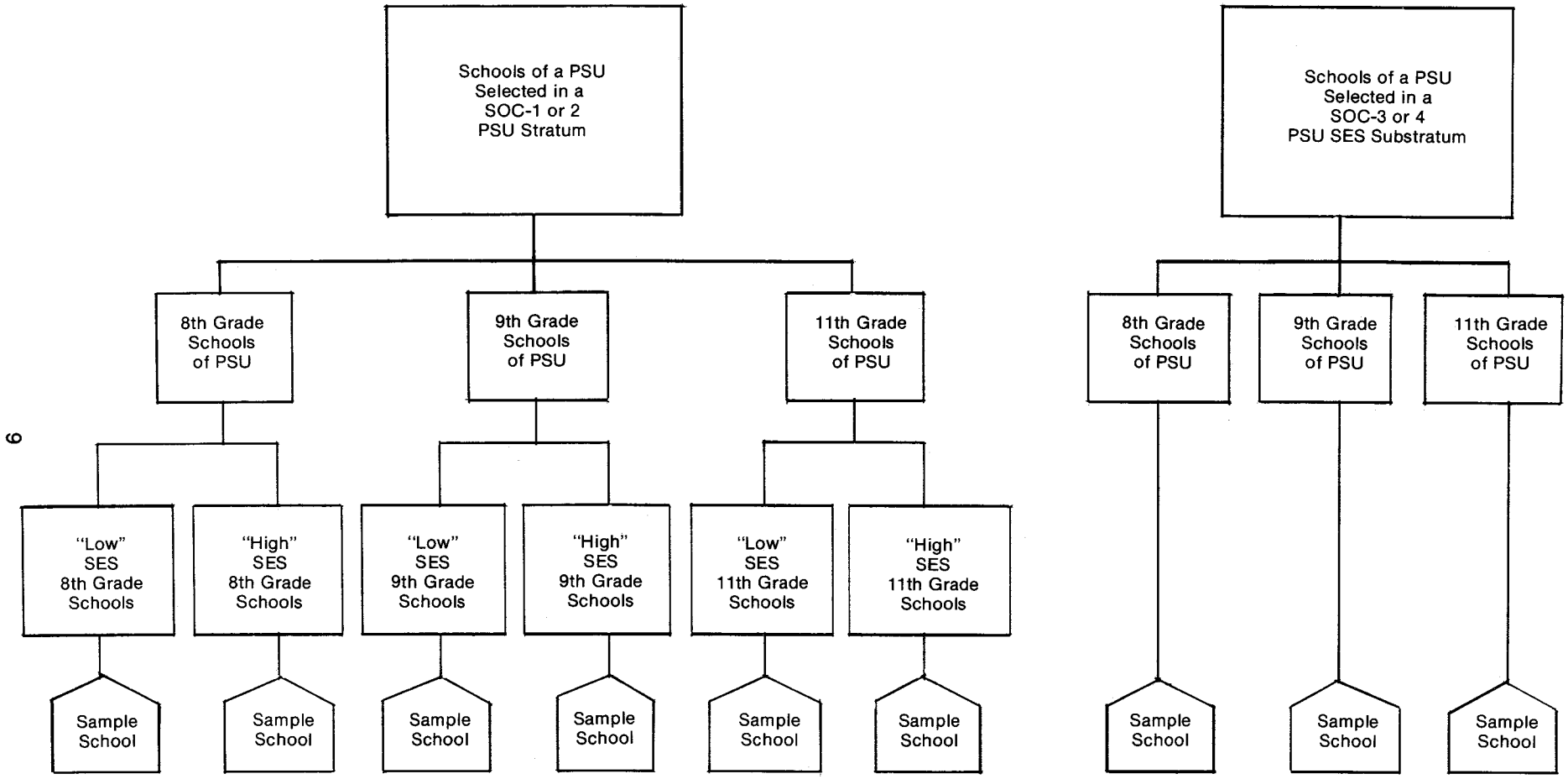

Fig. 1. Diagram of ACD/CPP 8-11 Norming Study Sample Design of PSUs.

Number of SES Substrata Required

The next step of the sampling process involved determining the number of SES substrata to be constructed, and the number of PSUs to be selected from each SES substratum. Table 8 gives the number of SES substrata within each region by SOC stratum. As mentioned above, substratification was not done at the PSU level within the SOC 1 and 2 strata in order to permit maximum SES stratification at the school sampling stage. An exception to this rule was made, however, in Region 1—SOC 2 which was divided into two SES substrata. This latter substratification was done because the urban fringe (SOC 2) in the Northeastern region (1) was allocated a large number of sample PSUs, the number of PSUs being proportional to the 17-year-old population. Table 9 gives the number of PSUs sampled from the region, SOC, and SES substrata for each grade level. The number of schools sampled from the region, SOC, and SES substrata for each grade level are presented in Table 10.

TABLE 7
Adjusted Proportional Allocation of Sample Schools to the Region and SOC Strata per Grade Level

	Size of community					
Region	*1SR*[a]	*1S*[b]	*2*	*3*	*4*	**Total**
Northeast	3	4	10	7	3	27
Southeast	N[c]	4	4	8	11	27
Central	4[d]	4	6	6	7	27
West	3	6	6	6	5	26
Total	10	18	26	27	26	107

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no schools were allocated to this substratum.

[d]Two schools were selected from Chicago and two schools from Detroit.

TABLE 8
Number of SES Substrata by Region and SOC Strata per Grade Level

	Size of community					
Region	*1SR*[a]	*1S*[b]	*2*	*3*	*4*	**Total**
Northeast	1	1	2	3	1	8
Southeast	N[c]	1	1	4	5	11
Central	2	1	1	3	3	10
West	1	1	1	3	2	8
Total	4	4	5	13	11	37

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no SES substrata were allocated to this substratum.

