ADJUSTING EDUCATIONAL SURVEY DATA

Robert C. Larson and Donald T. Searls

National Assessment of Educational Progress
The Education Commission of the States

1. ASSESSING EDUCATIONAL ACHIEVEMENT

When the United States Office of Education was founded in 1867, one charge set before its commissioner was to determine the nation's progress in education. Almost 100 years later, the National Assessment of Educational Progress began charting educational change under the guidance of Dr. Ralph Tyler and Dr. Francis Keppel. The project has now grown to where five years of field assessment have already been completed.

National Assessment is a project of the Education Commission of the States and was established to measure educational achievement. The project's goal is to provide reliable information describing what young Americans (at the ages of nine, thirteen, seventeen and twenty-six to thirty-five) know and can do. Specifically, the assessment is designed (1) to obtain census-like data on the knowledge, skills, and attitudes at regular intervals and (2) to measure the growth or decline in educational attainments.

Presently we have completed a first assessment in Citizenship, Reading, Literature, Music, Social Studies, Mathematics and Career and Occupational Development, and have completed second assessments in Science and Writing. The tasks included in each assessment have been judged important by representative panels of scholars, laypersons and educators and represent things that should be taught in American schools.

The reporting categories National Assessment uses were selected because they reflect groups where differences in achievement occur in the population. The four age levels essentially mark the end of primary, intermediate, secondary and post-secondary education. For each age we report results for groups defined by region, sex, color, level of parents' education and size-and-type of community. Of these, region and sex groups have traditionally shown large differences in educational attainments; schools are thought to vary with the size and type of community (STOC) they serve; and color and level of parental education (PED) are believed to differentiate socio-economic and home and family environments.

An estimate of the percent of people who can perform a task or group of tasks is the basic measure of educational achievement used by National Assessment. The difference between the percentage of a group and the percentage of people in the nation at that age who can perform the task is called the "group effect."

Observed group effects estimate achievement levels of subpopulations such as Southeast or Blacks as they exist in our country. Interpretations of observed group percentages, however, can be misleading in several ways. The fact that a group's relative performance is labeled as a Northeast or Southeast regional "effect" does not mean that differences in these effects occur solely because the respondents live in the Northeast or Southeast. For example, a large fraction of respondents in the Northeast live in large cities while a larger fraction of respondents in the Southeast live in rural areas. Consequently, size and type of community effects may be masquerading as part of an observed regional effect. Similarly, persons whose parents went beyond high school are more frequent in affluent communities than in the country as a whole and persons whose parents had no high school are more frequent in extreme rural communities. In this case parental education effects may be masquerading as size-and-type of community effects.

Confusion about group effects due to masquerading arise when the mixture of characteristics among groups are unbalanced. Since most of the groups are in fact distributed disproportionately in the national population, our weighted probability sample automatically ensures this imbalance in our percentage estimates.

Demands are continually placed on education to find solutions to difficult problems and provide a better education for everyone. At the same time, the pace of changing views and emerging opinion about the nature and the solutions to problems facing education far outdistances the aggregation of supporting data. Unfortunately, this pressure increases the tendency to go beyond the capability of observational data and decreases resistance to making inferences about causes which are at best uncertainties.

Much has been written about the differences between observation and experimentation and the danger of inferring cause from observational studies. Not much has been said, however, about the need to combine skilled observation with incisive analysis in order to generate suggestions for experimental studies. In his lecture on social experimentation, Frederick Mosteller (3) pointed out that "...we need both mechanisms, one to generate suggestions that might lead to improvement, the other to eliminate most suggestions as ineffective."

Analysis and reanalysis of observational data cannot dispel uncertainties about causes, but they can help us gain new perspectives, new hypotheses and perhaps a better understanding of the data. For example, a statistical adjustment which balances the distribution of group characteristics may lead to a better understanding of differences among group effects. This is the basic purpose for performing a "balanced analysis" of National Assessment data.

