CAB D0007188.A2/Final
December 2003

# On the Representativeness of Norming Samples for Aptitude Tests

William H. Sims • Catherine M. Hiatt

CNA
4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Approved for distribution: December 2003

Henry S. Griffis

Henry S. Griffis, Director
Workforce, Education and Training Team
Resource Analysis Division

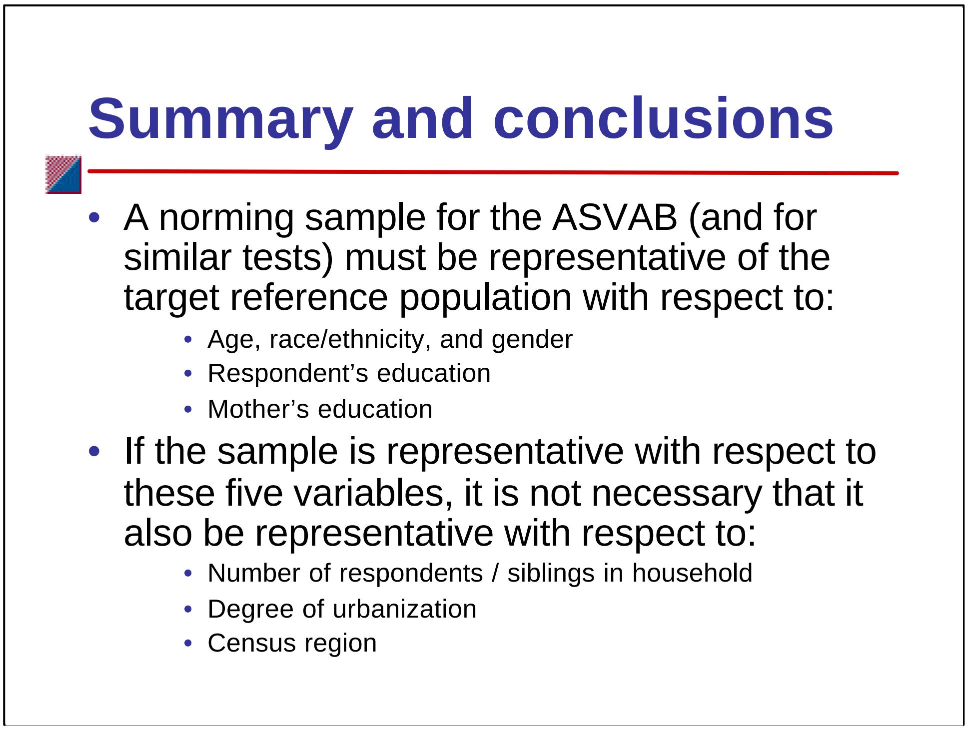
CNA

# On the Representativeness of Norming Samples for Aptitude Tests

31 December 2003 William H. Sims and Catherine M. Hiatt

This paper discusses the extent to which a sample intended for use in norming aptitude scores must be representative of the underlying population.

This document is part of CNA's support to the Defense Manpower Data Center (DMDC) on the National Longitudinal Survey of Youth (NLSY97).




Based on the results described in following slides, we conclude that:

- A norming sample for the Armed Services Vocational Aptitude Battery (ASVAB) (and similar tests) must be representative of the target population with respect to age, race/ethnicity, gender, respondent's education, and mother's education.
- It is not necessary that the sample be representative with respect to number of siblings in the household, degree of urbanization, or census region. Although these factors may be correlated to aptitude test scores, if the five other variables are representative, these factors need not be representative.

## Issue to be addressed

- What demographic variables must be representative of the population in order to have a satisfactory norming sample for aptitude tests?

We address the general question of what variables must be representative of the population in order to have a satisfactory sample of test scores that can be used to norm a test.

Norms for a test describe how a target reference population performs on the test. Therefore, to be useful, the norming sample must be fully representative of the target reference population group on any demographic variable that makes a unique contribution to the variance of test scores.

## Why are representative test norms important?

- If the norming sample is not representative, then:
  - Persons selected on the basis of the test scores may not really have been qualified
  - Persons denied selection on the basis of the test scores may really have been qualified
- Defense community plans to use data from NLSY97 to norm ASVAB

Representative test norms are important to any user of test score information. Users might include schools, employers, government, and the military services.

If the norming sample is not representative of the population of interest, persons selected on the basis of test scores may not really have been qualified. Conversely, persons denied selection on the basis of test scores may really have been qualified.

This issue is of particular importance to the defense community given current plans to use aptitude scores collected during the National Survey of Youth (NLSY97) [1] to produce new norms for ASVAB.

## Approach

- Regression analysis of a nationally representative sample of test scores and demographic information
  - Determine those demographic variables that make *unique contributions* to test score variance

Our approach is to conduct a regression analysis of a nationally representative sample of test scores and demographic information. We will determine those demographic variables that make unique contributions to test score variance.

We stress the phrase "make unique contributions" because it is important to distinguish between the rather large number of variables that are correlated with test scores and that smaller group that uniquely contributes to test score variance. One cannot specify the sample (or develop population weights) on the basis of a very large number of variables because the cell sizes for each combination would be so small that estimates would have large errors.

This work is an extension of our earlier work on the subject [2, 3]. In these earlier reports, we show evidence that age, race, gender, respondent's education, and mother's education are important predictors of test scores. However, these reports were very wide ranging and did not focus on the issue of representativeness of reference or norming populations. In this report, we narrow the focus to the issue of representativeness. We also include additional explanatory variables and develop results for various age and educational subgroups.

## Data

- We will use PAY80 data
  - Persons who were part of NLSY79 who tested on ASVAB in 1980 as part of joint DOD/DOL effort
  - 11,914 cases
  - Will focus on AFQT scores as a measure of general aptitude

We will explore the issue by identifying demographic variables that are correlated with a measure of general aptitude.

We consider the best available sample of nationally representative general aptitude scores to be that collected as part of the Profile of American Youth (PAY) 1980 [4].

