This is to certify that the

dissertation entitled

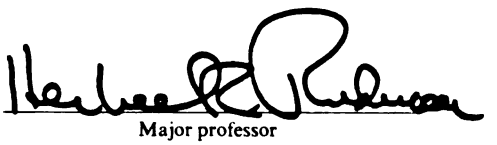
THE EFFECT OF INCREASING TESTING TIME ON THE RESULTS OF THE READING COMPREHENSION AND REFERENCE MATERIALS SUBTESTS OF THE IOWA TEST OF BASIC SKILLS

presented by

Henry G. Dulmage

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Educational Administration



Major professor

Date Dec. 22, 1992

THE EFFECT OF INCREASING TESTING TIME ON THE RESULTS OF THE READING COMPREHENSION AND REFERENCE MATERIALS SUBTESTS OF THE IOWA TEST OF BASIC SKILLS

By

Henry G. Dulmage

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree or

DOCTOR OF PHILOSOPHY

Department of Educational Administration

1992

achie

is e

impro

As t

incre

more

incre

Refere

Skills

altere

minute

random

sixth

distric

Classro

ABSTRACT

THE EFFECT OF INCREASING TESTING TIME ON THE RESULTS OF THE READING COMPREHENSION AND REFERENCE MATERIALS SUBTESTS OF THE IOWA TEST OF BASIC SKILLS

By

Henry G. Dulmage

Speededness is a major concern in standardized achievement tests. The high stakes use of test results is exploding as the United States looks for ways to improve education to compete better in a global economy. As these high stake decisions based on test results increase, the optimal time limit of each subtest becomes more important.

This research was designed to test the effect of increasing time on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills. The time limits for these two subtests were altered in five minute increments, five, ten, and fifteen minutes. The study was conducted using a balance, randomized block design to reduce the error. Twelve sixth grade classrooms, in a middle-sized school district, were divided into three blocks by ability. The classrooms were randomly assigned to one of the

tre

eff

con

dete

sign

five

Comp

Iowa

the

Readi

treat

Compr

speede

Stanfo

Skills

differ

treatments. The results included analyzing for the effects of sex, membership in Chapter I, and family configuration. A regression procedure was used to determine whether the effect of excess time was significant.

The study found that increasing testing time by five, ten, or fifteen minutes on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills does not significantly increase the score, however a block effect was found for both Reading Comprehension and Reference Materials and a treatment by block interaction was found in Reading Comprehension. It also determined that the focus and speededness of the Reading Comprehension subtests of the Stanford Achievement Test and the Iowa Test of Basic Skills are different; they may, in fact, measure different kinds of reading comprehension.

Herb
cont
wit
ext
me
man
per
wit
dis

co
an
He
mo
ma

f
H
e
S
c

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. Herbert C. Rudman, chairman of my committee, for his contributions to this study. His patience and guidance with the specifics of writing a research paper were extremely valuable. He took the time necessary to help me understand how to write and research in a scholarly manner. In addition, he was interested in me as a person, and really cared about the struggles I was having with the writing process. I could not have brought this dissertation to completion without his expert help.

I would like to thank Dr. Steven Raudenbush, committee member, for his help with the statistical analysis and interpretation and presentation of results. He was patient and understanding as he helped me learn more statistical analysis. Dr. Raudenbush has a way of making difficult concepts easier to understand.

To Dr. Frederick Ignatovic, I express my gratitude for his help in reviewing the fine points of this paper. His expertise in research design and writing were extremely valuable in editing this paper. He maintains a good sense of humor which helped me break the tension connected with this process.

see

me

pra

gra

sta

dif

my

wor

com

Lema

grad

devo

lear

profe

Dr. Louis Romano's steady influence and ability to see the big picture are greatly appreciated. This helped me relate my study to practice and gave my research added practical meaning.

In addition, I would like to thank Mr. Rafa Kasam, graduate student in psychometrics, who helped with the statistical analysis of this study. He helped make the difficult seem easy.

Finally, I would like to thank my wife, Marge, and my family for their patience and understanding as I worked toward completion of this study. I could not have completed this research without their support.

I would like to dedicate this study to Mrs. Mildred Leman, now deceased, one of the world's truly great first grade teachers and one of my heroes. Her faith in me and devotion to never saying a child can't learn and won't learn have been a constant source of guidance for my professional career.

LIS

LIS

Cha

II

III.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

Table Page

Fi

4.

LIST OF FIGURES

ach
cons
the
give
educ
stud
loss
to
cons
task
achi
(Hun
publ
have
demo
teste
tests
for t
resul

CHAPTER I

PURPOSE

The concern for speededness in standardized achievement tests has always been an important consideration. Achievement tests that accurately reflect the academic condition of individual students and can be given in an efficient amount of time are crucial to educators. Educators need to be able to demonstrate that students are learning, and learning well, with minimal loss of instructional time. Therefore, the time it takes to administer a battery of tests is an important consideration when choosing an achievement test. Time on task has been shown to be positively related to achievement by the research of Madeline Hunter and others (Hunter, 1978). The time limits established by test publishers and authors reflect the amount of time they have determined is appropriate for a student to demonstrate accurately their knowledge of the material tested. The internal design of standardized achievement tests, then takes into account a sensitivity for the need for timely administration of the test as well as accurate results.

The uses of standardized achievement tests have changed from the original purpose of evaluating a pupil's achievement and/or the condition of the curriculum to include the high stakes decisions of funding programs, granting of diplomas, the rating of teachers, bonuses, and the like. The importance of testing and the time limits prescribed by test publishers for achievement tests have taken on additional importance with the recent movements toward "excellence" in education and accountability. New laws, such as Michigan's P.A. 25, call for student achievement to be measured annually and reported to the public and for planning to be based on measured student outcomes. The promise of funding tied to performance, looms in the near future. In addition, P.A. 25 calls for the establishment of achievement goals for students which, if not reached in a reasonable amount of time, will result in a school's being taken over by the state. The movement toward whole language instruction in reading is causing reading achievement subtests to use longer passages that more accurately assess the comprehension of the learner. Clearly, test results are becoming the basis of more and more decision making at the local district, state, and national policy-making level. The need for accurate measures of student learning, both norm referenced and criteria referenced, are exploding. With this explosion in the number of

tests and uses of tests, the need for efficiency is also paramount.

The time limits set by publishers is important, and so is the accuracy of those time limits. It is valuable to know if established time limits can be altered in relatively minor ways and have a significant impact on individual or group scores or means. If the inaccurate timing of achievement tests by teachers or the inaccurate establishment of test times by publishers has a differential effect on individual or collective outcomes, then more careful administrative practices and test development procedures may be necessary. Classrooms, teachers, or school districts may be eligible for rewards of various kinds due to test results that are unfairly achieved.