2. THE BALANCING PROCEDURE

Sample survey data like the kind National Assessment has been collecting over the last five years is frequently adjusted in the sense of forcing sample frequencies or sample ratios to agree with population figures that are known from other sources (See Deming (1)). This is a desirable procedure because presumably the adjusted sample represents the population better and sampling variability of the adjusted frequencies is reduced to some extent.

As a result of this adjustment, the disproportionate distributions are not greatly altered, merely refined. For our purpose, however, we need to go beyond this. It is easier to think about a single marginal group effect if the distribution of the other factors in that group are represented in the same proportions as in the whole age population. Then a direct comparison of any two marginal groups is unconfounded with differing mixtures of these other variables.

Many analyses or data adjustments are possible and specific concerns often direct choices for adjusting data. John Tukey (5) specified a data adjustment procedure which defined marginal effects to be non-sensitive to simple disproportionate distributions of other variables. For reasons which favor reader/listener understandability rather than statistical efficiency Tukey chose to fit an additive linear model to the margins. This led to the "conditions for balance" where the observed number of successes equals the fitted (balanced) number of successes for each marginal group. These conditions for balance were first written by Tukey as follows:

$$(2.1) \quad \Sigma n_{ijk\ell m}(P_{nat} + \Delta P_i + \Delta P_j + \Delta P_k + \Delta P_\ell + \Delta P_m) = \Sigma C_{ijk\ell m}$$

The sum in (2.1) is taken in turn over all indices except one thereby generating a set of 21 equations. Each equation corresponds to a group denoted by one value of the indices $i = 1, \ldots, 4$; $j = 1,2$; $k = 1, \ldots, 7$; $\ell = 1,2,3$; and $m = 1, \ldots, 5$ belonging to Region, sex, STOC, color and PED respectively, and where:

- P_{nat} is the overall national percent correct for the age group;
- ΔP is the balanced group effect corresponding to each value of the indices;
- $N_{ijk\ell m}$ is the weighted number of observations in each 5-way cell; and
- $C_{ijk\ell m}$ is the weighted number of correct responses in each 5-way cell.

A solution to the 21 simultaneous equations gives a set of fitted group effects. These effects were designed so that when added together and with the national percentage, and when multiplied by the actual number of cases they would give a fitted number of successes equal to the observed number of successes. This set of equations is, however, not of full rank and cannot yield unique balanced ΔP's directly. The number of linearly independent equations is 16 but can be increased by appropriately replacing 5 of the 21 equations with the following usual side conditions imposed when an additive linear model is fitted to a multi-way crossed classification

$$\sum_i n_{i\cdots\cdot}\Delta P_i = \sum_j n_{\cdot j\cdots}\Delta P_j = \sum_k n_{\cdot\cdot k\cdot\cdot}\Delta P_k = \sum_\ell n_{\cdots\ell\cdot}\Delta P_\ell = \sum_m n_{\cdots\cdot m}\Delta P_m = 0$$

of the data where the dot "·" notation denotes the sum over the replaced subscript.

A solution of the independent set of equations results in a unique set of balanced group ΔP's. Exhibit 1.1 provides a simple example of the balancing equations and the computation of fitted group effects.

Tukey's balancing equations can be shown to be algebraically equivalent to the usual set of normal equations resulting from minimizing the error sum of squares for a five factor additive linear model

$$(2.2)\quad Y_{ijk\ell m} = \hat{P}_{nat} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\theta}_\ell + \hat{\phi}_m + e_{ijk\ell mr}$$

where $Y_{ijk\ell m}$ is the weighted response for the r-th person in the ijkℓmr-th cell, and α, β, γ, θ, and ϕ correspond to the ΔP's with the same subscript and $e_{ijk\ell mr}$ is random error associated with the r-th person. The solution to the normal equations provides simple least squares estimates of the group effects. If the variability of the percentages in the five-way cell combinations happen to be proportional to the reciprocal of the corresponding cell weights, then these group effects are minimum variance estimates.

Since the original conditions for balance are preserved in the algebraic equivalence, balanced group effects are also least squares estimates.