The PAY80 sample consists of persons who had participated in the NLSY79 and who agreed to be tested on ASVAB in 1980 as part of a joint effort of the Department of Defense (DOD) and the Department of Labor (DOL). A total of 11,914 persons were tested.

ASVAB contains a measure of general aptitude, known as the Armed Forces Qualification Test (AFQT), along with other tests that measure specific aptitudes.

This analysis will focus on the relationship of AFQT scores to demographic variables. We will assume that variables that correlate with AFQT in 1980 are likely to also correlate in later years.

## Sample/subsample size

| | Size | | |
|---|---|---|---|
| Sample/ subsample | Total tested[1] | All variables present | Case weighted |
| PAY80 | 11,878 | 10,419 | 31,452,444 |
| Age 18-23 | 9,173 | 7,801 | 25,585,172 |
| 4-yr college | 1,512 | 1,428 | 4,990,206 |
| 2-yr college | 742 | 667 | 2,169,072 |
| 12th grade | 1,216 | 1,192 | 3,397,710 |
| 11th grade | 1,277 | 1,256 | 4,061,013 |

1. Excludes 36 cases tested under non-standard conditions.

The PAY80 data set consists of 11,878 participants in NLSY79 who were tested on ASVAB in 1980 under standard conditions. We will examine the full data set and several subsamples made up of various age and educational levels.

An important subsample of PAY80 consists of 9,173 persons age 18-23 during 1980. They were used in developing the current ASVAB score scale (i.e., they were the sample used to norm the test).

The Department of Defense also develops norms for the Student Testing Program (STP) used in many high schools for vocational counseling. We will examine data for 11th and 12th grade students as well as those in 2- and 4-year colleges.

Only those cases with complete demographic information will be used in the regression analysis. This reduces the sample size (as shown in the slide).

# Statistical considerations

- Scale case weights by the design effect to approximate a simple random sample
  - Allows interpretation of standard regression statistics

Standard statistical packages produce statistics under the assumption that the data are from a simple random sample (SRS). Neither the 11,914 raw cases or the case weighted sample (approximately 30,000,000) for the PAY80 sample represent the number of cases in an SRS.

Clustering and oversampling both reduce sampling efficiency, but stratification increases sampling efficiency. All three procedures were used in PAY80 and are routinely used in other large sampling efforts.

The design effect is a factor that expresses the inefficiency of a sample relative to a simple random sample. A sample with a design effect of 1.0 is equivalent to an SRS. A sample with a design effect of 2.0 requires twice as many cases as an SRS to be statistically equivalent to an SRS.

We will scale the sample case weights by the design effect to approximate the size of an equivalent simple random sample. This procedure allows us to interpret the standard regression statistics.

## Scaling case weights

- Design effect
  = 1.441+ (.0005056)*(sample size)[1]
- Effective sample size
  = sample size/design effect
- Scaled case weight
  = (case weight/sum of case weights)*
  (effective sample size)

1. Relationship developed for the PAY80 data set. See [3].

Design effects were computed for PAY80 by the National Opinion Research Center (NORC) [5] for specific race and gender subsets of the data. We must generalize these data for our use with different subsets of the data. We do this by using a simple linear equation. The equation fits the NORC design effects very well, and the procedure is described in [3]. Supporting detail is given in appendix A of this report. The equation is:

Design effect = 1.441 + .0005056* (sample size)

We then use this equation to compute the design effect for our various subsamples and apply the result to estimate the size of an effective simple random sample as shown:

Effective sample size = sample size/design effect

We then scale the case weights of the sample or subsample as:

Scaled case weight = (case weight/sum of case weights)*(effective sample size).

# Calculation of SRS sample size

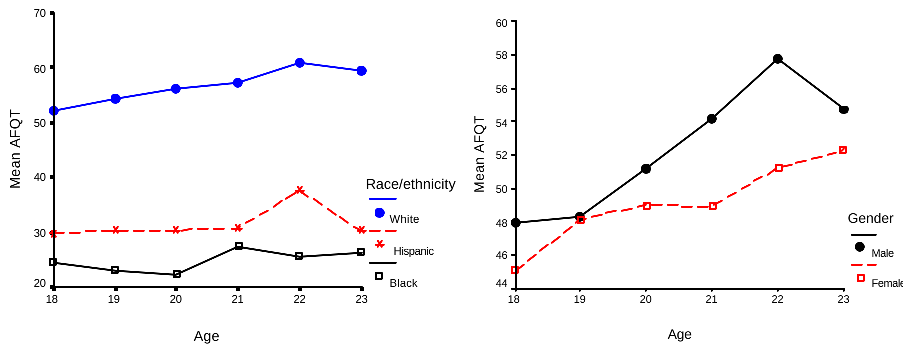
| Sample/ subsample | Cases[1] | Sum of case weights | Design effect[2] | SRS size[3] |
|---|---|---|---|---|
| PAY80 | 10,419 | 31,452,444 | 6.7088 | 1,553 |
| Age 18-23 | 7,801 | 25,585,172 | 5.3852 | 1,449 |
| 4-yr colleges | 1,428 | 4,990,206 | 2.1630 | 660 |
| 2-yr colleges | 667 | 2,169,072 | 1.7782 | 375 |
| 12th grade | 1,192 | 3,397,710 | 2.0437 | 583 |
| 11th grade | 1,256 | 4,061,013 | 2.0760 | 605 |

1. Cases with complete set of regression variables
2. Design effect = 1.441 + .0005056 (cases)
3. Equivalent simple random sample (SRS) size = cases/design effect

In this slide, we show the calculation of the design effect and equivalent simple random sample size for our sample and various subsamples. We used the equations described on the previous slide.

Note that the design effect ranges from 1.7782 to 6.7088 and that SRS sizes are rather modest in comparison to the raw number of cases. We specifically draw the reader's attention to the fact that the 10,419 PAY80 cases (with a complete set of regression variables) are statistically equivalent to an SRS of only 1,553 cases.

