DOCUMENT RESUME

ED 406 187 SE 059 877

AUTHOR	Boone, William J.; And Others
TITLE	Race, Gender, Test Length, and Missing Data: Why Estimates of Performance May Be Clouded.
PUB DATE	22 Mar 97
NOTE	9p.
AVAILABLE FROM	Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (Chicago, IL, March 22, 1997).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Data Analysis; Educational Change; Intermediate Grades; Junior High Schools; Middle Schools; *Minority Groups; *Racial Differences; *Sex Differences; *Test Results
IDENTIFIERS	Middle School Students

ABSTRACT

In 1991, Ohio received National Science Foundation (NSF) funding through its Statewide Systemic Initiative (SSI) program. One aspect of the reform effort involved evaluating the performance of middle school students with a test item bank of items from the National Assessment of Educational Progress (NAEP). This paper presents the results of evaluating these data. It explores how unanswered items can effect analysis of such data when it is used to calculate mean performance measures of groups. How "missing" data can influence calculations of group performance is significant, for if particular subgroups do not complete a test in much higher numbers than other subgroups, it is likely that analyzed data may not reflect reality. Analyzed data showed a great disparity in the percentage of blacks and whites answering the science test items. Noteworthy are black and white students' answering patterns toward the end of the science test. Findings indicate that male and female test takers exhibit some of the same trends as observed in the racial composition. It is concluded that the design of science tests can greatly influence the quality of achievement measures calculated for students. When tests are administered to students, it is critical to evaluate the influence missing data may have upon calculations. (Author/JRH)

Race, Gender, Test Length, and Missing Data. Why Estimates of Performance may be Clouded.

William J. Boone
Indiana University
School of Education, Room 3130
201 N. Rose Avenue
Bloomington, IN 47405

Steven R. Rogg
Miami University
420 McGuffey Hall
Oxford, OH 45056

Jane Butler Kahle
Miami University
420 McGuffey Hall
Oxford, OH 45056

Arta Damnjanovic
Miami University
420 McGuffey Hall
Oxford, OH 45056

Paper presentation at the 1997 Annual National Association for Research in Science Teaching (NARST) Conference in Chicago, March 22, 1997.

Race, gender, test length, and missing data. Why estimates of performance may be clouded.

William J. Boone, Indiana University

Steve Rogg, Jane Butler Kahle, and Arta Damnjanovic, Miami University

In 1991, Ohio received NSF funding through its SSI program. One aspect of the reform effort involved evaluating the performance of middle school students with a test item bank of NAEP items. This paper presents the results of evaluating these data. Specifically, how unanswered items can/can not effect analysis of such data when it is used to calculate mean performance measures of groups. How "missing" data can influence calculations of group performance (e.g. females -vs- males) is significant for if particular subgroups do not complete a test in much higher numbers than other subgroups it is likely that analyzed data may not reflect reality. If missing data does influence calculation of subgroup science performance, what are the implications with regard to the analysis and the construction of science tests? Analyzed data show a great disparity in the percentage of blacks and whites answering the science tests items. Noteworthy are black and white students' answering (and not answering) patterns toward the end of the science test. At the end of the test the disparity between blacks and whites attempting items increases significantly. Male and female test takers exhibit some of the same trends as observed in the racial comparison.

Race, gender, test length and missing data.

Why estimates of performance may be clouded.

Introduction:

In 1991, Ohio was one of the first ten states to receive National Science Foundation funding through its new Statewide Systemic Initiative (SSI) program. Because of Ohio's size and large population, its effort was deliberately restricted to middle school (grades five through nine) science and mathematics. Further, it focused on practicing teachers for whom it provided sustained professional development. Four years into the reform, a study was implemented to describe progress, particularly to assess administrator, teacher, and parent attitudes, teaching practices, and student learning. It attempted to describe the landscape of science and mathematics education in Ohio and, hence, was called the Landscape Study (Kahle and Rogg, 1995). This paper focuses on one component of the collected student data.

Objective:

One aspect of the reform effort carried out during Ohio's SSI was to evaluate the performance of students using a set of well-piloted NAEP items. The objective of this paper is to present the results of evaluating the science test item data. Specifically, how unanswered items can/can not effect analysis of such data when it is used to calculate mean performance measures of groups as a function of race and gender. The issue of how "missing" data can influence calculations of group performance (e.g. females -vs- males) is significant in science education for if one particular subgroup does not complete a test in much higher numbers than other subgroups it is likely that the picture painted with analyzed data may not reflect reality. If missing data does influence calculation of subgroup science performance, then what are the implications with regard to the analysis and the construction of science tests?