3. EXAMINING THE RESULTS OF BALANCING

An understanding of balanced group effects can be guided by thinking of a "conceptual" balanced population where the mixture of characteristics in each group is the same as the mixture in the age population. For example, in 1972 approximately 43% of all thirteen-year-olds in the nation had at least one parent who was educated beyond high school. In the Southeast, however, only 29% had at least one parent educated beyond high school. In the "conceptual" thirteen-year-old balanced population, the Southeast region would have effectively the same proportion of children with at least one parent having a post-high school education as the nation.

Consider the result of balancing. If persons having parents with post-high school education do well and are more frequent in one region than in others then we would expect that the balanced effect of that region to be less than its unadjusted effect. In contrast, if persons having parents with no high school education do poorly and are more frequent in another region, then we would expect that balanced effect to appear better than the unadjusted effect. If the magnitude of a group effect is generally reduced by the balancing adjustment, one might conclude that the factor named by the group label itself is not what "accounts" for the observed differences as much as the unbalanced representation of these other variables.

The proportion of an observed group effect that can be attributed to imbalance of other variables can be shown by considering two of the normal equations. First consider the normal equation obtained from (2.2) by summing over all subscripts:

$$(3.1)\quad n_{\cdot\cdot\cdot\cdot\cdot}\hat{P}_{nat} + \sum_i n_{i\cdot\cdot\cdot\cdot}\hat{\alpha} + \sum_j n_{\cdot j\cdot\cdot\cdot}\hat{\beta}_j + \sum_k n_{\cdot\cdot k\cdot\cdot}\hat{\gamma}_k + \sum_\ell n_{\cdot\cdot\cdot\ell\cdot}\hat{\theta}_\ell + \sum_m n_{\cdot\cdot\cdot\cdot m}\hat{\phi}_m = \sum_{ijk\ell m} n_{ijk\ell m} P_{ijk\ell m}\,.$$

Note that the right-hand side is equal to the observed weighted number of successes for an age group. Dividing both sizes by $n_{\cdot\cdot\cdot\cdot\cdot}$

$$(3.2)\quad \hat{P}_{nat} + \sum_i \frac{n_{i\cdot\cdot\cdot\cdot}}{n_{\cdot\cdot\cdot\cdot\cdot}}\hat{\alpha}_i + \sum_j \frac{n_{\cdot j\cdot\cdot\cdot}}{n_{\cdot\cdot\cdot\cdot\cdot}}\hat{\beta}_j + \sum_k \frac{n_{\cdot\cdot k\cdot\cdot}}{n_{\cdot\cdot\cdot\cdot\cdot}}\hat{\gamma}_k + \sum_\ell \frac{n_{\cdot\cdot\cdot\ell\cdot}}{n_{\cdot\cdot\cdot\cdot\cdot}}\hat{\theta}_\ell + \sum_m \frac{n_{\cdot\cdot\cdot\cdot m}}{n_{\cdot\cdot\cdot\cdot\cdot}}\hat{\phi}_m = \frac{\text{wted \# successes}}{\text{wted \# cases}}\,.$$

Since terms 2-6 of the left hand side are all equal to zero, then $\hat{P}_{nat}$ equals the ratio of successes to cases as expected. Note that the ratio of group weights to total weight in these five terms are the marginal group proportions in the observed populations. Now consider a normal equation corresponding to the summation over all subscripts except one, say i, then

$$(3.3)\quad n_{i\cdot\cdot\cdot\cdot}\hat{P}_{nat} + n_{i\cdot\cdot\cdot\cdot}\hat{\alpha}_i + \sum_j n_{ij\cdot\cdot\cdot}\hat{\beta}_j + \sum_k n_{i\cdot k\cdot\cdot}\hat{\gamma}_k + \sum_\ell n_{i\cdot\cdot\ell\cdot}\hat{\theta}_\ell + \sum_m n_{i\cdot\cdot\cdot m}\hat{\phi}_m = \sum_{jk\ell m} n_{ijk\ell m} P_{ijk\ell m}\,.$$