## Mean AFQT by age, gender, and race/ethnicity: age 18-23

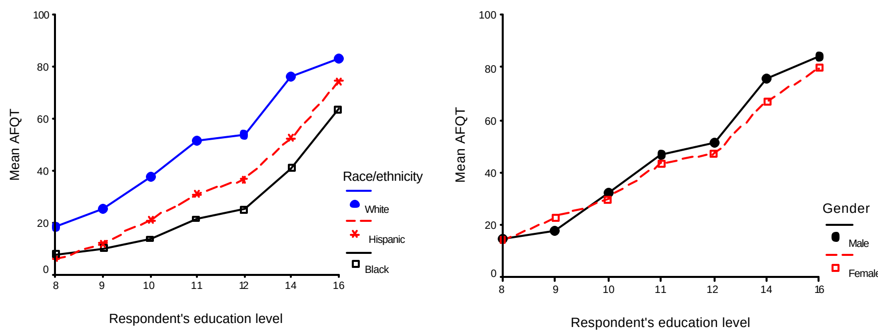



In the next few slides, we examine mean AFQT by various demographic slices in order to better formulate a regression equation. We focus on the age 18-23 subsample because this is the group of most interest to our sponsor. However, the insights gained will also apply to other subsamples in our study.

The left panel shows mean AFQT by age and by race/ethnicity. The data appear to be linear with age and race/ethnicity.

The right panel shows mean AFQT by age and by gender. There is some indication that the slope of AFQT by age may vary with gender. This result suggests that a cross product of age by gender may be appropriate to include in the regression equation.

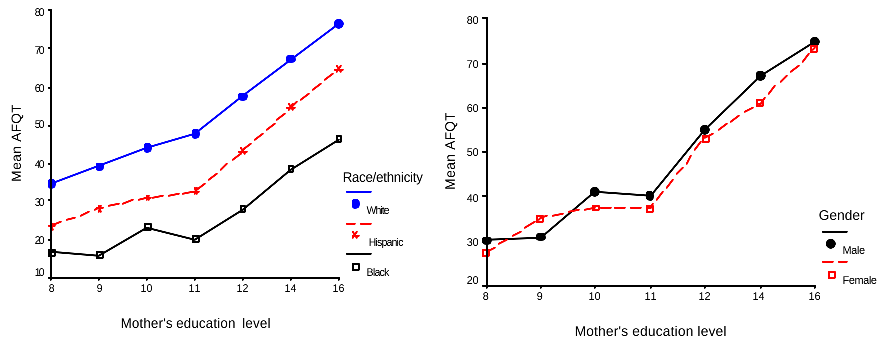
## Mean AFQT by respondent’s education, gender, and race/ethnicity: age 18-23




The left panel shows mean AFQT by respondent’s education level and race/ethnicity. The data are generally linear with respect to age, respondent’s education, and race/ethnicity. However, there is some indication that the slope of the line may differ for some race/ethnicity groups. This suggests that a race/ethnicity cross product with respondent’s education may be appropriate.

The right panel shows mean AFQT by respondent’s education and gender. The data appear to be linear with respect to respondent’s education and gender.

## Mean AFQT by mother's education, gender, and race/ethnicity: age 18-23




The left panel shows mean AFQT by mother's education level and race/ethnicity. The relationship appears to be generally linear.

The right panel shows mean AFQT by mother's education and gender. The relationship appears to be generally linear.

## Regression equation

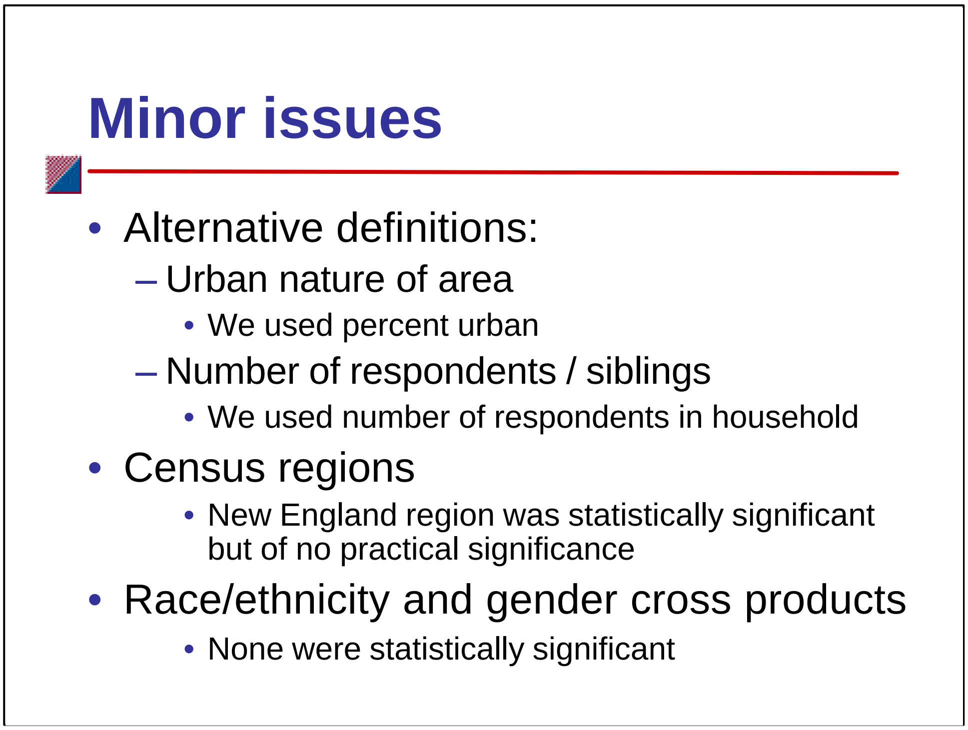
- AFQT = A + B*(age)
  + C*(Black)
  + D*(Hispanic)
  + E*(male)
  + F*(respondent's edu)
  + G*(mother's edu)
  + H*(number of respondent youth in HH)
  + I*(urban / rural)
  + J*(census region)

NOTE: 1. Several alternative measures were used to capture urban/rural and the number of youth in household (HH).
2. Cross terms between race/ethnicity groups, gender, and other variables were also examined in appendix A.