It is the purpose of this study to look at the effect of altering the time limits of selected subtests of the Iowa Test of Basic Skills on both individual and class outcomes. This is a replication of two studies done by Rudman and Raudenbush (1986; 1987) using the Stanford Achievement Test. They found a significant effect of testing time on the class mean in Reading Comprehension, but no significant effect on either Word Study Skills or Mathematics Applications subtests when the standardized time limits were exceeded (1986, p. 8; 1987, p. 9). Their results also suggested that the

opti
sugg
Word
of
time
othe
This
as w
rule
limi
The
of t
test
resea
can
at o
addre

truly
Compr
Achie
readi

effect
Test
added

optimal testing time may not necessarily be the time suggested by the test publisher. The time limits of the Word Study Skills, and Mathematics Applications subtests of the Stanford Achievement Test were not shown to be time-sensitive. This could mean that these subtests and others are given more time than is optimally necessary. This may also apply to other standardized test batteries as well. At the least, it brings into question the 90% rule commonly used by test developers when setting time limits for subtests of standardized achievement tests. The 90% rule establishes the testing time as the amount of time it takes for 90% of the test takers to finish the test under power conditions (Nunnally, 1978). Further research needs to be done to determine if these results can be confirmed for subtests of other test batteries and at other grade levels. The important research questions addressed by this dissertation follow:

1. Did the Rudman/Raudenbush research (1986; 1987) truly uncover a time sensitive area in Reading Comprehension, or was it an artifact of the Stanford Achievement Test? Will the Iowa Test of Basic Skills, reading subtest, yield similar results?

2. Will this study show if there is a quadratic effect of excess time on the two subtests of the Iowa Test of Basic Skills used in this study? The effect of added time may increase the scores for some amounts of

tir

inc

Iow

Tes

thi

our

set

bat

wil

stu

fol

tests

(Nunn

to me

time, but that effect may not be sustained for all tested increases in time.

3. Are the reading comprehension subtests of the Iowa Test of Basic Skills and the Stanford Achievement Test equivalent, that is, do they appear to test the same things?

Research Questions

These research questions are directed at furthering our knowledge of the accuracy of current practices in setting time limits for subtests of achievement test batteries. The time limit alterations used in this study will replicate those used in the Rudman/Raudenbush studies. The major research questions explored were the following:

1. Will increasing the time limits on the Reading Comprehension subtest of the Iowa Test of Basic Skills by five, ten, or fifteen minutes significantly increase student achievement?

2. Will increasing the time limits on the Reference Materials subtest of the Iowa Test of Basic Skills by five, ten, or fifteen minutes significantly increase student achievement?

Speededness

Tests can be categorized into speeded tests, power tests, time-power tests, and speed-difficulty tests (Nunnally, 1978, p. 632). Pure power tests are designed to measure knowledge in the absence of time limits, using

to

ite

tes

and

(Nu

ach

adm

clea

admi

limi

the

test

Nunna

(197

Test

limi

upon

the

of t

indi

for

Achie

time

a hig

time

to measure knowledge in the absence of time limits, using items of ordered difficulty, easy to hard. Pure speeded tests measure knowledge under the constraints of time, and the test items are more trivial in difficulty (Nunnally, 1978, p. 631). Standardized tests of achievement are time-power tests due to the necessary administrative functions involved in testing. It is clear that time limits are imposed because of administrative constraints (Kendall, 1964). The time limits are set for most tests by using some variation of the 90% rule, that is, the time is noted when 90% of the test takers are finished under power conditions. Nunnally refers to this as the "comfortable time limit" (1978). The time limits for the Stanford Achievement Test have been established using the 90% rule. The time limits for the Iowa Test of Basic Skills were set based upon a variation of the 90% rule, the authors looked at the percentage of students finishing 75% of the test, 80% of the test, and 100% of the test.

The results of the Rudman/Raudenbush studies indicate the optimal time limit may not have been found for the Reading Comprehension subtest of the Stanford Achievement Test by using the usual method of setting time limits for tests (1986; 1987). Their results showed a highly significant, positive, linear effect of excess time on Reading Comprehension test scores. Perhaps a

lo

ac

Co

If

ver

tes

add

of

the

stud

Boag

kinds

score

are l

1956)

6

Standa

1986)?

7

way to

longer testing time than that used by the publisher would actually be the optimal time limit for the Reading Comprehension subtest of the Stanford Achievement Test. If the results of the Rudman/Raudenbush study can be verified by further research, using different achievement tests, these ancillary questions and suggestions for additional study may be suggested:

1. Are the time limits set for the administration of the subtests on standardized achievement tests really the "optimal time limit" (Kendall, 1964)?

2. Is there a differential effect of time limits on students of differing abilities (Durso & Rubin, 1982; Boag & Neild, 1962)?

3. Are there different kinds of speed for different kinds of tasks (Lord, 1956)?

4. Will students who are capable but work slowly score better than students who work slowly because they are less able (Daly & Stahmann, 1968)?

5. Is there a speed factor in intelligence (Lord, 1956)?

6. How closely should the administration of standardized tests be monitored (Rudman & Raudenbush, 1986)?

7. Is the 90% rule, and its variations, the best way to set test time limits (Rudman & Raudenbush, 1987)?

for

Sta

One

cons

of

diff

the

plac

resu

inve

impor

subte

Compr

in fa

compa

to lo

teste

T

Public

The el

centers

8. Are there in fact different maximal time limits for different subgroups in the population (Daly & Stahmann, 1968)?

This study is designed to be a replication study. One purpose of a replication study is to determine consistency of results. In this case to test the results of the Rudman/Raudenbush (1986; 1987) studies across a different population and different achievement test. If the results are the same, additional confidence can be placed upon the results of the earlier studies; if the results are different, researchers may want to investigate the reasons (Borg, 1983, p. 383).

To determine if the results are the same, it is important to determine if the Reading Comprehension subtest of the Stanford Achievement Test and the Reading Comprehension subtest of the Iowa Test of Basic Skills, in fact, test the same thing. That is, are the two tests comparable? To make that determination, it is necessary to look at the internal makeup of the tests, the content tested, and standardization procedures.

Methodology

The research milieu is a middle sized Michigan Public School system of approximately 7,844 students. The elementary schools are organized around "elementary centers" that have replaced the traditional neighborhood

schoo

appro

one

high

grade

of t

in a

of p

Test

prog

to

conc

gene

for

5, a

12 s

volun

part

spring

who h

Iowa

3.1).

locati

School

school concept. These centers or complexes each house approximately 1,300 students. This system currently has one large high school (1,486 pupils) and one large junior high school (1,591 pupils). The average enrollment per grade is 562 students. The sixth grade is fairly typical of the characteristics of a Midwest middle sized school in all of the major demographics: sex, ethnicity, number of parents, and social class.

The school system used the 1986 edition of the Iowa Test of Basic Skills to monitor the progress of its programs. These tests were given in the fall of the year to provide the district with needed information concerning Chapter I eligibility, the status of the general curriculum, and to provide objective information for developing individual student's programs. Grades 3, 5, and 6 are tested with the Iowa Test of Basic Skills.