TABLE 9

Number of Sample PSUs by Region, SOC, and SES Substrata per Grade Level

	Size of community													
			2 SES substrata		*3 SES substrata*				*4 SES substrata*					
Region	1SR[a]	1S[b]	1	2	1	2	3	4	1	2	3	4	5	**Total**
Northeast	1	2	3	2	3	2	2	N	3	N	N	N	N	18
Southeast	N[c]	2	2	N	2	2	2	2	3	2	2	2	2	23
Central	2[d]	2	3	N	2	2	2	N	3	2	2	N	N	20
West	1	3	3	N	2	2	2	N	3	2	N	N	N	18
Total	4	9	11	2	9	8	8	2	12	6	4	2	2	79

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no PSUs were allocated to these substrata.

[d]Two schools were selected from Chicago and two schools from Detroit.

TABLE 10

Number of Sample Schools by Region, SOC, and SES Substrata per Grade Level

	Size of community													
			2 SES substrata		*3 SES substrata*				*4 SES substrata*					
Region	1SR[a]	1S[b]	1	2	1	2	3	4	1	2	3	4	5	**Total**
Northeast	3	4	6	4	3	2	2	N	3	N	N	N	N	27
Southeast	N[c]	4	4	N	2	2	2	2	3	2	2	2	2	27
Central	4[d]	4	6	N	2	2	2	N	3	2	2	N	N	27
West	3	6	6	N	2	2	2	N	3	2	N	N	N	26
Total	10	18	22	4	9	8	8	2	12	6	4	2	2	107

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no schools were allocated to these substrata.

[d]Two schools were selected from Chicago and two schools from Detroit.

PSU Sample Selection Procedure

The procedure for selecting the sample PSUs required selecting a sample of two PSUs (in some substrata three PSUs were selected) from each substratum using unequal selection probabilities strictly proportional to the number of 17-year-olds in the PSU. This sample selection procedure used was developed by M. R. Sampford (1967) and involves selecting PSUs without replacement. According to this procedure, a sample is drawn and accepted if it contains n different PSUs. Otherwise, the sample is rejected and the selection process repeated until an acceptable sample is drawn. Appendix B contains the technical details of the PSU sample selection procedure.

School Sample Selection Procedure

A school sampling frame was constructed for each sample PSU selected. Each sampling frame consisted of all public and Catholic schools in the PSU with enrollments in grades 8, 9, and 11. Certain enrollment data, e.g., grade range, total school enrollment, 8th and/or 9th and/or 11th grade enrollment(s), and the school zip code, were obtained for each school in this sampling frame.

A school SES index was defined as the percentage of families with income less than $3,000 in the zip code area of the school. This SES index for each school was determined from the RTI data file based on Internal Revenue Service income tax return data. Two components of the SES index were the PSU index based on county level data, and the school index based on zip code area for the area containing the school.

Within each PSU, three school samples, one each for grades 8, 9, and 11, were selected independently. School selection probabilities were proportional to the grade enrollment of the school.

In accordance with the overall sample design, only one school was selected per grade from each primary sampling unit (PSU) in SOC strata 3 and 4. Therefore, no stratification of schools was done within the PSU as the SES stratification was accomplished at the PSU stage of sampling.

To allow for SES stratification within the Big City and Suburban Fringe PSUs, the sample design called for the selection of two schools per grade level from each PSU in SOC substrata 1S and 2 and from Wayne County and Cook County in SOC substratum 1SR. Two school strata were constructed independently for each grade for each of these PSUs by dividing the total accumulated PSU enrollment for the grade of interest by two. Those schools whose cumulative enrollments were less than or equal to this result constituted the first school stratum and the remaining schools cumulated the second school stratum. Appendix C contains the technical details of the school selection procedure.

Each of the other two PSUs in SOC substratum 1SR, metropolitan New York and Los Angeles County, was allotted three schools per grade level. Therefore, three school strata were constructed in each.

IV. Participation of Sample Schools

Contact and Replacement

Each of the schools selected into the sample was sent a letter inviting the school to participate in the norming study. The principal was the contact person in all schools selected into the sample. Names and addresses were furnished by RTI using a current list of schools and principals. Enclosed with the letter were materials describing the instruments, as well as samples of the type of results that schools and students would receive. Also included was a reply form on which the school indicated whether it agreed or declined to participate. If the school decided to participate, the tentative testing date, the number of students to be tested, and an indication of whether a subsampling of the grade was to be done were to be included on the return form. This form was then to be sent back to ACT within a designated period of time.

If a school did not respond to the original letter, a follow-up letter was sent with a duplicate set of enclosures. Schools which did not respond to the follow-up were called by telephone and asked for their decision about participation. Each school was followed up to assure that all of these selected schools had been contacted and that each had indicated a decision to participate or not to participate.

For each of the 107 schools per grade originally selected into the sample, a backup school was designated to replace each one not participating. This backup school was similar to the original

school in that it was selected from the same substratum and in the same manner as the original.

Student Subsampling

Participating schools were asked to indicate whether an entire grade or a subsample of 60 students was to be tested. The option was offered to schools to accommodate different schedules and situations. It was hoped that one of the two options (to test either the entire grade or just the 60 students) would be found workable by most schools.

Schools which had decided to test a subsample of the grade were instructed to forward to ACT a list of all of the students in the designated grade in the school.

ACT numbered all of the students on the grade list sequentially, then using random number tables selected a simple random sample of 60 students and a supplementary simple random sample of 6 extra students to be tested in the event that some of those selected would not be present for the testing. ACT then made a label for each student, attached a label to each test booklet, and returned the original grade list (with the 66 names circled) in the shipment of test materials to the school.

Testing and Scoring Procedures

Normative testing for the ACD and CPP 8-11 took place between February and May of 1973. ACT had initially wanted to complete all testing by April, but as often is the case with schools involved in scheduling guidance tests, circumstantial delays were encountered and the testing time had to be extended by one month.