The regression equation will be of the form:

AFQT = A + B* (age)
+ C* (Black)
+ D* (Hispanic)
+ E* (male
+ F* (respondent's education)
+ G* (mother's education)
+ H* (number of respondent youth in household)
+ I* (percentage urban)
+ J* (census region).

Several alternative measures were used to capture the urban/rural nature of the region and the number of youth in the household. We also examined the effect of cross product terms involving race/ethnicity and gender with other demographic variables. These issues are discussed in more detail in the following slide and in appendix B.




In this slide, we discuss and dismiss a number of minor issues. Appendix B contains details of our findings.

We examined several alternative definitions of the urban nature of the residence and the number of siblings.

We chose to use percent urban rather than SMSA categories because it gave a slightly higher $r^2$ contribution in the regression.

We chose to use number of respondents in the household rather than number of siblings because the $r^2$ contributions were very similar and the number of respondents was much more straightforward to calculate.

We included census region as an explanatory variable in all regressions. Only the New England region showed statistical significance. It was of no practical significance, however, as the contribution to $r^2$ was negligible.

Race/ethnicity cross products with other demographics were also included in the regressions. None were found to be statistically significant.

## AFQT regression: PAY80 sample

| Variable | Coefficient | T-stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Constant | -26.0 | -4.6 | **.000** | .183 | .183 |
| Age | -2.6 | -7.7 | **.000** | | |
| Black | -25.0 | -15.1 | **.000** | | |
| Hispanic | -11.7 | -4.9 | **.000** | | |
| Male | 2.8 | 2.5 | **.012** | | |
| Respondent's edu. | 7.9 | 18.2 | **.000** | .384 | .201 |
| Mother's edu. | 3.3 | 12.0 | **.000** | .438 | .054 |
| Youth in household | -1.6 | -2.7 | **.007** | .440 | .002 |
| Urban area | 2.3 | 1.7 | .093 | .440 | .000 |

NOTE: All $r^2$ are adjusted $r^2$ and variables statistically significant at the .05 level are in bold type

This slide summarizes the regression results for the full PAY80 sample. The sample includes persons age 16 to 23 in 1980. These persons were age 15 to 22 in 1979 when the original NLSY79 survey data were collected.

The slide shows the regression coefficients, T-statistics, significance, cumulative adjusted $r^2$, and incremental change in adjusted $r^2$ as the variable, or groups of variables, were entered into the regression.

Age, race/ethnicity, and gender were entered as a group. They are all statistically significant and contribute 0.183 to the $r^2$. Respondent's education is statistically significant and adds 0.201 to the $r^2$, increasing it to 0.384. Mother's education is statistically significant and adds another 0.054 to the $r^2$, increasing it to 0.438. The number of youth in the household is also statistically significant but only adds a negligible 0.002 to the $r^2$. Percentage urban is not statistically significant.

The slide does not include any discussion of census regions or race/ethnicity cross products because they are either not statistically significant or they have a negligible effect on $r^2$. See appendix B for more detail on these issues.

## AFQT regression:
## Age 18-23 subsample

| Variable | Coefficient | T-stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Constant | -40.3 | -5.5 | **.000** | .174 | .174 |
| Age | -1.9 | -5.2 | **.000** | | |
| Black | -25.4 | -14.9 | **.000** | | |
| Hispanic | -12.4 | -5.0 | **.000** | | |
| Male | 3.9 | 3.5 | **.000** | | |
| Respondent's edu. | 8.3 | 20.0 | **.000** | .442 | .248 |
| Mother's edu. | 2.9 | 10.2 | **.000** | .461 | .039 |
| Youth in household | -1.6 | -2.7 | **.006** | .464 | .003 |
| Urban area | 2.1 | 1.5 | .132 | .464 | .000 |

This slide summarizes the regression results for the age 18-23 subsample. These individuals were age 18-23 when they were tested on ASVAB in 1980.

We see that age, race/ethnicity, gender, respondent's education, and mother's education are all statistically significant and make meaningful incremental contributions to $r^2$. The number of youth in the household is statistically significant but does not make a meaningful contribution to $r^2$.

## AFQT regression:
## 4-year college subsample

| Variable | Coefficient | T-stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Constant | 40.4 | 3.6 | **.000** | .252 | .252 |
| Age | 0.4 | 0.9 | .395 | | |
| Black | -29.4 | -12.8 | **.000** | | |
| Hispanic | -12.9 | -3.3 | **.001** | | |
| Male | 5.6 | 4.0 | **.000** | | |
| Respondent's edu. | NA | NA | NA | NA | NA |
| Mother's edu. | 2.2 | 6.5 | **.000** | .297 | .045 |
| Youth in household | -0.1 | -0.2 | .875 | .295 | -.002 |
| Urban area | -0.4 | -0.2 | .815 | .294 | -.001 |

This slide summarizes the regression results for the 4-year college subsample. The persons in this group were in 4-year colleges in 1980 when they were tested on ASVAB.

We see that race/ethnicity, gender, and mother's education are all statistically significant and make meaningful incremental contributions to $r^2$.

Respondent's education was not included in the regression because the subsample was selected on educational level (i.e., those attending a 4-year college).

Age, number of youth in the household, and the urban nature of the area are not statistically significant.

## AFQT regression: 2-year college subsample

| Variable | Coefficient | T-stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Constant | -7.5 | -0.4 | .682 | .248 | .248 |
| Age | 1.9 | 2.4 | **.018** | | |
| Black | -32.4 | -9.2 | **.000** | | |
| Hispanic | -19.0 | -4.2 | **.000** | | |
| Male | 6.9 | 3.1 | **.002** | | |
| Respondent's edu. | NA | NA | NA | NA | NA |
| Mother's edu. | 1.7 | 3.2 | **.002** | .271 | .023 |
| Youth in household | 0.2 | 0.2 | .849 | .270 | -.001 |
| Urban area | 10.1 | 3.1 | **.002** | .286 | .015 |

This slide summarizes the regression results for the 2-year college subsample. The persons in this group were in 2-year colleges in 1980 when they were tested on ASVAB or had been in 2-year colleges the previous year.