This study was based upon the test results from the 12 sixth grade classrooms, in the school system that volunteered to be part of the study. This study was not part of the regular testing program and took place in the spring of 1988. Within those classrooms, the students who had test results available from their fall, 1987, Iowa Test of Basic Skills testing were used (see Table 3.1). The sixth graders, included, were housed in two locations, School 1 had 157 students in 6 sections, and School 2 had 154 students in 6 sections. They were

simi

Tabl

stud

poss

ethn

focu

1978

chil

diff

forma

Erick

covar

Stanf

to a

varia

Raude

(Aid

blocki

T

This

achiev

Skills

similar regarding the other variables of interest (see Table 3.1).

Variables

The variables considered in the Rudman/Raudenbush study (1986; 1987) were replicated, to the extent possible, in this study. These variables--sex, ethnicity, and social class--have continued to be the focus of sociological research (Rehberg & Rosenthal, 1978, p. 4). In addition, it has been noted that children from families of low levels of education and different language and cultural styles do not do well on formal tests of ability and aptitude (Brookover & Erickson, 1975, p. 104).

The Rudman/Raudenbush research found the best covariate to be achievement on the previous years Stanford Achievement Test. No significant relationship to achievement was found for the other demographic variables (Rudman & Raudenbush, 1986, p. 6). Rudman and Raudenbush used the number of students receiving AFDC (Aid to Families of Dependent Children) as a secondary blocking variable.

This research used the same blocking variables. This research found the best covariate to be previous achievement on the same subtest of the Iowa Test of Basic Skills. Membership in Chapter I also proved to be a

sig

dis

12,

Comp

Usag

Mate

Comp

syst

avai

This

and

Basic

repli

group

publis

fiftee

subtes

Th

determi

significant covariate. No significant relationship was discovered for other variables.

Characteristics of the Iowa Test of Basic Skills

The Iowa Test of Basic Skills, 1986 edition, Level 12, is made up of 13 subtests: Vocabulary, Reading Comprehension, Spelling, Capitalization, Punctuation, Usage and Expression, Visual Materials, Reference Materials, Math Concepts, Math Problems, Math Computation, Social Studies, and Science. The school system included in the study administers all the available subtests, except Science and Social Studies. This research dealt with only the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills.

Time Alterations

The time alterations used in this research replicated those used by Rudman/Raudenbush. The control group used the time limits established by the test publisher. The experimental groups had five, ten, or fifteen minutes added to the testing time of each subtest.

Outcomes

The principal objective of this study was to determine if additional time will affect the outcome on

the
of t
sugg
subt
by a
was
Test

to b
the
time
using
this
are
uneq
neede
time

the Reading Comprehension and Reference Materials subtest of the Iowa Test of Basic Skills. Evidence was found to suggest that achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills is not affected by additional time. In addition, no significant effect was found for the Reference Material subtest of the Iowa Test of Basic Skills when additional time was added.

Other subtests, after additional study, may be found to be affected by additional time. If this is true, then the test publishers should look at alternate ways to set time limits for subtests. The standard procedure of using the 90% rule, to set time limits, or variations of this rule, is no longer sophisticated enough when tests are being used for "high stake" decisions or to compare unequal subgroup populations. Additional research is needed to create new and better procedures to establish time limits for all subtests of all achievement tests.

CHAPTER II

REVIEW OF THE LITERATURE

High Stakes Use of Test Results

The accountability movement in education has brought with it a new reliance on testing to demonstrate the effectiveness of educational programming at all levels. Effective education means a positive change in performance or increase in knowledge and a change that can be measured and reported. Schools are rewarded or criticized based upon test results. The stakes are high. The importance of administrative practices in the giving of standardized achievement tests has increased with the "high stakes" decisions that are attached to the outcome data.

Three reports from the U.S. Department of Education--A nation at risk (1983), First lessons (1986), and James Madison secondary school (1986)--all point to standardized test results as an indicator of slipping quality in America's schools and their use as a quality indicator for the future. The use of standardized test results to reward excellence and monitor remedial activities, designed to improve competence, began before

these

these

Educa

summi

goals

The

measu

testi

for f

testi

of sc

save

that

these reports, but they have increased dramatically since these reports were published.

The more recent Education 2000 National Goals for Education, developed at an historic governors' education summit in Charlottesville, Virginia, established six goals for education:

1. By the year 2000 we will bring all children to school ready to learn.

2. By the year 2000 we will raise the graduation rate to 90%.

3. By the year 2000 we will demonstrate student competency in all subjects and prepare students to be responsible citizens, productive workers, and lifelong learners.

4. By the year 2000 we will make the U.S. #1 in math and science.

5. By the year 2000 we will achieve 100% literacy.

6. By the year 2000 we will ensure drug-free, violence-free schools. (National Association of School Boards Bulletin).

The implications for testing are very clear. The measurement for the Goals 2000 will involve student testing at every level and create enormously high stakes for failure to perform.

Madaus (1985) reports, although he disagrees, that testing has been seen as the universal cure for the ills of schooling. It is the source of information that will save the world from illiterate graduates. School reforms that rely on tests to certify students or to assess the

qua

(Sa

tes

tes

poli

Stud

them

notic

compe

educa

begir

or e

quali

gradu

the

upon

"judi

quality of schools have proliferated in recent years (Salganik, 1985).

Several kinds of high stake decisions are based on test results. A review of the literature suggests that tests are used in several ways:

1. Student accountability.
2. Development of state and local educational policies.
3. Comparing schools and school districts.
4. Determining program or activity eligibility.
5. Teacher Certification or advancement.
6. Curriculum change and development.

Student Accountability

Have the students learned what the system has taught them? Should they be given a diploma or promoted? The notion that advancement should be based upon demonstrated competence as shown by test results is very common in education today. Indeed, students' test scores are beginning to be used as the data base for entrance into or exiting from academic programs as well as a major quality control factor in determining readiness for graduation. A nation at risk (1983), while certainly not the initial catalyst for change, has focused attention upon education as perhaps never before and calls for the "judicious use of achievement tests at key transition

poin

Resn

Basi

legi

of e

Sena

acad

call

they

(Pip

requ

next

Educa

state

kinde

estab

norm-

incre

points throughout the educational system (Resnick & Resnick, 1983). Sandifer reports that the South Carolina Basic Skills Assessment Program, established by legislation in 1978, requires that the state department of education:

1. Establish statewide educational objectives in the basic skills with minimum standards of student achievement for readiness and for grades 1, 2, 3, 6, 7, and 11;

2. Select a readiness test to be administered at the beginning of grade 1; and 3; and

3. Select or develop criterion-referenced tests in reading and mathematics to be administered at the end of grades 1, 2, 3, 6, 8, and 11 (Popham, Cruse, Rankin, Sandifer, & Williams, 1985).