Design & Analysis:

In 1996 a 28 item science test was administered to a random sample of 1866 students throughout the state of Ohio. The sample consisted of 520 african-americans and 1346 whites. The breakdown in terms of gender was 1008 females, and 858 males. Following data collection the responses were evaluated utilizing a probabilistic model (Rasch, 1960). This model enabled students performance to be calculated on a linear scale, which allowed parametric tests to be utilized (Wright and Stone, 1979).

Following this analysis, an evaluation of the percentage of students not answering items as a function of race and gender was made. Figure 1 presents the results of the evaluation as a function of gender, while figure 2 presents the data as a function of race.

Findings & Significance:

The data presented in figure 1 are significant for they show that there is a great disparity in the percentage of blacks and whites answering the science tests items. Perhaps most noteworthy are black and white students' answering (and not answering) patterns toward the end of the science test. At the end of the test the disparity between blacks and whites attempting items increases significantly. The data comparing the answering of items by male and female test takers is presented in figure 2. The gender data exhibits some of the same trends as observed in figure 1. However, there are some differences as well. Figure 2 shows that during the early and mid parts of the test there is no significant difference in the percentage of males and females answering/not answering the test items. The difference in test item answering pattern is apparent however, toward the end of the test. Once the answering patterns past item 22 are examined, it becomes apparent that a significantly greater percentage of females do not answer items than the males.

Implications:

There are important implications of the patterns present in figure 1 and 2. Implication #1: When the final items in tests similar in construction and length to this test are very difficult, then the counting of unanswered items as wrong will not greatly effect the overall performance calculated for a group that does not complete the final items of a test. In this case the overall performance of the females would not have been poorly estimated when a comparison was made to males. Implication #2: When the items at the end of a test of similar length and construction are "easy" then the counting of "not answered" items against test takers could greatly effect the performance measure calculated for males and females. In the case in which "easy" items at the end of a test are not answered by females, the net effect is that their performance is underestimated. Clearly, the same situation exists when the performance of african americans is compared to that of whites.

Conclusion:

The design of science tests can greatly influence the quality of achievement measures calculated for students. When tests are administered to students, it is critical to evaluate the influence missing data may have upon calculations.

In many cases the counting of "not attempted items" as wrong on science tests can cause the following problems:

1. Underestimate performance of slower test takers.
2. Overestimate the achievement gap between females-males, whites-african americans.

Although these pitfalls could have effected Ohio's SSI data a key step way taken - statistical tests that did not count missing data as wrong were used. This meant that the test taking pattern of respondents would not influence the achievement measures calculated for each individual student. For others calculating "science achievement", similar methods should be utilized unless test taking strategies are also being evaluated. Also a range of science item difficulties should be present throughout tests. This would mean that neither a preponderance of only easy or only hard items would not be attempted by students, solely because of location within a test.

A second key issue not only has to do with the placement of items throughout a test as a function of difficulty, but also with regard to the distribution of items which might define one of many subscales on a test. If science test items are being used not only for an overall "science" measure, as well as subscale measures - then the items defining the subscale should be evenly placed throughout the test. If this is not done, then one subscale may have items that were not attempted in greater numbers than other subscales.

References:

Kahle, J.B. and Rogg, S.R. 1995. A Pocket Panorama of the Landscape Study. Oxford, OH: Miami University.

Rasch, G. 1960. Probabilistic Models for Some Intelligence Tests. Danmarks paedogogiske Institute. (Reprinted, University of Chicago Press, Chicago, 1980)

Wright, B.D. and Stone, M.H. 1979. Best Test Design. University of Chicago, MESA Press.