We see that age, race/ethnicity, gender, and mother's education are all statistically significant and make meaningful incremental contributions to $r^2$.

Respondent's education was not included in the regression because the subsample was selected on educational level (i.e., those attending a 2-year college).

The number of youth in the household is not statistically significant. Urban area is statistically significant. It contributes 0.015 to $r^2$.

## AFQT regression: 12th grade subsample

| Variable | Coefficient | T-stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Constant | 128.9 | 6.5 | **.000** | .245 | .245 |
| Age | -7.1 | -6.2 | **.000** | | |
| Black | -25.9 | -9.4 | **.000** | | |
| Hispanic | -9.7 | -2.4 | **.016** | | |
| Male | 3.8 | 2.0 | **.044** | | |
| Respondent's edu. | NA | NA | NA | NA | NA |
| Mother's edu. | 3.1 | 7.0 | **.000** | .305 | .060 |
| Youth in household | -0.6 | -0.6 | .556 | .304 | .001 |
| Urban area | 1.9 | 0.9 | .389 | .304 | .000 |

This slide summarizes the regression results for the 12th grade subsample. The persons in this group were expected to enter the 12th grade in the fall of 1980, having been tested on ASVAB during the summer of 1980.

We see that age, race/ethnicity, gender, and mother's education are all statistically significant and make meaningful incremental contributions to $r^2$.

Respondent's education was not included in the regression because the subsample was selected on a specific educational level (i.e., those expected to be in the 12th grade in the fall of 1980).

Number of youth in the household and the urban nature of the area are not statistically significant.

## AFQT regression: 11th grade subsample

| Variable | Coefficient | T-stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Constant | 97.9 | 2.1 | **.035** | .183 | .183 |
| Age | -6.6 | -2.3 | **.020** | | |
| Black | -26.3 | -9.8 | **.000** | | |
| Hispanic | -11.2 | -2.8 | **.005** | | |
| Male | -1.5 | -0.8 | .409 | | |
| Respondent's edu. | NA | NA | NA | NA | NA |
| Mother's edu. | 4.7 | 10.2 | **.000** | .307 | .124 |
| Youth in household | -1.8 | -1.7 | .082 | .309 | .002 |
| Urban area | 4.5 | 2.1 | **.038** | .313 | .004 |

This slide summarizes the regression results for the $11^{th}$ grade subsample. The persons in this group were expected to enter the $11^{th}$ grade in the fall of 1980, having been tested on ASVAB during the summer of 1980.

We see that age, race/ethnicity, and mother's education are all statistically significant and make meaningful incremental contributions to $r^2$.

Respondent's education was not included in the regression because the subsample was selected on a specific educational level (i.e., those expected to be in the $12^{th}$ grade in the fall of 1980).

Number of youth in the household is not statistically significant. Urban area is statistically significant but contributes a negligible amount to $r^2$.

Interestingly, gender is not statistically significant for $11^{th}$ grade, although it was for $12^{th}$ grade. This result suggests that strong gender effects begin to emerge late in high school.

## Summary of regression coefficients

| Sample/ subsample | Age | Black | Hisp | Male | Resp. edu | Mom's edu | Youth/ HH | Urban |
|---|---|---|---|---|---|---|---|---|
| PAY80 | -2.6 | -25.0 | -11.7 | 2.8 | 7.9 | 3.3 | -1.6 | NS |
| Age 18-23 | -1.9 | -25.4 | -12.4 | 3.9 | 8.3 | 2.9 | -1.6 | NS |
| 4-yr col | NS | -29.4 | -12.9 | 5.6 | NA | 2.2 | NS | NS |
| 2-yr col | 1.9 | -32.4 | -19.0 | 6.9 | NA | 1.7 | NS | 10.1 |
| 12th grade | -7.1 | -25.9 | -9.7 | 3.8 | NA | 3.1 | NS | NS |
| 11th grade | -6.6 | -26.3 | -11.2 | NS | NA | 4.7 | NS | 4.5 |

NOTE: NS = not statistically significant at the .05 level, NA = not applicable

Here, we draw together the coefficients from regressions on all samples. For example, one additional year of mother's education is associated with an increase in AFQT of 4.7 percentile points for 11th grade youth. The results are generally consistent, and the trends that emerge appear reasonable.

The coefficient on age is generally negative. This finding is reasonable to expect when respondent's educational level is held constant either by regression (as in the PAY80 sample and age 18-23 subsample) or by selection (as in the other subsamples). Presumably, the older persons in a particular educational group are more likely to have been held back for lack of performance and, hence, would be expected to have lower AFQT scores. The reason for the positive age coefficient for the 2-year college sample is unclear but it does represent persons in the first and second year of college.

Coefficients for race and ethnicity are generally constant over all samples. Males do better than females except for the 11th grade subsample. This finding is consistent with an onset of strong gender differences late in the high school.

Respondent's education is consistently important where applicable. Mother's education is always a factor but seems to be most important in the high school subsamples, particularly in the 11th grade.

The number of youth respondents in the household is statistically significant only for the entire PAY80 sample and for the age 18-23 subsample.

Urban area is statistically significant for 2-year colleges and 11th grade. The lack of consistency over subsamples makes this result somewhat suspect.