Other states have joined the movement. Florida's Senate Bill 6B includes performance standards for each academic course in Grades 9-12 (Pipho, 1986). Arkansas called for holding students at the eighth grade unless they could pass a competency test in the basic skills (Pipho, 1986). Texas banned social promotion and required a grade of 70% for passing from one grade to the next (Pipho, 1986). Georgia's 1985 Quality Basic Education Act addressed the testing of students and stated that a readiness instrument be administered during kindergarten and early first grade, that competencies be established for each student K-12, and that the use of norm-referenced and criterion-referenced tests be increased at all levels (Pipho, 1986). Oklahoma's Reform

Bill
stan
(Phi
loca
Soci
Stud
resu
high

chan
cont
whet
prog
cont
curr
(And
Detr
high
Prof
demo
rece
Mary
Gradu
Writ
1985

Bill of 1985 called for the use of norm-referenced standardized testing for students in grades 3, 7, and 10 (Phipho, 1986). It is now common for states, as well as local districts, to base promotion upon test results. Social promotion is becoming a thing of the past. Students are also being confronted with the use of test results to determine whether or not they graduate from high school.

The focus is on outcomes in current educational change. Unlike earlier reforms that relied on process controls, the new reforms use test scores to determine whether students should receive diplomas or whether program quality is adequate, and rely on "output controls" (Salganik, 1985). Nineteen states are currently implementing tests for high school graduation (Anderson & Pipho, 1984). Rankin reports that the Detroit Public Schools have established requirements for high school graduation called the Detroit High School Proficiency Program. Students in Detroit need to demonstrate mastery of these requirements in order to receive a board-endorsed diploma (Popham et al., 1985). Maryland has used a reading test as a prerequisite for graduation since 1982 and planned to add mathematics and writing in 1987 and citizenship in 1988 (Popham et al., 1985).

that
in T
progr
appro
In a
each
(Poph
for r
& Pip

atten
exami
or c
Resni
the m
(Resn

estab
years
1985)
minim
This
to sta

Remedial programming is another high stakes decision that is being determined by test scores. Senate Bill 350 in Texas directs that the results of the assessment program are to be used to design and implement appropriate compensatory services in grades 3, 4, and 9. In addition each school is to receive an allotment for each disadvantaged student enrolled in the district (Popham et al., 1985). Pennsylvania has mandated a test for remediation purposes for grades 3, 5, and 8 (Anderson & Pipho, 1984).

A related issue, tracking, also has received attention. In England, a complex system of national examinations permit schools to assign students to tracks or classes according to test performance (Resnick & Resnick, 1983). Tracking is also identified as one of the major elements effecting U.S. educational standards (Resnick & Resnick, 1983).

These are growing trends. In 1975 not one state had established minimum competency testing programs; four years later, 37 states had done so (Doyle & Hartle, 1985). Now 40 states are actively pursuing some form of minimum competency testing (Anderson & Pipho, 1984). This high stakes use of test results appears to be here to stay.

De
Lo

is
mar
tha
edu
adm
qua
suc
on
resu
Mich
part
be
Stan
univ
rela
avai
symb
(Mada

Polic
confi
simpli
Qualit

Development of State and Local Educational Policy

The accountability and reform movement in education is using test results to create policy decisions. In mandating tests, policymakers have created the illusion that test performance is synonymous with the quality of education (Madaus, 1985). Politicians and educational administrators have changed the definition of educational quality from process based to outcome based. Examples, such as the MAP (Michigan Accreditation Program), focuses on outcome measures and utilizes standardized test results as a measure of quality. In addition, P.A. 25 in Michigan requires progress to be reported at least in part as standardized test scores and policy decisions to be made by analyzing standardized test scores. Standardized testing, says one critic, is regarded as the universal cure for educational ills because it is relatively inexpensive, well developed, readily available, and administratively simple. Moreover, the symbolic value of tests is attractive to policymakers (Madaus, 1985).

As public confidence in education has eroded, policymakers have looked for ways to revive that confidence. Testing has introduced a deceptive simplicity to the task of restoring both educational quality and public confidence in the schools. Few people

a
e
b
r

re
de
th
te
of
are
dat
to

fro
of
impl
Init
infor
of e
logic
curri
variou
discre
majori
during

are willing to argue with the use of tests as a means of ensuring quality control (Salganik, 1985). Testing has become the "authoritative" evidence that quality is rising (Salganik, 1985).

As school systems across the United States began to react to public pressures in the 60's and 70's, they developed policies related to change and set policies that are based on evidence created through the use of test data. Today state legislatures, state departments of education, local school districts, and test publishers are all working together to bring about more comparative data. Testing is becoming the preferred means of trying to effect change in education (Anderson & Pipho, 1984).

The use of testing to effect policy change has moved from the local to the state and national level. The use of tests has assumed a central role in establishing and implementing state and federal education policy. Initially, the results of local tests were used merely to inform state and federal policymakers about the condition of education. The shift away from local use was a logical outgrowth of the huge federal expenditures for curriculum development in the 1960s. Advocates for various racial and ethnic groups have begun to cite discrepancies in test results between minorities and the majority. The idea of national assessment was promoted during the 1960s so that, according to Ralph Tayler, the

fou

Pro

col

bro

pro

res

lev

of

to

rath

sphe

a m

stan

test

we a

we m

they

admi

must

Compa

schoo

ident

founder of the National Assessment of Educational Progress, "Necessary information might be periodically collected to furnish a basis for public discussion and broader understanding of educational progress and problems." The Coleman report was based largely on the results of a standardized test of verbal ability.

All of these examples show the trend toward state level and national level testing. They represent the use of test scores as a source of information when lobbying to influence within the development of public policy rather than using test results as an influence within the sphere of pedagogical practice (Madaus, 1985).

Tests and examinations have traditionally served as a major means of setting and maintaining educational standards (Resnick & Resnick, 1983). Additional uses of tests demand that they become more and more reliable. If we are to base policy decisions upon test results, then we must be confident that tests measure what they say they measure and that administrative practice in administering tests do not effect the results. Tests must be made as error free as possible.

Comparing School or School Districts

There is a tremendous amount of pressure to compare schools at all levels--local, state, and national. The identification of the best schools is viewed as important

in th
past
favor
Schol
Progr
cause
eleme
Such
an ar

agenc
is g
Tenne
score
state
Pipho
in th
Educa
Chief
endors
expres
establ

Determ
Activi

I
to pr

in the nation's quest for excellence. Terrel Bell, a past Secretary of Education, has gone on record as favoring a 50-state educational ranking system using the Scholastic Aptitude Test and American College Testing Program (Anderson & Pipho, 1984). The political pressure caused by such comparison is thought to be an important element of successful programs (Odden & Anderson, 1986). Such a listing of state-by-state comparisons appears in an article by Stellman and Powell (1985).

Comparisons are being considered by other groups and agencies as well. The Southern Regional Education Board is going ahead with a pilot program in Florida, Tennessee, and Virginia to make minimum competency test scores comparable across state lines and to publish state-by-state comparison data on tests (Anderson & Pipho, 1984). In Michigan, school districts are compared in the newspapers based on the results of the Michigan Educational Assessment Program test. The council of Chief State School Officers voted in November, 1984, to endorse such state-by-state comparisons, but they expressed concern that comparisons be fair and established a group to study the issue (Salganik, 1985).