The actual amount of time required to administer and complete the ACD and CPP 8-11 was 4 1/4 hours. In the normative testing, both assessment instruments were combined into one test booklet. Two distinct administration models were designed, one for schools following a 40-minute period schedule and the other for schools on a 55-minute period schedule. Each of these models divided the administration time into two 2-hour sessions. Schools were urged to follow one of these suggested models. While it appears that the majority of schools did indeed follow one of these procedures, many schools felt that such a time arrangement was unsuitable for their daily schedule. These schools, then, administered the ACD and CPP 8-11 by using one class period a day until all of the testing was completed. The arrangement of test units for administration within individual class periods appears to be a better method for testing than the traditional procedure which usually requires one or two blocks of testing time, especially in norm group schools which are volunteering their services. In each situation, however, time limits for all test units were followed.

When each school completed the ACD and CPP 8-11 normative testing, it returned all answer folders for scoring. Upon completion of scoring, a tape of all item responses and scale scores for each student in the ACD/CPP 8-11 norming study was prepared. Normative distributions for each scale and for selected items on both instruments were then constructed, as described in Part V of this report. Finally each school was mailed a report of its results. For the ACD, group summary reports and student list reports were furnished; for the CPP 8-11, this included an individual report and a student booklet for each student. Both sets of reports represented the full services available for the measurement components of each instrument.

Norm Group Obtained

A total of 201 schools participated in the ACD/CPP 8-11 norming study. However, because of administrative problems (e.g., answer folders being temporarily lost in the mail), the final norm group consisted of 197 schools. There were 61 8th, 64 9th, and 72 11th grade schools in the sample (see Part V for a discussion of adjustments made for non-response). Schools are tallied by grade and size of community in Table 11, and by grade and region in Table 12. Tables 13, 14, and 15 show the distributions of schools for grades 8, 9, and 11 by region, size of community, and socioeconomic substrata. The 197 schools represent 55, 60, and 62 PSUs from grades 8, 9, and 11 respectively. These PSUs are presented by grade and size of community in Table 16 and by grade and region in Table 17

Although approximately 33,000 students from the 201 schools participated in the study, not all students completed both the ACD and CPP 8-11. As might be expected, absence during one of the testing dates was one of the main reasons some students did not complete all of the units for both instruments. In addition, answer folders from students in the four schools mentioned previously were not available for scoring. Thus, because of these circumstances, data for all of the units of both the ACD and CPP 8-11 were available for 28,298 students. This final total of students is presented by grade and sex in Table 18.

TABLE 11

Number of Schools in Sample by Grade and SOC

Size of community	Grade 8	Grade 9	Grade 11	Total
Counties containing big cities (SOC 1)	13	17	14	44
Counties located in fringes around big cities (SOC 2)	11	9	19	39
Counties containing medium-sized cities (SOC 3)	13	16	15	44
Remaining counties (SOC 4)	24	22	24	70
Total	61	64	72	197

TABLE 12

Number of Schools in Sample by Grade and Region

Region	Grade 8	Grade 9	Grade 11	Total
Northeast	11	13	16	40
Southeast	18	17	20	55
Central	16	15	19	50
West	16	19	17	52
Total	61	64	72	197

TABLE 13

Grade 8 Sample Schools by Region, SOC, and SES Substrata

	Size of community													
			2 SES substrata		3 SES substrata				4 SES substrata					
Region	1SR[a]	1S[b]	1	2	1	2	3	4	1	2	3	4	5	Total
Northeast	1	3	N	2	1	1	0	N	3	N	N	N	N	11
Southeast	N[c]	1	4	N	1	1	1	1	3	1	1	2	2	18
Central	2	1	2	N	1	1	2	N	3	2	2	N	N	16
West	1	3	3	N	1	1	1	1	3	2	N	N	N	16
Total	4	8	9	2	4	4	4	2	12	5	3	2	2	61

[a] 1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b] 1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c] N indicates that no schools were allocated to these substrata.

TABLE 14

Grade 9 Sample Schools by Region, SOC, and SES Substrata

	Size of community													
			2 SES substrata		*3 SES substrata*				*4 SES substrata*					
Region	1SR[a]	1S[b]	1	2	1	2	3	4	1	2	3	4	5	**Total**
Northeast	1	2	2	2	1	2	1	N	2	N	N	N	N	13
Southeast	N[c]	3	1	N	1	1	1	1	3	1	1	2	2	17
Central	2	3	2	N	0	1	1	N	3	1	2	N	N	15
West	2	4	2	N	2	2	2	N	3	2	N	N	N	19
Total	5	12	7	2	4	6	5	1	11	4	3	2	2	64

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no schools were allocated to these substrata.

TABLE 15

Grade 11 Sample Schools by Region, SOC, and SES Substrata

	Size of community													
			2 SES substrata		*3 SES substrata*				*4 SES substrata*					
Region	1SR[a]	1S[b]	1	2	1	2	3	4	1	2	3	4	5	**Total**
Northeast	1	1	5	2	2	1	1	N	3	N	N	N	N	16
Southeast	N[c]	3	3	N	2	1	1	0	3	1	2	2	2	20
Central	3	2	4	N	2	1	1	N	2	2	2	N	N	19
West	1	3	5	N	1	1	1	N	3	2	N	N	N	17
Total	5	9	17	2	7	4	4	0	11	5	4	2	2	72

[a]1SR indicates that PSUs in these region by SOC 1 substrata were "self-representing" (i.e., Metropolitan New York City, Chicago, Detroit, and Los Angeles).

[b]1S indicates that PSUs in these region by SOC 1 substrata were "sampled" independently of those in the SOC 1SR substrata.

[c]N indicates that no schools were allocated to these substrata.