The right hand side is equal to the weighted number of successes in group i, and $n_{i\cdot\cdot\cdot\cdot}$ is the weighted number of cases in group i. Dividing both sides by $n_{i\cdot\cdot\cdot\cdot}$ and subtracting $\hat{P}_{nat}$ gives

$$(3.4)\quad \hat{\alpha}_i + \sum_j \frac{n_{ij\cdot\cdot\cdot}}{n_{i\cdot\cdot\cdot\cdot}}\hat{\beta}_j + \sum_k \frac{n_{i\cdot k\cdot\cdot}}{n_{i\cdot\cdot\cdot\cdot}}\hat{\gamma}_k + \sum_\ell \frac{n_{i\cdot\cdot\ell\cdot}}{n_{i\cdot\cdot\cdot\cdot}}\hat{\theta}_\ell + \sum_m \frac{n_{i\cdot\cdot\cdot m}}{n_{i\cdot\cdot\cdot\cdot}}\hat{\phi}_m = \frac{\text{wted \# successes group i}}{\text{wted \# cases group i}} - P_{nat}\,.$$

The right hand side is the observed

group i effect, and $\hat{\alpha}_i$ is the balanced group i effect. In a balanced population the marginal proportions of the other groups in group i are equal to the corresponding proportions in the population as shown in equation (3.5). That is, the proportion in (3.4) are equal to their corresponding proportions in (3.2) as shown below:

$$(3.5) \quad \frac{n_{ij\cdots}}{n_{i\cdots\cdot}} = \frac{n_{\cdot j\cdots}}{n_{\cdots\cdot}},$$

$$\frac{n_{i\cdot k\cdot\cdot}}{n_{i\cdots\cdot}} = \frac{n_{\cdot\cdot k\cdot\cdot}}{n_{i\cdots\cdot}}, \quad \frac{n_{i\cdot\cdot\ell\cdot}}{n_{\cdots\cdot\cdot}} = \frac{n_{i\cdots m}}{n_{i\cdots\cdot}}$$

$$\text{and} \quad \frac{n_{i\cdots m}}{n_{i\cdots\cdot}} = \frac{n_{\cdots\cdot m}}{n_{\cdots\cdot\cdot}}.$$

If the proportions of groups in subpopulation i are the same as the proportions of groups in the total population, then terms 2-4 of the left hand side in (3.4) are zero and the balanced group i effect $\hat{\alpha}_i$ equals the observed group i effect. To the extent that subpopulation and population proportions in (3.5) differ, the balanced group effect $\hat{\alpha}_i$ will differ from the observed group i effect.

Note that the observed group i effect is equal to the balanced group i effect plus four terms. Each term estimates how much of the balanced effects of one variable are transmitted to the observed group effect i through imbalance of that factor. For example,consider the imbalance of the STOC groups in the Southeast group. If i denotes the Southeast region and k sums over the size-and-type of community variable (STOC), then $\sum_k \frac{n_{i\cdot k\cdot\cdot}}{n_{i\cdots\cdot}} \hat{\gamma}_k$ is the portion of the balanced effects of STOC that is transmitted through the imbalance of the distribution of STOC groups in the Southeast when compared to the distribution in the nation as a whole. Exhibit (3.1) provides an example showing transfers of balanced effects when different variables are included in the balancing equations.

4. LIMITATIONS ON THE INTERPRETATION OF BALANCED EFFECTS

There are several kinds of limitations on the interpretation of balanced results. The group names National Assessment uses for data analysis are labels standing not only for the factor indicated by its name but also for a variety of other factors National Assessment did not (or could not) measure--factors associated or correlated with the named factor. Like observed results, balanced group effects do not show what is caused by the labelled factor. They show only the part of the unadjusted effect that can be conveniently named and attached to a group for bookkeeping purposes. They provide a means of comparing groups of individuals, free of confounding due to various mixtures of other groups.