## Summary of explained variance ($r^2$)

| | Increment to $r^2$ by indicated variable | | | | | |
|---|---|---|---|---|---|---|
| Sample/ subsample | Age, gender and race/ethnicity | Resp. edu | Mom's edu | Youth/ HH | Urban | Total |
| PAY80 | .183 | .201 | .054 | .002 | .000 | .440 |
| Age 18-23 | .174 | .248 | .039 | .003 | .000 | .464 |
| 4-yr col | .252 | NA | .045 | -.002 | -.001 | .294 |
| 2-yr col | .248 | NA | .023 | -.001 | .015 | .286 |
| 12th grade | .245 | NA | .060 | .001 | .000 | .304 |
| 11th grade | .183 | NA | .124 | .002 | .004 | .313 |

On this slide, we draw together the contribution to explained variance for the sample and subsamples. Again, the results are generally consistent across groups:

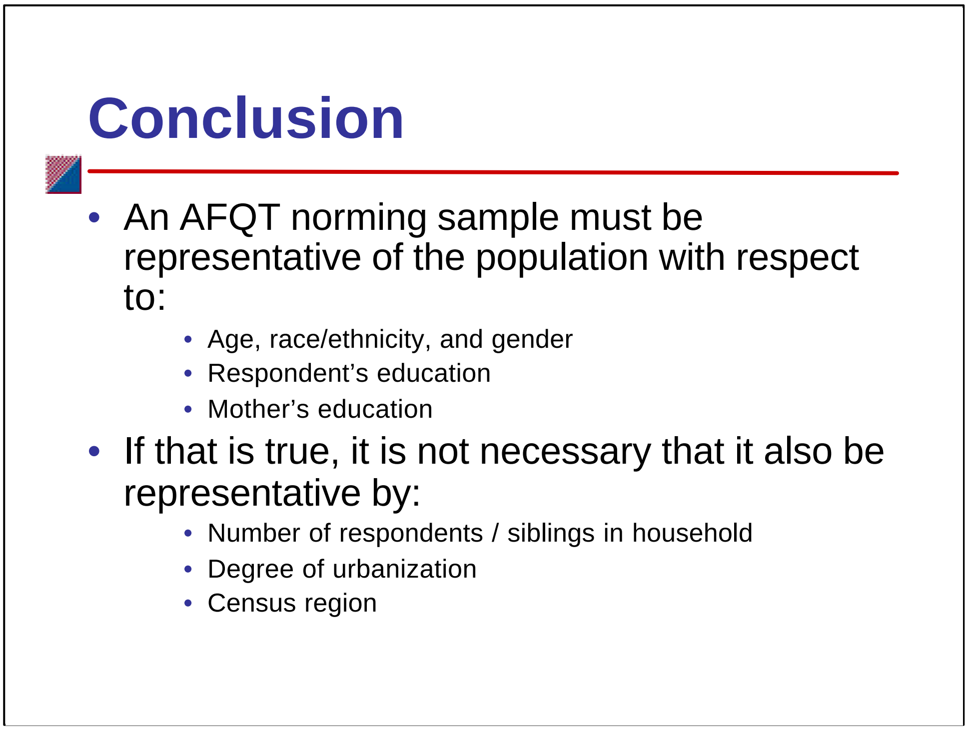
1. The combination of age, gender, and race/ethnicity consistently contributes about 0.2 to the $r^2$.

2. Respondent's education contributes another 0.2 to $r^2$.

3. Mother's education contribution to $r^2$ ranges from a low of 0.023 for 2-year college students to 0.124 for 11th grade students. This variable appears to be more important for high school students than for others.

4. The contribution to $r^2$ by number of respondents per household is consistently negligible.

5. The urban nature of the area makes a negligible contribution to $r^2$ except for 2-year college students. The lack of consistency in this result suggests that it should be viewed with some skepticism.




Based on the results described above, we conclude the following.

An AFQT norming sample must be representative of the target population with respect to age, race, gender, respondent’s education, and mother’s education. Mother’s education is particularly important for high school norms.

If the sample is representative on the five variables noted above, it is not necessary that it also be representative by number of respondents, degree of urbanization, or census region.

# Appendix A: Design effect

In this appendix, we include details on the estimation of design effects for the various subsamples. NORC computed the design effect for the PAY80 sample and for several race and gender subsamples. However, for our analysis, we needed to generalize the design effect to other subsamples.

## What is the design effect?

- It is a factor that expresses the inefficiency of a sample relative to a simple random sample:
  - Clustering reduces sampling efficiency
  - Oversampling reduces sampling efficiency
  - Stratification increases sampling efficiency
- Effective sample size is estimated as:
  - Actual sample size /design effect
- Why do we need to know it?
  - We need it to estimate statistical errors in PAY80

The design effect is a factor that expresses the inefficiency of a sample relative to a simple random sample (SRS). A sample with a design effect of 1.0 is equivalent to an SRS. A sample with a design effect of 2.0 requires twice as many cases as an SRS to be statistically equivalent to an SRS.

Both clustering and oversampling reduce sampling efficiency, but stratification increases sampling efficiency. All three procedures were used in PAY80 and are routinely used in other large sampling efforts.

Effective sample size (i.e., size of an equivalent simple random sample) is the actual sample size divided by the design effect.

The PAY80 data set is based on about 12,000 cases and weighted by case weights to approximate the total youth population of about 30,000,000. Neither the raw number of cases nor the weighted number of cases is appropriate for use in statistical tests because neither represents an SRS (which is assumed by most common statistical packages). For this reason, we must use the design effect to estimate new scaled case weights that will approximate an SRS.

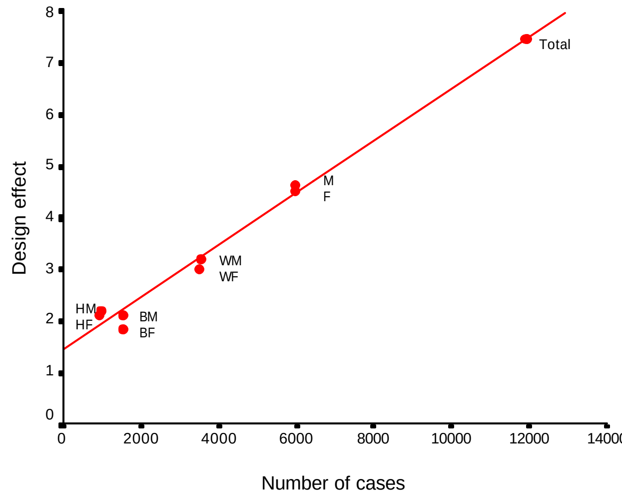
## Design effect for mean AFQT in PAY80[a]

| Gender | Race/ethnicity | Number of cases | Design effect |
|---|---|---|---|
| Male | White | 3,544 | 3.2164 |
| | Black | 1,517 | 1.8253 |
| | Hispanic | 908 | 2.1018 |
| | Subtotal | 5,969 | 4.6307 |
| Female | White | 3,499 | 2.9946 |
| | Black | 1,511 | 2.1147 |
| | Hispanic | 935 | 2.2091 |
| | Subtotal | 5,945 | 4.5057 |
| Total | | 11,914 | 7.4373 |

a. NORC, *Profile of American Youth, User's Guide and Codebook*, March 1982

This slide shows the design effects calculated by NORC [5] for major race and gender subsamples within the PAY80 sample.