C:\discover\narst.97

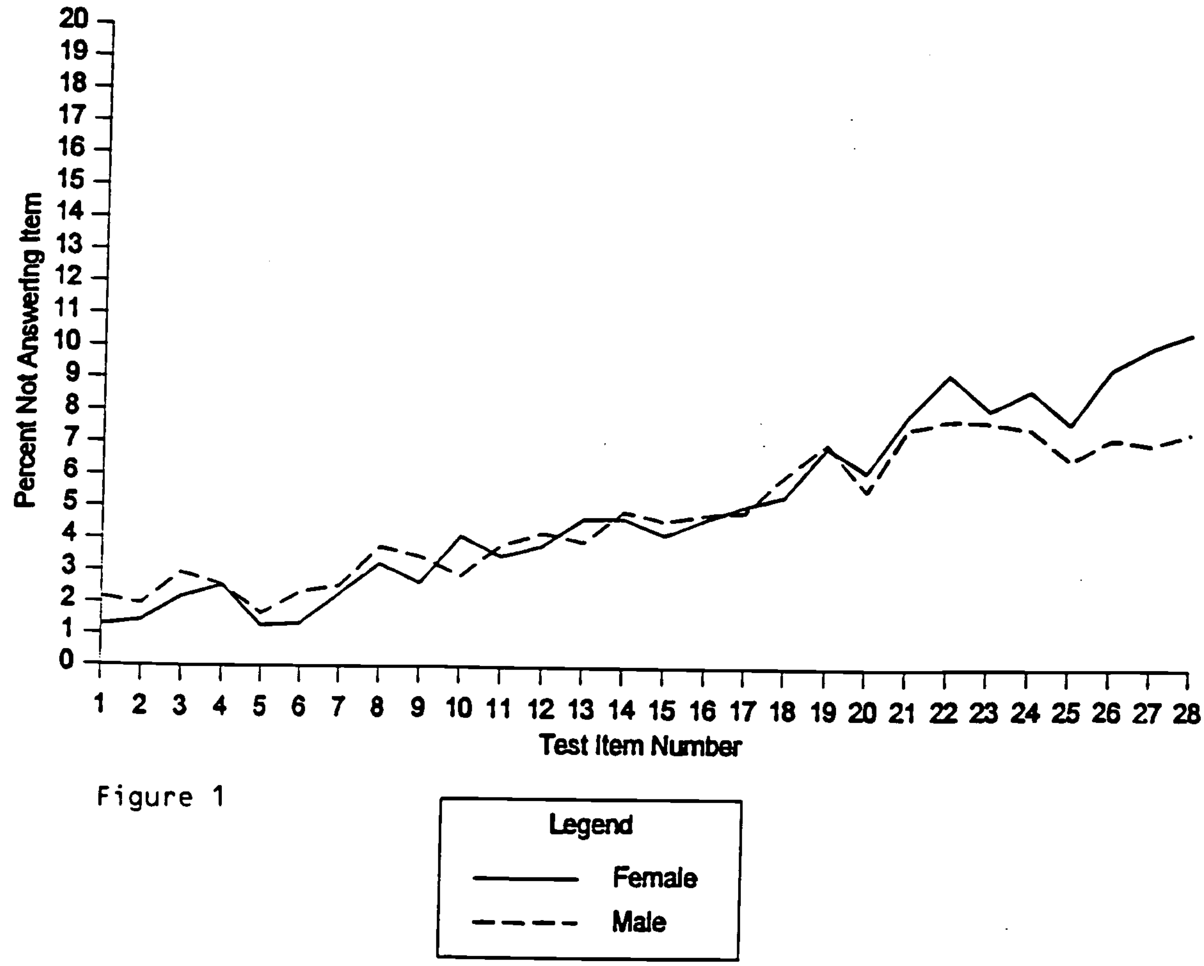

Figure 1

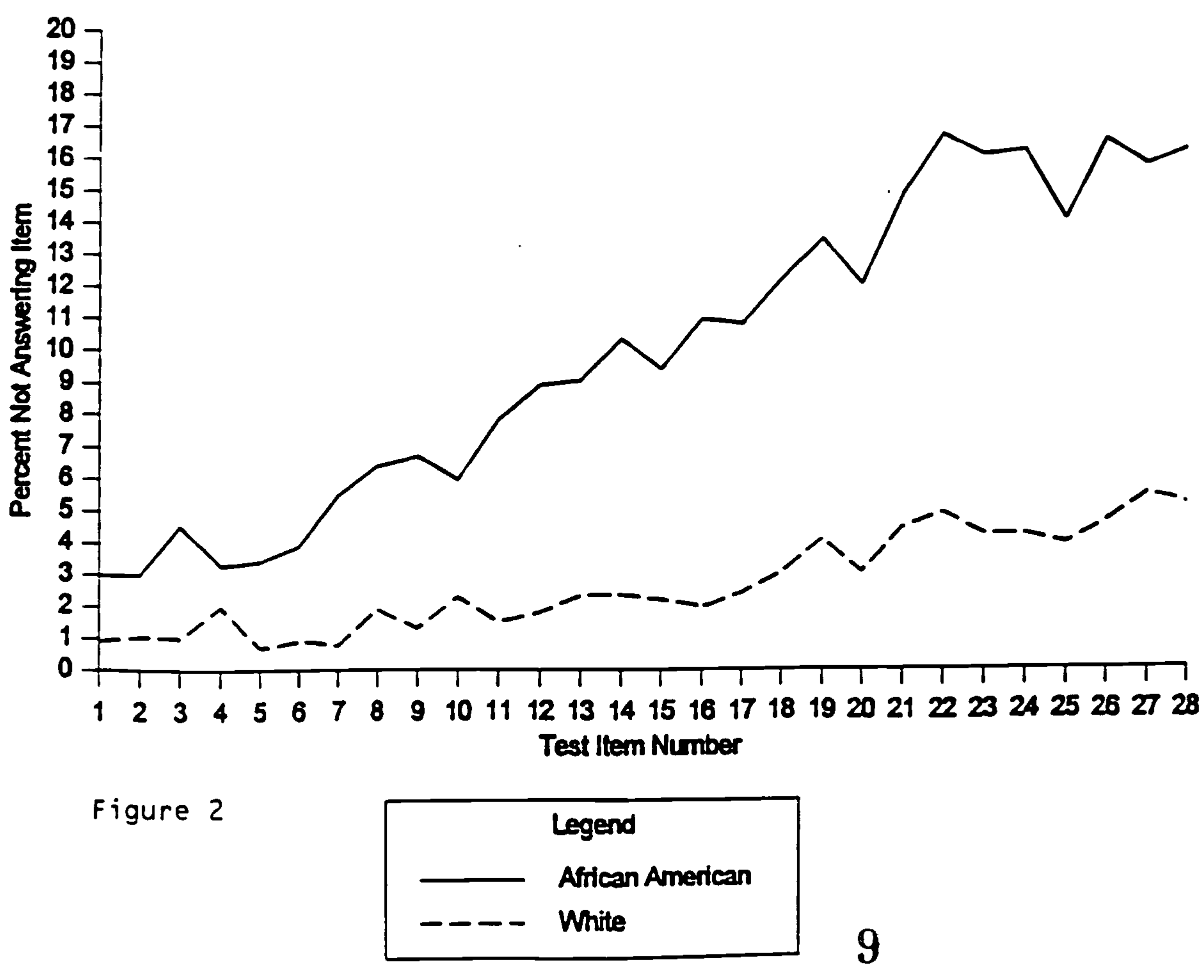

Figure 2

SE059077

ERIC

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Race, Gender, Test Length, and Missing Data. Why Estimates of Performance may be Clouded.

Author(s): William J. Boone, Steven R. Rogg, Jane Butler Kahle, Arta Damnjanovic

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

[X] Check here **For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all Level 2 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] Check here **For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: William J. Boone

Printed Name/Position/Title: William J. Boone, Assistant Professor of Educ.

Organization/Address: Indiana University
School of Education, Rm. 3130
201 N. Rose Avenue
Bloomington, IN 47405

Telephone: 812-856-8132

FAX: 812-856-8440

E-Mail Address: wboone@indiana.edu

Date: 3/27/97

(over)

Share Your Ideas With Colleagues Around the World

Submit your conference papers or other documents to the world's largest education-related database, and let ERIC work for you.

The Educational Resources Information Center (ERIC) is an international resource funded by the U.S. Department of Education. The ERIC database contains over 850,000 records of conference papers, journal articles, books, reports, and non-print materials of interest to educators at all levels. Your manuscripts can be among those indexed and described in the database.

Why submit materials to ERIC?

- **Visibility.** Items included in the ERIC database are announced to educators around the world through over 2,000 organizations receiving the abstract journal, *Resources in Education* (*RIE*); through access to ERIC on CD-ROM at most academic libraries and many local libraries; and through online searches of the database via the Internet or through commercial vendors.

- **Dissemination.** If a reproduction release is provided to the ERIC system, documents included in the database are reproduced on microfiche and distributed to over 900 information centers worldwide. This allows users to preview materials on microfiche readers before purchasing paper copies or originals.