Balancing of National Assessment data is limited to the five variables assessed. Some important factors may be partially represented in our factors and others not represented at all. Factors may exist which have smaller "proxy" bundles of other factors. If other factors had been included in the balancing equations then the balanced effects and the portion of the balanced effects transmitted though imbalance would be different. The difficulty in interpreting balanced group effects is the same as those encountered in interpreting regression coefficients--that is, the meaning of a balanced group effect depends heavily on what other factors are balanced at the same time. Thus, a balanced group effect represents the influence of unnamed background variables that are still associated with the balanced group name after considering the other known variables included in the balancing set.

Our balancing procedure utilizes an additive model which emphasizes balancing of marginal group effects and ignores balancing on interactions or effects of combinations of groups. For example, the fraction of rural Blacks living in the Southeast is greater than the fraction of rural Blacks living in the Northeast. If rural Blacks living in the Southeast do poorly compared to all rural Blacks, then we would expect the balanced Southeast region to do poorly. Though problems of confounding also exist from disproportionality of combinations of groups, one hopes they exist to a lesser extent. Similar disproportionate representation exists for the other two-, three-, and four-way group interaction effects.

5. SUMMARY AND CONCLUSIONS

The observed group effects from our surveys are facts describing the extent and level of educational performance for subpopulations in our country. Insightful analyses and data adjustments can be helpful in understanding these facts but they cannot change them.

We have shown that balancing is a combination of data adjustment and marginal main effect analysis. Typically, a five dimension data adjustment forces observed sample sizes to known marginal population totals leaving existing disproportionate marginal distributions within a single group pretty near the same. The balanced fit was shown to be equivalent to a least-squares fit of an additive linear model. A unique solution of the corresponding normal equations yields balanced group effects consistent with what one would obtain if the distribution of marginal proportions in each group were the same as the distribution of the marginal groups in the population.

Although a statistically more sophisticated data adjustment could have been employed, we can begin to see how simple adjustments help in removing effects of masquerading and the portion of a balanced group effect that may be transferred to the observed group effect due to imbalance of other characteristics. The greatest limitation is that other important background factors which were not included may still be masquerading as balanced group effects.

For National Assessment (4) the balancing of marginal baseline data is just the beginning. As we obtain repeated measures over time, it will be helpful to know if the proxy bundles maintain their relative importance. As characteristics of the population change, which is expected in our ever changing society, it could be that the "factors" we measure by these variables are changing too. In the future it may be even more important to have well-measured variables and well-conceived data adjustments if we are to go beyond the observed data and obtain some guidance about the mechanisms involving the complex set of factors affecting education. Though the problems in education are far reaching and urgent, one must guard continually against overstating conclusions about the results of sample surveys and data adjustments. The full value of adjusting data depends on careful, clear documentation of the limited steps we have taken, what we have learned, and what we still don't know.

REFERENCES

(1) Deming, W. Edwards, Statistical Adjustment of Data, 2nd publication, New York: Dover Publications Inc., 1964.

(2) Larson, Robert; Wayne Martin; Todd Rogers; Donald Searls; Susan Sherman; and David Wright, A Look at the Analysis of National Assessment Data presented by J. Stanley Ahmann in Frontiers of Educational Measurement and Information Systems-1973, ed. William C. Coffman (Boston: Houghton Mifflin Company, 1973), pp. 89-111.

(3) Mosteller, Frederick, "Social Experimentation," Fifth S.S. Wilks Memorial Lecture, Princeton University, Princeton, NJ., 1973.

(4) National Assessment of Educational Progress, Report 7: Science Group Results B 1970, Washington, D.C.: (U.S. Government Printing Office), 1973.

(5) Tukey, J.W. "Technique for Analysis of Groups". A personal memo to National Assessment staff and the Analysis Advisory Committee, 1970.

EXHIBIT 11, AN EXAMPLE OF THE PROCEDURE FOR OBTAINING A BALANCED FIT.

Consider the simple example of two variables, variable A having 3 levels and variable B having 2 levels. The layout for the number of cases and number of successes is shown below. Note that the number of successes in each cell are to be fitted to the marginals and are left blank.