Determining Program or Activity Eligibility

It has already been noted that test results are used to provide information for tracking students in the

se

co

mo

di

bas

to

equ

thi

pra

the

ano

Sena

funn

othe

Give

comp

of s

(198

an

reme

rule

have

the

durin

secondary schools. Resnick has noted that one way to control educational standards is to use test results to monitor individual student's access to programs and diplomas (Resnick & Resnick, 1983). This movement is based on the idea that a common criterion applied equally to every one is just, since it makes all individuals equal before the law (Popham et al., 1985). However, this does not take into account the administrative practices in giving tests, nor the differing talents of the various subgroups of the population.

Required admission to remedial programming is yet another use of test results. As was noted earlier, Senate Bill 350 in Texas directs that students be funneled into remedial classes (Popham et al., 1985). In other states, such as Michigan, additional funds are given to some districts to establish remedial programs or compensatory programs when they have a significant number of students who score low on assessment tests. Madaus (1985) notes that the placing of children who fall below an artificial cutoff line on a statewide test into remedial programs is one of the problem areas in testing.

The National Collegiate Athletic Association has a rule that a student who attends a Division I school must have a combined verbal and math score of 700 or higher on the Scholastic Aptitude Test to be eligible to play during his freshman year (Madaus, 1985). Local school

di

sir

act

leg

stu

Tea

of

insu

cred

stat

exam

certi

merit

teach

genera

would

T

Career

districts and some states, such as Texas, have taken similar stands and define eligibility for extracurricular activities on academic performance. In New Mexico a 1986 legislation established eligibility requirements for many student programs (Pipho, 1986).

Teacher Certification or Advancement

Student achievement is not alone as a high-stake use of test results. Teachers are also being tested to insure competency before being granted teaching credentials.

Action in this area has been taken in the following states:

Oklahoma--a 1985 reform bill that established an examination in the basic skills for initial teacher certification (Pipho, 1986).

Florida--a 1983 bill that called for a statewide merit pay plan for teachers (Pipho, 1986).

Arkansas--Act 76 created a competency test for teachers that required all practicing teachers to pass a general test of academic skills before their certificates would be renewed (Pipho, 1986).

Tennessee--Senate Bill-1 in 1984 that created a Career Ladder Law (Pipho, 1986).

to

(Pi

pro

tes

suc

for

and

lev

sta

meas

the

succ

test

admi

whic

on a

of t

close

and i

distr

studer

California--Senate Bill-813 included pilot program to reward high schools for improved student achievement (Pipho, 1986).

Michigan--P.A. 267 and P.A. 25--accreditation programs, such as Michigan's MAP program use standardized test results as an output measure in determining the success of a school. Schools must develop annual reports for the public that show current achievement test results and comparisons to past testing. In addition, a two-tier level of teacher competency tests must be passed for state licensure.

Anderson and Pipho (1984) have noted that the use of measured outputs is one of the notions that came out of the accountability movement. Test results mean continued successful operation in some school districts. "What do test results mean to a school board member, administrator, or teacher . . . in a school district in which certification is in jeopardy because of low results on a statewide exam?" (Madaus, 1985). The superintendent of the Detroit Pubic Schools announced that she would close schools that do not measure up to set performance and improvement standards.

It is clear that teachers, students, and school districts have high stakes in the measured outcomes of student achievement. Certification, accreditation, and

inc

and

Cur

cha

sta

the

198

the

proc

curr

dist

spec

Test

curr

comm

the

schoo

state

caref

as we

is de

criter

increased revenues are strong motivators for districts and teachers to have students who perform well on tests.

Curriculum Change and Development

Testing has become a powerful tool for effecting change in education (Resnick & Resnick, 1983). Popham states, "a high stakes test of educational achievement then serves as a powerful curricular magnet" (Popham, 1987). Information to drive curriculum change is one of the oldest uses for educational community. It is a process tool used to determine the effectiveness of curriculum content and teaching methodology. School districts refer to standardized tests to determine what specific changes they should make in their curriculum. Test developers try to design tests that fit the general curriculum of the nation.

Test results are important to the educational community. They are used to make decisions that effect the lives of children and the fate of teachers, schools, school districts, and even the educational well being of states. These high stakes demand that professionals carefully scrutinize all aspects of test administration as well as test validity and curriculum match. Too much is dependent upon the outcome measures of norm and criterion referenced tests.

ach

the:

proc

is

diff

popu

know

perf

test

(Nunn

to m

item

test

and

(Nunn

achi

admi

clea

admi

Sett

under

Speednesses in Test Construction

Speed is an affecting factor in almost all achievement and aptitude tests (Lord, 1956). Speed, therefore, is a major factor in test administration procedures, and in fact, is used to categorize tests. It is a possible factor in fairness as speed may differentially affect various subgroups within the population. Anastasi (1961) states, "It is important to know the extent to which speed and power enter into performance on any particular test."

Tests can be categorized into speeded tests, power tests, time-power tests, and speed-difficulty tests (Nunnally, 1978, p. 632). Pure power tests are designed to measure knowledge in the absence of time limits, using items of ordered difficulty, easy to hard. Pure speeded tests measure knowledge under the constraints of time, and the test items are more trivial in difficulty (Nunnally, 1978, p. 631). Standardized tests of achievement are time-power tests due to the necessary administrative functions involved in testing. It is clear that time limits are imposed because of administrative constraints (Kendall, 1964).

Setting Time Limits

The setting of time limits is not clearly understood.

Som
is
def

var
whe
con
tim
com
lim
test
not
the
1964
time
is n

Spee

speed
have

> Much remains to be learned about speed, in spite of the fact that it is commonly an element in tests and may impact upon test scores. Is speed on cognitive tests a unitary trait? Or are there different kinds of speed for different kinds of tasks? If so, how highly correlated are these different kinds of speed? How highly correlated are speed and level on the same task? How do various criteria relate to speed, and how speeded should tests to predict these criteria be? (Lord, 1956, p. 31).

Some researchers, Thorndike and Rimoldi, have questioned is there is a difference between test takers that can be defined by rate of completion (Daly & Stahmann, 1968).

The time limits are set for most tests by using some variation of the 90% rule, that is, the time is noted when 90% of the test takers are finished under power conditions. Nunnally refers to this as the "comfortable time limit" (1978). In other situations, this comfortable limit is checked by using a preconceived time limit and then checking to see what percentage of the test takers are finished within that time. There does not appear to be any empirical evidence for the use of the 90% figure, it appears to be convention (Kendall, 1964). Kendall (1964) contends that there is a "maximum time limit" for each test, the time beyond which, "there is no further increase in validity."