TABLE 16

Number of PSUs in Sample by Grade and SOC

Size of community	Grade 8	9	11
Counties containing big cities (SOC 1)	10	13	11
Counties located in fringes around big cities (SOC 2)	8	9	12
Counties containing medium-sized cities (SOC 3)	13	16	15
Remaining counties (SOC 4)	24	22	24
Total	55	60	62

TABLE 17

Number of PSUs in Sample by Grade and Region

Region	Grade 8	9	11
Northeast	10	13	13
Southeast	16	16	18
Central	15	14	16
West	14	17	15
Total	55	60	62

TABLE 18

Number of Students in Sample by Grade and Sex

Sex	Grade 8	9	11	Total
Male	4,384	5,342	4,623	14,349
Female	4,438	4,827	4,684	13,949
Total	8,822	10,169	9,307	28,298

V. Sample Weighting to Assure Representativeness

Student Probability

Once the three stage probability sample of 8th, 9th, and 11th grade students was selected, it was possible to designate for each student in the sample the probability that he or she would be selected. This overall probability of selection was obtained by first determining the probability of selecting the student's PSU, then figuring the probability of selecting the student's school, given that his or her PSU had been selected, and finally calculating the probability of selecting the student, given that his or her school was selected into the sample. This latter probability was equal to one in cases in which the school elected to test the entire grade, and equal to 60/(number of students in designated grade in the school) in situations in which the school elected to test a subsample of 60 students. Once these three probabilities—the probability of selecting the PSU, the school within the PSU, and the student within the school—were determined, they could be combined as follows to determine the overall probability of selecting a student into the sample:

probability of selecting PSU X probability of selecting school within PSU X probability of selecting student within school = probability of selecting a student into the sample

Definition and Illustration of a Statistical Weight

In order to estimate from the sample students the normative distribution of a given scale for the population of interest, the statistical weight for each sample student must be determined. The statistical weight for a given student is simply the reciprocal of the overall probability of the student being included in the sample. The formula for the statistical weight of each sample student is therefore:

$$\text{Student weight} = \frac{1}{\text{probability of selecting PSU}} \times \frac{1}{\text{probability of selecting school within PSU}} \times \frac{1}{\text{probability of selecting student within school}} = \frac{1}{\text{probability of selecting the student into the sample}}.$$

The interpretation of a weight value of 500, for example, is that the student in the sample with that weight value represents a total of 500 students in the population from which the sample was drawn. Students selected into the sample with smaller probabilities have larger weights while students selected into the sample with large probabilities have smaller weights.

By summing the student weights for different categories of students, estimates of population totals may be obtained. The sum of the weights of all female 8th grade students selected into the sample, for example, provides an estimate of the total number of female 8th grade students in the population. The sum of the weights for all 11th grade students selected into the sample provides an estimate of the total number of 11th grade students in the population.

Adjusting for Nonresponse

The above description of the summing of student weights to obtain population estimates assumed that every student selected into the sample completed the test batteries. This is rarely, if ever, the case in large scale norming surveys and was not the case in the ACD/CPP 8-11 norming study. While 321 schools, located in 79 PSUs, were selected into the sample and invited to participate, testing was completed in only 201 schools. The remaining schools were either unable or unwilling to participate in the ACD/CPP 8-11 norming study. The reasons for this decision were varied. Many schools indicated that they did not have the time for this testing; others reported that they could not accommodate any additional guidance testing; some schools felt that their students had no need for testing of this type. Certainly time constraints are a major factor in institutional planning and scheduling and many schools allot only a limited amount of time for guidance testing.

Since not all selected schools participated in the norming study and since not all selected students completed the ACD and CPP 8-11, some adjustment was necessary to account for this nontesting. The adjustment was made by means of an additional weighting process. Within a school, tested students were assigned a weight to account for those sample students who had not completed the ACD and CPP 8-11. In cases in which no students were tested by a school, that school was matched with a similar nearby school that had participated in the norming study, and students in the participating school were assigned a weight to account for the nontested school. The matching was based upon the structure of the sample design. Thus, the weighting to correct for nonresponse was done by matching schools by region, size of community, and socioeconomic characteristics. This procedure should have a substantial effect on reducing any biases that would otherwise occur in the estimates due to the number of nonresponding schools in the study. Ideally, the best method to examine such bias is to test a subsample of the nonresponding schools. However, this procedure was not possible in the ACD/CPP 8-11 norming study because of time constraints and other reasons related to the schools' initial inability to participate. It is important to note that the results of the ACD and CPP 8-11 were not used in any manner to match schools.

Thus, each of the 28,298 sample students who completed the ACD and CPP 8-11 was assigned a final weight made up of the product of the weight components.

$$\text{Final student weight} = \frac{1}{\text{overall probability of selecting the student into the sample}} \times \text{weight to adjust for nontesting of some of the sample schools} \times \text{weight to adjust for nontesting of some of the sample students}$$

Estimating Normative Distributions

Using the "final student weight" which reflected both the probability of selecting the student and adjustments to compensate for nontesting, estimates of population characteristics were made. For example, the population proportion of 11th grade students planning to go to college (an ACD item response) was estimated from the following ratio:

$$\frac{\text{Sum of "final student weights" for all tested 11th grade sample students planning to go to college}}{\text{Sum of "final student weights" for all tested 11th grade sample students for all of the item responses}} = \text{the estimated population proportion of 11th grade students planning to go to college}$$

Similarly, the population proportion of 9th grade students who would obtain a score of 20 on a particular scale is estimated as follows:

$$\frac{\text{Sum of "final student weights" for all tested 9th grade sample students with a score of 20}}{\text{Sum of "final student weights" for all tested 9th grade sample students for the entire scale}} = \text{the estimated population proportion of 9th grade students who would have had a score of 20 had all 9th graders in the population been tested}$$

Thus, by computing the ratios of the sums of "final student weights" for different groups (grade by sex) of the tested sample students, the normative distributions for the ACD and the CPP 8-11 were obtained.