Number of Cases

		B 1	B 2	
A	1	100	100	200
	2	50	150	200
	3	0	200	200
		150	450	600

Number of Successes

		B 1	B 2	
A	1			80
	2			90
	3			100
		80	190	270

The representation of the fitted number of successes in each cell is shown below where marginal effects are denoted by the letters and the national percentage is 45%.

		B 1	B 2
A	1	$100\ (45\% + a_1 + b_1)$	$100(45\% + a_1 + b_2)$
	2	$50(45\% + a_2 + b_1)$	$150(45\% + a_2 + b_2)$
	3	$0(45\% + a_3 + b_1)$	$200(45\% + a_3 + b_2)$

The balancing equations are formed by combining cell representations of successes and equating to each marginal number of successes in turn.

Cell Combinations	Balancing Equations
$a_1b_1 + a_1b_2$	(1) $100(45\% + a_1 + b_1) + 100(+5\% + a_1 + b_2) = 80$
$a_2b_1 + a_2b_2$	(2) $50(45\% + a_2 + b_1) + 150(45\% + a_2 + b_2) = 90$
$a_3b_1 + a_3b_2$	(3) $0(45\% + a_3 + b_1) + 200(45\% + a_3 + b_2) = 100$
$a_1b_1 + a_2b_1 + a_3b_1$	(4) $100(45\% + a_1 + b_1) + 50(45\% + a_2 + b_1) + 0(45\% + a_3 + b_1) = 80$
$a_1b_2 + a_2b_2 + a_3b_2$	(5) $100(45\% + a_1 + b_2) + 150(45\% + a_2 + b_2) + 200(45\% + a_3 + b_1) = 190$
	Usual Side Conditions
	(6) $200\ a_1 + 200\ a_2 + 200\ a_3 = 0$
	(7) $150\ b_1 + 450\ b_2 = 0$

Only 3 of the set of 5 balancing equations are linearly independent. One way to obtain full rank is to replace equations (3) and (5) by equations (6) and (7). The unique solution for the fitted balanced effects is shown below.

$a_1 = -10\%$ $\quad$ $b_1 = 15\%$

$a_2 = 0\%$ $\quad$ $b_2 = -5\%$

$a_3 = 10\%$

EXHIBIT 3.1a. EXAMPLES FROM THE YEAR 01 SCIENCE ASSESSMENT ILLUSTRATING MEDIAN TRANSFERS OF BALANCED GROUP EFFECTS TO AN UNADJUSTED SOUTHEAST GROUP EFFECT FOR SEVERAL BALANCING COMBINATIONS.