Speed Factors

Earlier research has investigated several aspects of speed and possible relationships to covariates. Studies have found several speededness factors. All of these

factors are thought to influence the results of achievement and aptitude tests. A 1951 study by French found that there is a perceptual-speed factor, quickly finding the correct answer in the midst of distracting material (Lord, 1956). He breaks this factor down into two additional factors, speed of symbol discrimination, recognizing familiar symbols, and form perception, making sense of unfamiliar symbols (Lord, 1956). Also in his 1951 study, French found these factors to be related to speed; finger dexterity, fluency of expression, ideational fluency, reaction time, speed of association, speed of judgment, and tapping word fluency. In a 1953 study, French noted factors of speed of closure and speed of cognition (Lord, 1956).

Rimoldi, in a 1951 found that speed of judgment, speed of cognition, and a personal tempo are factors related to speededness in test taking (Lord, 1953). Lord refers to a number-speed factor in his 1956 study.

A study done in 1982 by Wild, Durso, and Rubin addressed the following research question relating to the verbal and quantitative experimental sections of the Graduate Record Examinations Aptitude Test (GRE).

> Does increasing the amount of time per question have a differential effect on the score (after controlling for initial ability) of examine subgroups as defined by sex, by race, (Black and White), and by the number of years that have elapsed since the baccalaureate degree was obtained?

Th
of
ad
of
Th
int
sub

fac
fac
fac
He
the
conc
the
admi
(Lor

rese
and
to
abil

These researchers discovered although a larger proportion of examinees completed the experimental tests when given additional time, this extra time does not help any of one of the subgroups more than any other subgroup studied. They suggest that further research is needed on the interaction of ability and tests timing within various subgroups (Wild, Durso, & Rubin, 1982).

Lord, in a 1956 study, looked at a total of 10 factors, four of which were factors of speed. The speed factors were number-speed factor, perceptual-speed factor, verbal-speed factor, and spatial-speed factor. He found that all correlations between course grades and the four speed factor were positive, but not large. He concluded that speed of various kinds plays some part in the course grades studied and that speededness in the admissions examinations is to this extent justified (Lord, 1956).

A 1962 study by Boag and Nield focused on the research question, "Is there a relationship between time and untimed test scores when vocabulary tests are given to secondary school students grouped according to ability. These researchers found that:

1. "Speed and power scores cannot be used interchangeably" as supported by the result that high school students within each group made marked changes in their relative standings when they given additional time.
2. "Changes in relative scores occurred with considerably greater frequency in the regents and

ave

com

suf

It

very

affe

look

Expr

proc

that

the

resu

Engl

test

dete

the

cand

typic

amour

nonregents group (average groups) than in the scholarship (above average) group." The biggest change was in the regents group (average academic track students).
3. "There is a relationship in the results of these scores to the general academic ability of the student" (Boag & Neild, 1962).

They concluded that "under power conditions, the average student, that is the slow and accurate student, comes out nearer the top when given plenty of time but suffers when there is a time limit" (Boag & Neild, 1962). It appears from this study that setting time limits from very speeded condition to a high power condition will affect how at least some subgroups perform on a test.

In 1968 Daly and Stahmann conducted a study that looked at the time limits of the Cooperative English Expression Test that were used as part of the admissions procedure for the University of Utah. The results showed that if 41% of the students were given additional time on the placement test, they would have achieved the same result they achieved after being placed in the remedial English course. This study raised a special subgroup of test takers, "slow working."

A 1972 study by Evans and Reilly was constructed to determine "(1) if the Reading Comprehension section of the Law School Admissions Test (LSAT) is more speeded for candidates from predominantly black colleges than for a typical candidate population, and (2) if reducing the amount of speededness has a differential effect on the

two c
two w
the
admir
the s
Their
speed
speed
group
signi
Reil
this
some
what
meas

cons
limi
cost
time
show

dif
This
tak
adm

two candidate populations." These researchers identified two ways to investigate speededness: (1) to administer the same test and vary the time limits, and (2) to administer tests with different numbers of questions in the same time limit. They chose the second alternative. Their findings were that the LSAT was somewhat more speeded for fee-free candidates, that reducing the speededness increases the test scores somewhat for both groups, and that decreasing speededness is not significantly beneficial to fee-free candidates (Evans & Reilly, 1972). There appears to be some evidence from this study that decreasing speededness does have at least some effect on reading achievement scores, as this is what the subsection of the test used is intended to measure.

Kendall has identified two kinds of time limits when considering the speedness of tests. The optimal time limit is that time for which the ratio of benefits to the cost associated with testing is maximum. The maximal time limit is that time beyond which validity does not show any reliable increase (Kendall, 1964).

This evidence supports the idea that speed may differentially affect subgroups within a population. This effect may show up in the means of different groups taking standardized achievement tests. Therefore, the administration of standardized tests is an important

issue

if re

acros

same

issue. The search for the maximal time limit is crucial if results are to be as accurate as possible and fair across all subpopulations and between groups taking the same test.

CHAPTER III

METHODOLOGY AND INSTRUMENTATION

Sample

This study included the sixth grade classrooms in a middle sized school district that volunteered. The sixth graders included in the sample were housed in two locations: School 1 had 128 students in six sections, while School 2 had 147 students in six sections. They were similar regarding the other variables of interest (see Table 3.1).

Twelve classrooms in two large elementary schools volunteered to be part of the study. A similar condition was true for the Rudman/Raudenbush research. Within the classrooms that volunteered, all students took the test, but only the scores for students who took the Fall 1987 Iowa Test of Basic Skills and had results for at least the Reading Comprehension subtest were used for the study. This accounts for the differences between student counts per building, 157 and 154, respectively; and the sample counts per building, 128 and 147, respectively.

Tabl

Desc

Demo
Vari

1. Pe
in
sa

2. Pe
ex

3. Per
sch
exp

4. Cha

5. Fam
Exp

Samplir

1. Sixt
Incl

Table 3.1

Description of Sample

Demographic Variables	Factors	Relative Frequency
1. Percent of each sex in experimental sample	Male	50.55
	Female	49.45
2. Percent ethnicity in experimental sample	Black	17.09
	Other Minority	6.91
	Total Minority	24.00
	White	76.00
3. Percent minority per school site in experimental sample	Total Minority School 1	26.56
	Total Minority School 2	21.77
4. Chapter I Population	Chapter 1	22.55
	Regular program	77.45
5. Family Structure of Experimental Sample	% of 2 parent families	70.54
	% of 1 parent families	29.45

Sampling Variable		Count
1. Sixth Graders Included in Sample	Enrollment	275
	Buildings	2
	Building 1	128
	Building 2	147

Instrumentation

Characteristics of the Iowa Test of Basic Skills

The Reading Comprehension subtest in Level 12 "consists of selections varying in length from a few sentences to a full page. The passages were chosen in an attempt to represent as completely as possible all of the types of materials encountered by the pupils in their everyday reading" (Hieronymus & Hoover, 1986a, pp. 78-79). The subject areas included are social studies, science, literature, and general information. The authors indicate that there is a gradual ranking of higher level thinking skills as a student moves from level to level. The Reading Comprehension subtest of the Iowa Test of Basic Skills is included in this study to see if any relationship between time and achievement exists, such as that in the Rudman/Raudenbush study (1986; 1987) using the Reading Comprehension subtest of the Stanford Achievement Test.