Distributions That Were Calculated

Weighted and unweighted normative distributions for each scale and selected items on the ACD and CPP 8-11 were calculated for grade-by-sex groups and for the combined group of males and females in each grade. Two examples of these distributions are shown. Table 19 presents the normative distribution for the ninth grade on a true-false ACD item. Included are weighted and unweighted frequencies, and weighted percentages for males, females, and the total group. Table 20 presents the normative distribution of an ACD scale for grade 11. This table contains both weighted and unweighted frequencies, weighted cumulative frequencies, weighted percentages, and weighted cumulative percentages for the total group.

Normative distributions for grade 10 were obtained by interpolation of 8th, 9th, and 11th grade results for all items and scales except the Reading Skills and Numerical Skills ability measures on the CPP 8-11. Interpolation was not possible for these scales because 8th and 9th grade students completed different Reading Skills and Numerical Skills ability scales than did 11th graders. Thus, extrapolation of 8th and 9th grade results on these two scales was performed to obtain 10th grade Reading Skills and Numerical Skills scale distributions.

Additional Normative Distributions

As previously stated, the normative distributions constructed for the ACD and CPP 8-11 were for males, females, and the total group for grades 8, 9, and 11. However, because of the stratification in the sample design, a considerable number of other normative distributions are possible. These might include, for example, the variables that were used to stratify the ACD/CPP 8-11 sample, such as region of the country and size of community.

Final Remarks

This report has presented some of the major considerations in conducting normative studies along with illustrations of procedures that can be used and problems that may be encountered. It is hoped that this report will contribute to a more thorough consideration of test norming procedures, both by test publishers and by test users.

TABLE 19

Distributions of a Selected ACD Item for Grade 9

Item response	Males			Females			Total		
	Weighted[a] frequency	*Unweighted[b] frequency*	*Weighted percentage*	*Weighted[a] frequency*	*Unweighted[b] frequency*	*Weighted percentage*	*Weighted[a] frequency*	*Unweighted[b] frequency*	*Weighted percentage*
False[c]	1,077,723	2,404	55.2	1,300,226	2,798	62.8	2,377,949	5,202	59.1
True	874,479	1,961	44.8	770,246	1,637	37.2	1,644,725	3,598	40.9
Total	1,952,202	4,365		2,070,472	4,435		4,022,674	8,800	

Note.—This ACD item asks for a True-False response to the statement, "Few women work outside of the home after marriage."

[a]Each student received a unique statistical weight to insure representativeness of the sample. This weight, applied to all item responses of the student, was used in calculating frequency distributions, percentages, etc.

[b]Unweighted indicates that no statistical weight was applied.

[c]The correct response.

TABLE 20

Distributions of a Selected ACD Scale for Grade 11

Item	Total group				
	Weighted[a] *frequency*	*Unweighted*[b] *frequency*	*Weighted cumulative frequency*	*Weighted percentage*	*Weighted cumulative percentage*
18	4,921	2	4,032,418	0.1	100.0
17	9,637	14	4,027,497	0.2	99.9
16	31,024	72	4,017,860	0.8	99.7
15	101,266	183	3,986,836	2.5	98.9
14	199,798	410	3,885,570	5.0	96.4
13	291,789	652	3,685,772	7.2	91.4
12	442,528	956	3,393,983	11.0	84.2
11	478,851	1,041	2,951,455	11.9	73.2
10	499,987	1,146	2,472,604	12.4	61.3
9	495,448	1,065	1,972,617	12.3	48.9
8	391,146	859	1,477,169	9.7	36.6
7	335,072	746	1,086,023	6.0	26.9
6	243,982	556	750,951	6.1	18.6
5	203,116	452	506,969	5.0	12.6
4	157,690	343	303,853	3.9	7.6
3	79,366	172	146,163	2.0	3.7
2	45,167	104	66,797	1.1	1.7
1	19,499	43	21,630	0.5	0.6
0	2,131	6	2,131	0.1	0.1
Total	4,032,418	8,822			

Note.—This was the ACD "Knowledge of Occupational Preparation Requirements" scale.

[a]Each student received a unique statistical weight to insure representativeness of the sample. This weight, applied to all item responses of the student, was used in calculating frequency distributions, percentages, etc.

[b]Unweighted indicates that no statistical weight was applied.

REFERENCES

The American College Testing Program. *Handbook for the ACT Assessment of Career Development.* Iowa City, Iowa: Author, 1974. (a)

The American College Testing Program. *Handbook for the ACT Career Planning Program* (Grades 8-11). Iowa City, Iowa: Author, 1974. (b)

American Psychological Association. *Standards for educational and psychological tests and manuals.* Washington, D.C.: Author, 1966.

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement.* Washington: American Council on Education, 1971.

Cornell, F. G. Sampling methods. In C. W. Harris (Ed.), *Encyclopedia of educational research.* (3rd ed.) New York: Macmillan, 1960.

Hansen, M. H., Hurwitz, W. H., & Madow, W. G. *Sample survey methods and theory.* New York: Wiley, 1953.

Kish, L. *Survey sampling.* New York: Wiley, 1965.

Moore, R. P., Chromy, J. R., & Rogers, W. T. National assessment's approach to sampling—Year 02. *National Assessment of Educational Progress* (Monograph), 1974, in press.

Prediger, D. J., Roth, J. D., & Noeth, R. J. *A nationwide study of student career development: Summary of results.* ACT Research Report No. 61. Iowa City, Iowa: The American College Testing Program, 1973.

Rand Corporation. *A million random digits with 100,000 normal deviates.* Glencoe, Ill.: The Free Press, 1955.

Sampford, M. R. On sampling without replacement with unequal probabilities of selection. *Biometrika,* 1967, **54,** 499-512.

Wilks, S. S. *Mathematical statistics.* New York: Wiley, 1962.