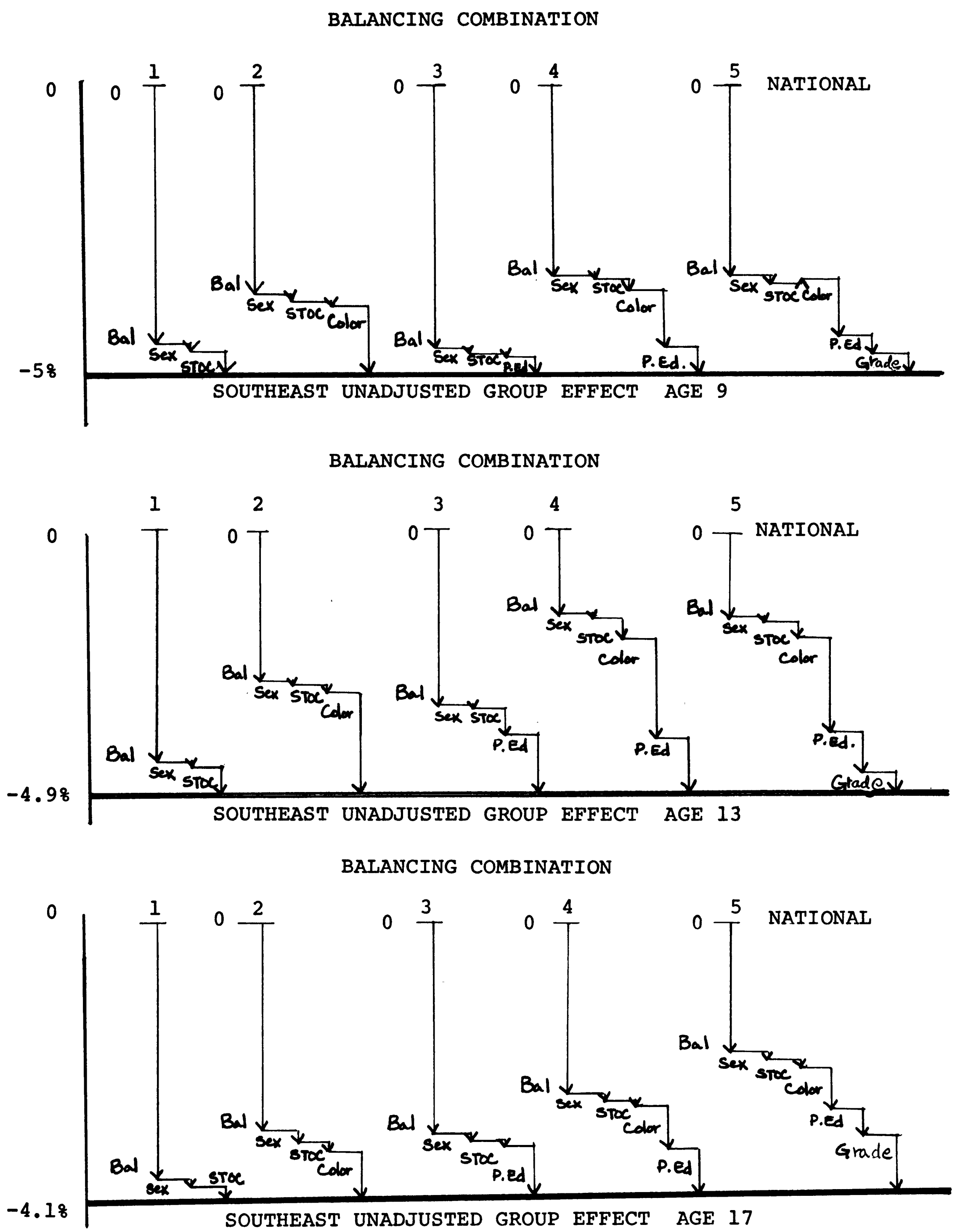
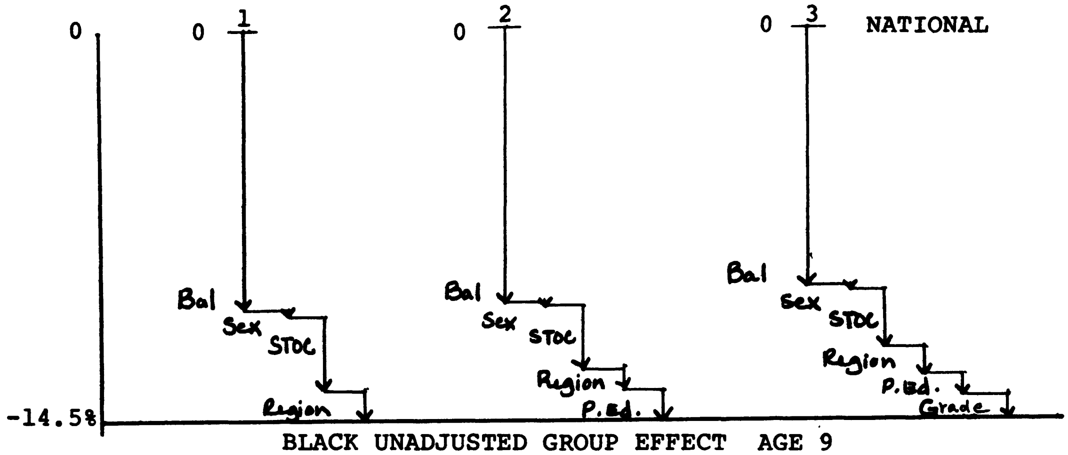