The Reference Materials subtest of the Iowa Test of Basic Skills measures the following skills: using a dictionary, using an encyclopedia, using general references, alphabetizing, using dictionary guide words, using key words, and using an index (Hieronymus & Hoover, 1986a, p. 85). This subtest was chosen because it tests application skills that are based upon reading and the

lan

to

the

in

cont

add

on va

Setti

t

t

m

c

de

1.

2.

3.

The comp

in Table

Davi

distingui

1.

2.

language arts. These skills are similar to those needed to score well on the Reading Comprehension subtest, and therefore, this subtest may also be sensitive to changes in time limits. It is a subtest that does not parallel content tested by Rudman/Raudenbush (1986; 1987) and may add to an understanding of the effect of increasing time on various subtests of standardized achievement tests.

Setting Time Limits

> An important characteristic of achievement tests is that they should be "power" tests rather than "speeded" tests. Pupils should be given ample time to complete the test in order to provide for a true measure of their skill development. . . . The power characteristics of the Iowa Test of Basic Skills are demonstrated in three ways:
> 1. by the percents of pupils who complete the tests,
> 2. by the percents of pupils who complete 75% of the tests, and
> 3. by the percents of items completed by 80% of the pupils (Hieronymus & Hoover, 1986b, pp. 42-43).

The completion rates of the relevant subtests are given in Table 3.2 (Hieronymus & Hoover, 1986b, p. 50).

A Replicated Experimental Design

David Lykken (Borg, 1983, pp. 383-385) has distinguished three types of replication:

1. Literal replication, or an exact duplication of the sampling procedure, experimental conditions, measuring techniques, and methods of analysis of an earlier study.

2. Operational replication, duplicates only the sampling and experimental procedures.

Table

Speede

Basic

Subtes

Reading

Referen
Skills

*Ma
Basic S
136).

Table 3.2

Speeded Considerations in Standardizing the Iowa Test of Basic Skills.

	Completion Rates Level 12 Spring*		
Subtest	Percent of Pupils Completing Test	Percent of Pupils Completing 75% of Items	Percent of Items Completed by 80% of Pupils
Reading	91	96	100
Reference Skills	97	99	100

*Manual for School Administrators, Iowa Test of Basic Skills, Riverside Publishing Company, 1986, p. 136).

sampl

Rudma

opera

Analyt

race,

progra

These

the inf

Ea

change

assigne

hypothe

3. Constructive replication, uses only a statement of the empirical fact the first investigator claimed to have established, and uses different methods of sampling, measurement, and data analysis.

Since this study was designed to duplicate the sampling procedures and measurement techniques of the Rudman/Raudenbush studies (1986; 1987), it is an operational replication.

Analytical Design

When considering an analytical design an experimenter needs to consider the following points:

1. Will the actual data collected contain all the information needed to make inferences, and can the information be extracted from the data.

2. Can the important hypotheses be tested validly and separately.

3. Will the level of precision reached in estimation, and the power of the statistical tests be satisfactory for the purpose (Hayes, 1981, p. 404).

The data collected included demographic data on sex, race, ability, social status, and membership in Chapter I programming. The testing statistics used was raw score. These data provided all the information necessary to make the inferences required by the hypothesis.

Each hypothesis was represented by an experimental change in time limit. Each experimental time limit was assigned to each block. Therefore, the important hypothesis were accounted for in the design of the study.

assur

confi

pupil

The

the

redu

for

bloc

In

Part

popu

Read

Tes

rec

The researcher used the .05 level of confidence to assure precision in the study. This is the level of confidence usually used in social science research.

The basic analytical design of this study involved pupils nested within block-by-treatment combinations. The blocks were used to increase the statistical power of the experiment by reducing the "error." The error is reduced by partitioning the error into the sum of squares for the blocks and the sum of squares residual, and the block interaction with treatment effect.

> One can often lower within factor level variability of the outcome measures by drawing the experimental units from a homogeneous subpopulation--e.g., persons all the same age, sex, IQ, and socioeconomic level. This may reduce the "error" considerably . . . (Glass & Stanley 1970, p. 491).

In this study the subpopulations were created by partitioning the experimental groups within the population by ability, defined as the group mean on the Reading Comprehension subtest on the previous Fall's Iowa Test of Basic Skills and the proportion of children receiving AFDC. Analysis issues involved were:

1. establishing blocks
2. choosing covariates
3. handling unequal sample sizes
4. choosing an appropriate error term for hypothesis testing

Block

block
score
Test
block
from
by si
was c
was r
were
block
from
on th
study
highe
middl
lowes

assig
numbe
for a
produ
block
are d

Blocking

The participating classrooms were divided into blocks using classroom means on the fall achievement test score on the Reading Comprehension subtest of the Iowa Test of Basic Skills Form H Level 12 as the primary blocking variable. The classrooms were then rank ordered from high to low. They were put into one of three blocks by simply dividing the ranked list. After this ordering was complete, the percentage of children receiving AFDC was reviewed for each classroom. The block assignments were also reviewed using this blocking variable. The blocks were established to be as different as possible from one another on the achievement blocking variable and on the proportion of children receiving AFDC. In this study there were three blocks: one contained the four highest achieving classrooms, one contained the four middle achieving classrooms, and one contained the four lowest achieving classrooms.

The classrooms within each block were then randomly assigned to treatments by assigning each classroom a number and then using a published list of random numbers for assignment. The resulting experimental procedure produced a balanced randomized block design with three blocks and four treatments within each block. The groups are described in Table 3.3.

Table 3.3

Experimental Group Descriptive Statistics by Block.

BLOCK	CLASS	TREATMENT	COUNT	% of Other Non-White	% Male	% of AFDC	% of Chap	% of one Prnt	Pre-Read Mean Raw Score	Pre-Read Mean Adjusted
1	1	5	26	0.0	46.2	3.8	0.0	0.0	38.84	44.88
1	2	15	27	29.6	70.4	7.4	11.1	22.2	34.78	42.74
1	3	0	24	16.7	54.2	33.3	0.0	4.2	33.17	37.33
1	4	10	29	20.7	44.8	20.7	0.0	17.2	34.17	39.48
MEAN			26.5	16.7	53.9	16.3	2.8	10.9	35.29	41.15
2	5	10	25	28.0	52.0	40.0	24.0	36.0	30.20	34.04
2	6	0	23	30.4	43.5	56.5	17.4	52.2	27.83	27.87
2	7	15	24	28.0	40.0	32.0	48.0	32.0	28.92	32.68
2	8	5	23	27.3	59.1	40.9	18.2	40.9	27.04	34.00
MEAN			23.8	28.4	48.7	42.2	26.9	40.3	28.54	32.18
3	9	15	15	33.3	60.0	93.3	0.0	26.7	25.40	28.53
3	10	5	14	35.7	71.4	100.0	78.6	28.0	19.43	19.26
3	11	10	22	18.2	36.4	72.7	40.9	63.6	26.73	34.55
3	12	0	23	21.7	39.1	60.9	34.8	43.5	26.30	37.26
MEAN			18.5	27.2	51.7	81.7	38.6	40.7	24.95	31.26

diffe
perce
stude
progr
and
demog
progra
especi
abilit
childr
abilit
for gi
done,
advanta
substar
from
familie
educati
groupin
ethnic,
(Brookov
reflecte
advantag
This was
and 15 s