APPENDIX A

Statistical Precision Requirements of the ACD/CPP 8-11 Norming Study Design

The sampling specifications for the study required the following level of statistical precision for each estimated normative distribution: the 95% confidence interval for a scale value corresponding to any percentile rank, P, which exceeds 74, should cover at most 6 percentile rank units. That is, in replicated samplings according to the ACT sample design, the scale values corresponding to P ± 3 (when P exceeds 74) should include the true scale value, $\hat{X}p$, at least 95% of the time. This requirement was directly related to the report mode used for CPP 8-11 ability and interest scale scores, i.e., stanine bands with high probability of including the true standing of a student in the target population.

It is possible to restate the statistical precision requirement of the sample design in a more familiar form (Kish, 1965; Hansen, Hurwitz, & Madow, 1953). It is required that the 95% confidence interval for an estimated percentile rank, Px, which exceeds 74%, should span no more than 6 percentage points. Using the normal approximation, the sample size n which meets this precision requirement is determined as:

$$n \geq \frac{P(1-P)}{(0.3)^2} \ (2)^2 \ . \qquad (1)$$

The sample size n thus determined is the number of students needed to satisfy the precision requirements, if a simple random sample of students were to be selected. The sample design planned, however, was that of a multi-stage, stratified, cluster sample of students. Since the size of the error variance depends not only on the sample size but on the sample design as well, the concept of the design effect (DEFF) was used.

$$\text{DEFF} = \frac{\text{error variance of an estimate based on the sample design used}}{\text{error variance of an estimate based on a simple random sample the same size}}$$

In practice, stratification usually has the effect of slightly decreasing the error variance while clustering usually has the effect of substantially increasing the error variance when each is compared to a simple random sample. In order to take into consideration the sample design that was planned, formula (1) was revised to be:

$$n \geq \frac{(\text{DEFF})\ (P)\ (1-P)}{(.03)^2} \ (2)^2$$

or, for P=74:

$$n \geq (\text{DEFF})\ (855).$$

Since it was planned that in most of the sample schools all students in a particular grade would be tested, the sample design for the norming study could be considered to be a two-stage (PSU and school) cluster sample with two schools selected from approximately one half of the primary sampling units (PSUs) and one school selected from the remaining PSUs. Three independent school samples were to be drawn in each PSU; one sample of schools containing the 8th grade, another containing the 9th, and a third containing the 11th grade. If s denotes the average number of schools per PSU and k the average number of students per school for a particular grade by sex subpopulation, the design effect for such a sample takes the form:

$$\text{DEFF}\ (s,k) = [1 + (sk-1)\ \delta_1 + (k-1)\ \delta_2\ (1-\delta_1)]$$

where δ_1 is the between student within PSU cluster coefficient and δ_2 the between student within school cluster coefficient. Based on RTI's experience with the National Assessment of Educational Progress data, average values of $\delta_1 = .015$ and $\delta_2 = .086$ are found. Thus, the design effect would be:

$$\text{DEFF}\ (s,k) = [1 + (sk-1)\ (.015) + (k-1)\ (.086)\ (1-\delta_1)] \ .$$

Assuming an average 8th grade school enrollment of 125 students distributed equally by sex, 62.5

[Continued]

males and 62.5 females, the design effect for 8th grade males (or females) would be:

$$\text{DEFF}\ (1.5, 62.5) = 7.62\ .$$

This requires a total sample size of:

$$n \geq (7.62) \times (855) = 6{,}515.10$$

to guarantee that a 95% confidence interval for a percentile in excess of 74% will span no more than 6 percentage points. If an average of 62.5 males per school is expected, then a total of (6515/62.5) = 104 schools will be required to meet the stated statistical precision requirement. Thus, 104 schools for each of grades 8, 9, and 11 was set as the desired number of schools for the sample. This was later rounded to 107 in school allocation to the region and size of community strata.

APPENDIX B

The PSU Sample Selection Procedure

A computer program was used to perform calculations necessary for actual PSU sample selection.

Let M_i = Number of 17-year-olds in PSU(i) of a given substratum and

$\sum_{i=1}^{N} M_i = M$ = Total number of 17-year-olds in the substratum.

The values P_i and λ_i were calculated for each PSU such that:

$$P_i = \frac{M_i}{\sum_{i=1}^{N} M_i} = \frac{M_i}{M} \quad \text{and} \quad \lambda_i = \frac{P_i}{1-nP_i} \quad ,$$

when N is the number of PSUs in the substratum and n is the number of sample PSUs selected out of N in the substratum. These values were used to find:

$$\sum_{i=1}^{N} \lambda_i \quad , \quad \theta_i = \frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i} \quad , \text{ and } \sum_{i=1}^{n} \theta_i \quad .$$

These values were listed on the computer as follows: (See Figure B-1, which is a copy of a *labeled* computer printout listing.)

"# 17S" = 17-year-old population of PSU(i) = M_i ;

"PROP. 17S" = proportion of substratum 17-year-old population contained in PSU = P_i ;

"ACCUM. 17S" = cumulative sum of PSU 17-year-old population = ΣM_i ;

"ACCUM. % 17S" = cumulative sum of "Prop. 17S," i.e., cumulative sum of $P_i = \Sigma P_i$;

"LAMBDA" = $\lambda_i = \frac{P_i}{1-2P_i}$;

"PROP. LAMBDA" = $\theta_i = \frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i}$;

"ACCUM. LAMBDA" = $\sum_{i=1}^{N} \lambda_i$ = cumulative sum of λ_i ;

"ACCUM. % LAMBDA" = $\sum_{i=1}^{N} \theta_i$ = cumulative sum of θ_i .

To begin the sample selection process, a group of four four-digit random numbers (Rand, 1955) was recorded on the "PSU Sample Selection Form." (See Figure B-2.) The groups of random numbers and the two columns of the computer listing labeled "ACCUM. % 17S" and "ACCUM. % LAMBDA" were used in the following way to select the sample of n=2 PSUs.