## Design effect and sample size: PAY80




This slide shows that the design effects calculated by NORC for the PAY80 sample are approximately linear with sample size. Consequently, we fit the relationship with a simple linear equation as shown on the next slide.

## Regression on PAY80 design effect

| Variable | Coefficient | Standard error | T-statistic | Significance |
|---|---|---|---|---|
| Constant | 1.441 | .117 | 12.275 | .000 |
| Number of cases | .0005056 | .000 | 22.430 | .000 |

NOTE: The $r^2$ for the fit was .99 and the standard error of estimate was 0.23

This slide shows the details of the regression on design effect in PAY80. Based on these results, we will use the following equation to estimate design effects for the various subsamples in our analysis:

$$\text{Design effect} = 1.441 + 0.0005056\ (\text{number of cases})\ .$$

# Appendix B: Statistical detail

This appendix contains backup slides with additional statistical detail.

## Means for main variables: PAY80 sample and subsamples

| Variables | PAY80 | Age 18-23 | 4-yr. col. | 2-yr. col. | 12th grade | 11th grade |
|---|---|---|---|---|---|---|
| AFQT | 48.83 | 51.08 | 76.69 | 60.51 | 47.12 | 42.73 |
| Age | 19.17 | 20.23 | 20.79 | 20.52 | 16.47 | 16.06 |
| Black | .13 | .13 | .10 | .11 | .14 | .14 |
| Hisp. | .06 | .06 | .03 | .07 | .06 | .06 |
| Male | .50 | .49 | .51 | .43 | .51 | .51 |
| Resp. edu. | 11.28 | 11.97 | NA | NA | NA | NA |
| Mom's edu | 11.79 | 11.80 | 13.07 | 12.43 | 12.00 | 11.83 |
| Youth/hh | 1.89 | 1.89 | 1.93 | 1.90 | 1.93 | 1.89 |
| Urban | .78 | .79 | .84 | .88 | .77 | .75 |

This slide shows means for the main variables in the PAY80 sample and various subsamples.

The 11th grade subsample appears to be about 0.5 year older than we would expect.

## Standard deviations for main variables: PAY80 sample and subsamples

| Variables | PAY80 | Age 18-23 | 4-yr. col. | 2-yr. col. | 12th grade | 11th grade |
|---|---|---|---|---|---|---|
| AFQT | 28.87 | 28.95 | 21.06 | 24.58 | 26.82 | 27.12 |
| Age | 2.39 | 1.77 | 1.40 | 1.37 | 0.84 | 0.33 |
| Black | .34 | .34 | .31 | .31 | .35 | .35 |
| Hisp. | .24 | .24 | .18 | .25 | .25 | .24 |
| Male | .50 | .50 | .50 | .50 | .50 | .50 |
| Resp. edu. | 1.91 | 1.66 | NA | NA | NA | NA |
| Mom's edu | 2.19 | 2.19 | 2.11 | 2.09 | 2.20 | 2.09 |
| Youth/hh | .94 | .94 | .92 | .91 | .94 | .92 |
| Urban | .41 | .41 | .37 | .33 | .42 | .43 |

This slide shows the standard deviations of the main variables in the PAY80 sample and subsamples.

Note that the standard deviation for the 11th grade sample is 0.3. This small standard deviation, coupled with the higher than expected mean age shown on the previous slide, suggests that the youngest of the 11th grade youth may be missing.

## Correlation matrix for main variables: age 18-23 subsample

| | AFQT | Age | Black | Hisp. | Male | Mom's edu. | Resp. edu. | Urban | Youth/ hh |
|---|---|---|---|---|---|---|---|---|---|
| AFQT | 1.00 | **.12** | **-.35** | **-.17** | **.05** | **.45** | **.53** | **.06** | **-.06** |
| Age | **.12** | 1.00 | -.03 | -.01 | -.02 | .00 | **.46** | .03 | **-.11** |
| Black | **-.35** | -.03 | 1.00 | **-.10** | -.01 | **-.12** | **-.08** | **.06** | **.08** |
| Hisp. | **-.17** | -.01 | **-.10** | 1.00 | .00 | **-.23** | **-.10** | **.09** | .02 |
| Male | **.05** | -.02 | -.01 | .00 | 1.00 | .03 | **-.05** | -.00 | .04 |
| Mom's edu. | **.45** | .00 | **-.12** | **-.23** | .03 | 1.00 | **.35** | **.11** | .02 |
| Resp. edu. | **.53** | **.46** | **-.08** | **-.10** | **-.05** | **.35** | 1.00 | **.09** | .00 |
| Urban | **.06** | .03 | **.06** | **.09** | -.00 | **.11** | **.09** | 1.00 | **.05** |
| Youth/hh | **-.06** | **-.11** | **.08** | .02 | .04 | .02 | .00 | **.05** | 1.00 |

NOTE: correlations significant at the .05 level are shown in **bold** type.

This slide shows the correlation matrix for the main variables in the age 18-23 subsample. We focus on the age 18-23 group in this and the following slides because it is of most interest to our sponsor. The data for other subsamples are similar.

Those correlations that are significant at the .05 level are shown in bold type.

Mother's education and respondent's education are both strongly correlated with AFQT. Respondent's education is strongly correlated with respondent's age and mother's education. Mother's education is strongly correlated with respondent's education but not with respondent's age. Race/ethnicity also correlates strongly with AFQT.