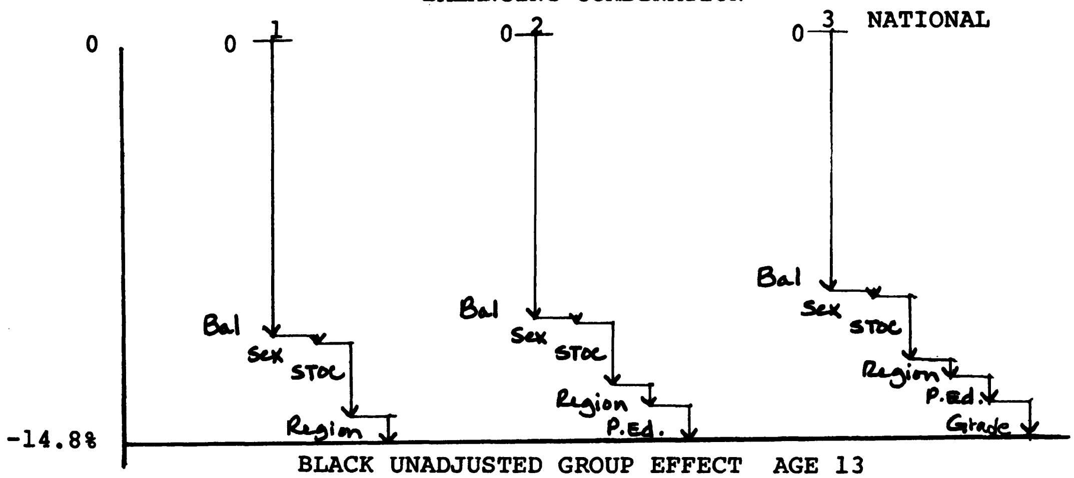
EXHIBIT 3.1b. EXAMPLES FROM THE YEAR 01 SCIENCE ASSESSMENT ILLUSTRATING MEDIAN TRANSFERS OF BALANCED GROUP EFFECTS TO AN UNADJUSTED BLACK GROUP EFFECT FOR SEVERAL BALANCING COMBINATIONS.

BALANCING COMBINATION



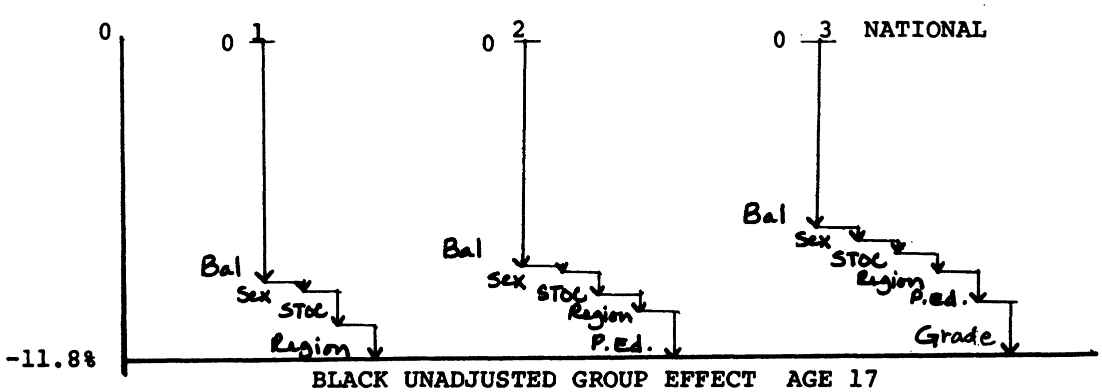
BLACK UNADJUSTED GROUP EFFECT AGE 9

BALANCING COMBINATION



BLACK UNADJUSTED GROUP EFFECT AGE 13

BALANCING COMBINATION



BLACK UNADJUSTED GROUP EFFECT AGE 17