The first random number listed on the sample selection form was matched to the "ACCUM. % 17S" column of the computer output (see Figure B-1). The PSU whose "ACCUM. % 17S" value forms the upper bound of the range in which the random number falls was the sample PSU selected. The second random number in the group was then matched to the output column labeled "ACCUM. % LAMBDA," and the PSU whose "ACCUM. % LAMBDA" value forms the upper bound of the range in which the random number falls was the second sample PSU selected. If the same PSU was selected each time, the sample was discarded and the entire process repeated using the third and fourth random numbers appearing on the sample selection form. When an acceptable sample was obtained, a six-digit RTI PSU number was assigned each sample PSU. The first two digits consisted of the PSU's region and SOC code respectively. A two-digit state code comprised the third and fourth digits of the RTI PSU number. The fifth digit was the SES substratum code assigned within each region by SOC sampling stratum, the higher numbered codes representing those PSUs having the largest estimated proportions of people with incomes of less than $3,000. A "1," "2," or "3" labeling the selected PSUs within the substratum was assigned to the last position of the RTI PSU number. This six-digit RTI PSU code was recorded on the top of the PSU sample selection form (Figure B-2). This sampling process was repeated for each substratum until a sample with the required number of PSUs was obtained.

[*Continued*]

APPENDIX B [*Continued*]

# 17S	PROP. 17S	ACCUM. 17S	ACCUM. % 17S	LAMBDA	PROP. LAMBDA	ACCUM. LAMBDA	ACCUM. % LAMBDA	RECORD NO.	PSU NO.
10953	0.0655	10953	0.0655	0.0754	0.0570	0.0754	0.0570	200	4
18399	0.1101	29352	0.1756	0.1412	0.1068	0.2166	0.1638	400	7
18252	0.1092	47604	0.2848	0.1397	0.1056	0.3563	0.2694	600	2
19491	0.1166	67095	0.4014	0.1521	0.1150	0.5084	0.3844	800	10
12294	0.0736	79389	0.4750	0.0862	0.0652	0.5946	0.4496	1000	3
16498	0.0987	95887	0.5737	0.1230	0.0930	0.7176	0.5426	1200	1
11228	0.0672	107115	0.6409	0.0776	0.0587	0.7952	0.6013	1400	9
9952	0.0595	117067	0.7004	0.0676	0.0511	0.8628	0.6524	1600	8
16367	0.0979	133434	0.7983	0.1218	0.0921	0.9846	0.7445	1800	6
33700	0.2017	167134	1.0000	0.3379	0.2555	1.3225	1.0000	2000	5

Fig. B-1. Computer printout of PSU sample selection procedure program.

Stratum Number 311

Number of PSUs in Stratum 10

Number of PSUs to Be Sampled 2

Total Enrollment of Stratum 167134

No.	Random Number	PSU Serial Number	Record Number	State No.	No. Counties	Population	RTI PSU No.	State Name	County Name
1	7890	6	1800	39	1	16367	313911	Ohio	Franklin
2	5970	9	1400	39	1	11228	313912	Ohio	Summit
3	3718		2416						
4									
5									
6									
7									
8									

Source of Random Numbers:

Rand Million Random Digits _____ Page 300 Row 23 Column 6.

Sampler Name _____ _____ Date

Checked by _____ _____ Date

Fig. B-2. PSU sample selection form.

APPENDIX C

School Sample Selection Procedure

A computer program listing the following information by school, in ascending order of school-SES index, was developed for school sample selection purposes.

1. RTI PSU number and school identification number
2. Zip code and SES index for the school
3. Grade range
4. Total enrollment of the school and cumulative total enrollment
5. Eighth grade enrollment of the school and cumulative eighth grade enrollment
6. Ninth grade enrollment of the school and cumulative ninth grade enrollment
7. Eleventh grade enrollment of the school and cumulative eleventh grade enrollment

The resulting listings were used to select the samples of schools. Within each sample PSU, three separate samples of schools were drawn, one sample for each of grades 8, 9, and 11. For each grade the sample schools were selected with probability proportional to the school enrollment for that grade.

For PSUs in SOC 3 and 4, the stratification by SES occurred at the PSU selection stage and not at the school selection stage. In these PSUs, the school selection process involved choosing a random number between "1" and the total enrollment in the PSU for the grade of interest and matching this number to the grade enrollment "ACCUM" column shown on the computer printout. The school whose "ACCUM" value formed the upper bound of the range in which the random number fell was the sample school selected. This school was assigned a two-digit RTI school number, the first digit a "0" indicating no stratification of schools within the PSU, the second digit a "1," "2," or "4" indicating the grade for which the school was sampled, i.e., grade 8, 9, or 11. This school selection process was repeated (for a PSU) until an independent school sample was obtained for each of grades 8, 9, and 11.

For PSUs in SOC 1 and 2, the stratification by SES occurred at the school selection stage. The schools in a PSU were listed on the computer printout in ascending order of SES index. The first stratum was considered to be made up of those schools with the lower SES percentages, i.e., those schools with a lower percentage of families with incomes less than $3,000. The second stratum was considered to be made up of those schools with the higher SES codes. For each of the three grades, one school was selected from the first stratum by matching a random number between "1" and the total enrollment of the "low" SES school stratum for that grade to the "ACCUM" column on the computer printout which corresponded to the grade level being sampled. The school whose "ACCUM" value formed the upper bound of the range in which the random number fell was the sample school selected. The school was assigned a two-digit RTI school number, with the first digit a "1," representing the first school stratum, and the second digit a "1," "2," or "4" indicating the grade for which the school was sampled, i.e., grade 8, 9, or 11. For each of the three grades, one school was selected from the second school stratum by the same technique. New accumulated enrollments for the second stratum schools were calculated by subtracting the total enrollment of the first stratum from each of the second stratum school values; this new accumulated enrollment data was recorded on the computer printout. A random number between "1" and the total grade enrollment of the second school stratum was selected and matched with the cumulative grade enrollments. The school whose cumulative value formed the upper bound of the range in which the random number fell was the sample school selected to represent those schools having the higher percentage of families with incomes less than $3,000. A two-digit school number was assigned the selected school as before, with the first digit a "2" representing the second school stratum and the second digit a "1," "2," or "4" indicating the grade for which the school was sampled, i.e., grade 8, 9, or 11.