## $r^2$ by various definitions of urban and siblings: age 18-23 subsample

| | Urban = % urban | | | Urban = SMSA groups | | |
|---|---|---|---|---|---|---|
| Variables | Youth = # sibs | Youth = # resp. | Youth = # resp.18-23 | Youth = # sibs | Youth = # resp. | Youth = # resp. 18-23 |
| Gender, race, age | .189 | .189 | .189 | .188 | .188 | .188 |
| Above + resp. edu | .435 | .435 | .435 | .432 | .432 | .432 |
| Above + mom's edu | .474 | .474 | .474 | .472 | .472 | .472 |
| Above + # youth/hh | .478 | .477 | .478 | .475 | .475 | .475 |
| Above + urban | .479 (NS) | .478 (NS) | .479 (NS) | .477 (NS) | .476 (NS) | .477 (NS) |

Denotes results for variable definitions used in this analysis.
Sample sizes are slightly different from those in the main analy sis.

We estimated the regression equation:

$$AFQT = A + \Sigma_i (B_i X_i) \ ,$$

where A and $B_i$ are constants and $X_i$ are independent variables.

Regression results are shown for six combinations of measures of numbers of respondent youth and urban nature of the region. For number of youth, we use the total number of siblings of all ages, the total number of respondent youth in the survey, and the total number of respondent youth age 18-23. For urban nature, we use the urban / rural designation as well as the four SMSA groups. The four SMSA groups are as follows: not SMSA, SMSA not center city, SMSA center city, and SMSA unknown center city. All combinations gave essentially the same results.

The slide shows cumulative percentage of variance explained ($r^2$) as different variables are added to the regression. At the first stage we include the basic variables of gender, race, and age. We then add respondent's education, then mother's education, then a measure of the number of youth in the household, and finally a measure of the urban nature of the region. All variables were statistically significant at the .05 level except for measures of the urban nature of the region.

We decided to use percentage urban as the measure of urbanizatio n because it is simple to use and gave a slightly larger $r^2$. We decided to use the number of respondent youth in the household as a measure of siblings because it is easiest to calculate.

## Regression results for census regions: age 18-23 subsample

| Variable | Coefficient | T-Stat. | Signif. | Cum $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| Others[1] | NA | NA | NA | .464 | .464 |
| CR Other | 9.9 | 0.6 | .563 | .467 | .003 |
| CR New England | 6.2 | 2.2 | **.028** | | |
| CR East North Central | -0.7 | -0.4 | .699 | | |
| CR West North Central | 1.9 | 0.7 | .456 | | |
| CR South Atlantic | -2.8 | -1.4 | .150 | | |
| CR East South Central | -3.9 | -1.4 | .162 | | |
| CR West South Central | -0.5 | -0.2 | .828 | | |
| CR Mountain | -1.9 | -0.7 | .507 | | |
| CR Pacific | -3.8 | -1.7 | .084 | | |

1. Age, race/ethnicity, gender, respondent's edu, mom's edu, youth/HH, urban

This slide summarizes the effect of adding dummy variables to represent census regions. Census region Mid Atlantic is subsumed in the constant.

The first row shows the cumulative $r^2$ for the main variables of age, race/ethnicity, gender, respondent's education, mother's education, number of respondent youth per household, and percent urban. Other rows show the effect of adding the census region dummy variables.

Only the variable for census region New England was statistically significant. However, all of the census region variables together added only 0.003 to the $r^2$. We consider that effect to be negligible. Census region variables were not included in the final regressions shown in the main text.

## Race/ethnicity and gender cross terms: age 18-23 subsample

| Variables | Coeff. | T-Stat. | Signif. | Cum. $r^2$ | Delta $r^2$ |
|---|---|---|---|---|---|
| All other variables | NA | NA | NA | .467 | .467 |
| Black x age | 0.4 | 0.4 | .711 | | |
| Black x male | -1.6 | -0.5 | .636 | | |
| Black x respondent's. education | -2.3 | -2.0 | .051 | | |
| Black x mom's education | -1.2 | -1.3 | .179 | | |
| Black x urban | -2.8 | -0.6 | .548 | | |
| Black x youth/HH | -1.1 | -0.7 | .489 | .467 | .000 |
| Hisp x age | 0.7 | 0.4 | .657 | | |
| Hisp x male | 0.9 | 0.2 | .845 | | |
| Hisp x respondent's education | -1.4 | -0.9 | .354 | | |
| Hisp x mom's education | 0.1 | 0.1 | .909 | | |
| Hisp x urban | -3.4 | -0.3 | .727 | | |
| Hisp x youth/HH | 0.9 | 0.4 | .689 | | |
| Male x age | 0.2 | 0.3 | .768 | | |

This slide summarizes the effect of adding cross products of race/ethnicity and gender with other demographics. We examined all cross products with race/ethnicity for completeness. However, based on an examination of the data shown in the main text, the only cross product that we considered for gender was age.

The first row shows the cumulative $r^2$ for the regression, including the variables of age, race/ethnicity, gender, respondent's education, mother's education, number of respondent youth in the household, percent urban, and census region.

The other rows show the effect of adding the cross products. None of the cross product terms were statistically significant. However, we note that the cross product of Black with respondent's education was almost statistically significant. Cross product terms were not included in the final regressions reported in the main text.

# References

[1] U. S. Department of Labor, *NLSY97 Users Guide 2001*, 2001

[2] William H. Sims and Catherine M. Hiatt. *Analysis of NLSY97 Test Scores*, Jul 1999 (CNA Annotated Briefing 99-66)

[3] William H. Sims and Catherine M. Hiatt. *Follow-on Analysis of PAY97 Test Scores*, Jul 2001 (CNA Annotated Briefing D0003839.A2)

[4] Office of the Assistant Secretary of Defense (MRA&L). *Profile of American Youth*, Mar 1982

[5] National Opinion Research Center (NORC). *Profile of American Youth, User's Guide and Codebook*, Mar 1982

CAB D0007188.A2/Final

CNA