- **Retrievability.** This is probably the most important service ERIC can provide to authors in education. The bibliographic descriptions developed by the ERIC system are retrievable by electronic searching of the database. Thousands of users worldwide regularly search the ERIC database to find materials specifically suitable to a particular research agenda, topic, grade level, curriculum, or educational setting. Users who find materials by searching the ERIC database have particular needs and will likely consider obtaining and using items described in the output obtained from a structured search of the database.

- **Always "In Print."** ERIC maintains a master microfiche from which copies can be made on an "on-demand" basis. This means that documents archived by the ERIC system are constantly available and never go "out of print." Persons requesting material from the original source can always be referred to ERIC, relieving the original producer of an ongoing distribution burden when the stocks of printed copies are exhausted.

So, how do I submit materials?

- Complete and submit the *Reproduction Release* form printed on the reverse side of this page. You have two options when completing this form: If you wish to allow ERIC to make microfiche and paper copies of print materials, check the box on the left side of the page and provide the signature and contact information requested. If you want ERIC to provide only microfiche or digitized copies of print materials, check the box on the right side of the page and provide the requested signature and contact information. If you are submitting non-print items or wish ERIC to only describe and announce your materials, without providing reproductions of any type, please contact ERIC/CSMEE as indicated below and request the complete reproduction release form.

- Submit the completed release form along with two copies of the conference paper or other document being submitted. There must be a separate release form for each item submitted. Mail all materials to the attention of Niqui Beckrum at the address indicated.

For further information, contact...

Niqui Beckrum
Database Coordinator
ERIC/CSMEE
1929 Kenny Road
Columbus, OH 43210-1080

1-800-276-0462
(614) 292-6717
(614) 292-0263 (Fax)
ericse@osu.edu (e-mail)