Three schools were selected from metropolitan New York and three from Los Angeles County (SOC 1SR). Three substrata were constructed within each of these PSUs by application of techniques previously discussed for the case of two substrata. One school was then selected from each substratum as before and a two-digit RTI selection number was

assigned the selected schools. The first digit was a "1," "2," or "3" depending on the school stratum represented, and the second digit again indicated the grade for which the school was sampled.

A final computer listing of these sample schools with the following necessary information was produced:

1. RTI PSU number (six-digit code)

2. School district identification including the district name, address, telephone number, and superintendent's name

3. Information for sample schools within the district such as school name and address (street, city, state, and zip code) and telephone number

4. Principal's name

5. School number (two-digit RTI school selection number)

6. School grade range, grade for which the school was sampled, and that grade enrollment.

ACT Research Reports

This report is Number 65 in a series published by the Research and Development Division of The American College Testing Program. The first 26 research reports have been deposited with the American Documentation Institute, ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20540. Photocopies and 35 mm. microfilms are available at cost from ADI; order by ADI Document number. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress. Beginning with Research Report No. 27, the reports have been deposited with the National Auxiliary Publications Service of the American Society for Information Science (NAPS), c/o Microfiche Publications, 305 East 46th Street, New York, New York 10017. Photocopies and 35 mm. microfilms are available at cost from NAPS. Order by NAPS Document number. Advance payment is required. Printed copies ($1.00) may be obtained, if available, from ACT Publications , The American College Testing Program, P.O. Box 168, Iowa City, Iowa 52240. A check or money order must accompany the request.

The reports since October 1970 in this series are listed below. A complete list of the reports can be obtained by writing to ACT Publications.

No. 37 *Practices and Outcomes of Vocational-Technical Education in Technical and Community Colleges,* by T. G. Gartland, & J. F. Carmody (NAPS No. 01441; photo, $6.80; microfilm, $2.00)

No. 38 *Bayesian Considerations in Educational Information Systems,* by M. R. Novick (NAPS No. 01442; photo, $5.00; microfilm, $2.00)

No. 39 *Interactive Effects of Achievement Orientation and Teaching Style on Academic Achievement,* by G. Domino (NAPS No. 01443; photo, $5.00; microfilm, $2.00)

No. 40 *An Analysis of the Structure of Vocational Interests,* by N. S. Cole, & G. R. Hanson (NAPS No. 01444; photo, $5.00; microfilm, $2.00)

No. 41 *How Do Community College Transfer and Occupational Students Differ?* by E. J. Brue, H. B. Engen, & E. J. Maxey (NAPS No. 01445; photo, $5.50; microfilm, $2.00)

No. 42 *Applications of Bayesian Methods to the Prediction of Educational Performance,* by M. R. Novick, P. H. Jackson, D. T. Thayer, & N. S. Cole (NAPS No. 01544; photo, $5.00; microfilm, $2.00)

No. 43 *Toward More Equitable Distribution of College Student Aid Funds: Problems in Assessing Student Financial Need,* by M. D. Orwig (NAPS No. 01543; photo, $5.00; microfilm, $2.00)

No. 44 *Converting Test Data to Counseling Information,* by D. J. Prediger (NAPS No. 01776; photo, $5.00; microfiche, $2.00)

No. 45 *The Accuracy of Self-Report Information Collected on the ACT Test Battery: High School Grades and Items of Nonacademic Achievement,* by E. J. Maxey, & V. J. Ormsby (NAPS No. 01777; photo, $5.00; microfiche, $2.00)

No. 46 *Correlates of Student Interest in Social Issues,* by R. H. Fenske, & J. F. Carmody (NAPS No. 01778; photo, $5.00; microfiche, $2.00)

No. 47 *The Impact of College on Students' Competence to Function in a Learning Society,* by M. H. Walizer, & R. E. Herriott (NAPS No. 01779; photo, $5.00; microfiche, $2.00)

No. 48 *Enrollment Projection Models for Institutional Planning,* by M. D. Orwig, P. K. Jones, & O. T. Lenning (NAPS No. 01780; photo, $5.00; microfiche, $2.00)

No. 49 *On Measuring the Vocational Interests of Women,* by N. S. Cole (NAPS No. 02071; photo, $5.00; microfiche, $1.50)

No. 50 *Stages in the Development of a Black Identity,* by W. S. Hall, R. Freedle, & W. E. Cross, Jr. (NAPS No. 02072; photo, $5.00; microfiche, $1.50)

No. 51 *Bias in Selection,* by N. S. Cole (NAPS No. 02073; photo, $5.00; microfiche, $1.50)

No. 52 *Changes in Goals, Plans, and Background Characteristics of College-Bound High School Students,* by J. F. Carmody, R. H. Fenske, & C. S. Scott (NAPS No. 02074; photo, $5.75; microfiche, $1.50)

No. 53 *Toward an Integration of Theory and Method for Criterion-Referenced Tests,* by R. K. Hambleton, & M. R. Novick (NAPS No. 02075; photo, $5.00; microfiche, $1.50)

No. 54 *College Student Migration,* by R. H. Fenske, C. S. Scott, & J. F. Carmody (NAPS No. 02215; photo, $5.00; microfiche, $1.50)