The blocks are similar in percent of males. The differences are greater in number of students per block, percent of non-White students per group, percent of students on AFDC, percent of students in the Chapter I program, percent of students from one-parent families, and pretest mean. Group mean differences in the demographic variables of AFDC, membership in Chapter I programming, and students from one-parent families were especially large. These differences suggest that the top ability classrooms contained markedly more advantaged children. The classrooms reflect an ordering of rooms by ability within the school system that is done to program for gifted and talented students. When this ordering is done, related social factors also are ordered from advantaged to less advantages. This finding substantiates an observation by Brookover, "Since people from minority groups and lower-socioeconomic status families are more likely to be placed in lower educational level groups or tracks, the extensive grouping practice is clearly related to the racial, ethnic, and socioeconomic stratification of the society" (Brookover, 1975, p. 130). Class size differences reflected the opposite attention, that is, the least advantaged classes had the least number of students. This was especially true of the two smallest classes (14 and 15 students). The smallest class also had a large

differ
fairly
and p
statis
blocks
less a
W
treatm
statis
statis
smalle
variab
betwee
studen
non-wh
half,
by a
treatm
group;
maximu
to 44.
doublec
experim
close
proport
15.6%,

difference in raw score per test mean. The fact that fairly large differences in number of students per block and per treatment existed was taken into account in the statistical process used in analyzing the data. The blocks are arranged from advantaged and high achieving to less advantaged and lower achieving.

Were there any apparent differences between the treatment groups? Table 3.4 shows the same descriptive statistics arranged by treatment. The descriptive statistics arranged by experimental group showed somewhat smaller differences between treatment groups on the variables included in the study than the differences between blocks. Class size varied a maximum of 4.3 students and a minimum of two students; the proportion of non-White students varied from 30.3% to 16.6% or almost half, a large difference; the proportion of males varied by a maximum of 14.5% from 58.9% in the five-minute treatment group to 44.4% in the ten-minute treatment group; the proportion of AFDC children varied by a maximum of 6% and was highest in the control group (50.2% to 44.2%), the proportion of Chapter I students almost doubled between the control group and the five-minute experimental group, 17.4% to 31.9%, the other groups were close to the five-minute experimental group; the proportion of children from one-parent families varied by 15.6%, 38.9%, to 23.3%. The spread of differences on the

Tabl

Desc

T	C
R	I
E	A
A	S
T	S
Contr	
3	
6	
12	
MEANS	
Five M	
1	
8	
10	
MEANS	
Ten Mi	
4	
5	
11	
MEANS	
Fifteen	
2	
7	
9	
MEANS	

Table 3.4

Descriptive Statistics by Treatment Group.

TREATS	CLASS	COUNT	% Other Non-White	% Male	% of AFCD	% of Chap.	% of One Parent Homes	Mean Pre-Test Raw Score	Mean Pre-Test Adjusted
Control									
	3	24	16.7	54.2	33.3	0.0	4.2	33.17	37.33
	6	23	30.4	43.5	56.5	17.4	52.2	27.83	27.87
	12	23	21.7	39.1	60.9	34.8	43.5	26.30	37.26
MEANS		23.3	22.9	45.6	50.2	17.4	33.3	29.16	
Five Minutes									
	1	26	0.0	46.2	3.8	0.0	0.0	38.84	44.88
	8	23	27.3	59.1	40.9	18.2	40.9	27.04	34.00
	10	14	35.7	71.4	100.0	78.6	28.9	19.43	19.26
MEANS		21.0	21.0	58.9	48.2	31.9	23.3	30.30	
Ten Minutes									
	4	29	20.7	44.8	20.7	0.0	17.2	34.17	39.48
	5	25	28.0	52.0	40.0	24.0	36.0	30.20	34.04
	11	22	18.2	36.4	72.7	40.9	63.6	26.73	34.55
MEANS		25.3	16.6	44.4	44.4	21.6	38.9	30.71	
Fifteen Minutes									
	2	27	29.6	70.4	7.4	11.1	22.2	34.78	42.74
	7	24	28.0	40.0	32.0	48.0	32.0	28.92	32.68
	9	15	33.3	60.0	93.3	0.0	26.7	25.40	28.53
MEANS		22.0	30.3	56.8	44.2	19.7	26.9	30.52	

achiev

score

betwee

of stu

in the

childr

highes

males.

anothe

differ

on th

experi

review

establ

Best C

single

pretest

Referen

covaria

achievement variable was small, a maximum of 1.55 raw score points. The differences in variables was spread between the experimental groups.

1. The control group was highest in the proportion of students on AFDC.

2. The five-minute experimental group was highest in the proportion of student in Chapter I.

3. The ten-minute experimental group was highest in children from one-parent families.

4. The fifteen-minute experimental group was highest in the proportion of non-White children and males.

The blocks were significantly different from one another and the experimental groups, while showing differences on most variables are similar to one another on the achievement variable. The random selection of experimental groups was verified. This completes the review of the demographic data concerning the establishment of blocks and treatment groups.

Best Covariate

For the Reading Comprehension subtest, the best single covariate proved to be the Reading Comprehension pretest, r=.80. For the Reference Materials subtest, the Reference Materials pretest proved to be the best covariate, r=.75 (see Table 3.5). The covariates (sex,

Tabl

Pear

Prere
Compr

Prere
Skill

family

prove

signi

the c

and

subte

Readi

the W

covar

reduce

Table 3.5

Pearson Correlation Coefficients

	Post Reading Comprehension	Post Reference Skills
Prereading Comprehension	.80 (N=275) P=.000	.66 (N=275) P=.000
Prereference Skills	.71 (N=275) P=.000	.75 (N=275) P=.000

family configuration, ethnic group, membership) did not prove to be significant. Membership in Chapter I was significant for Reading Comprehension. The analysis of the covariates is shown in Table 4.2.

This result agreed with the findings of the Rudman and Raudenbush research, "for Reading Comprehension subtest, the best single covariate proved to be the Total Reading pretest . . . for the Word Study Skills subtest, the Word Study Skill pretest proved to be the best single covariate" (Rudman & Raudenbush, 1986, p. 6).

An ANOVA using regression techniques was employed to reduce the error associated with unequal sample sizes.

Cho

Dete

were

Raude

relat
and
resea

T
member
were i
have c
(Rehbe
been n
educati

Choosing the Appropriate F-Test

> In nested designs of this type, the appropriate F test for the treatment contrasts typically uses the unexplained variation between classes as the mean square error. In such an analysis, the residual effects of classrooms are viewed as random effects. When the null hypothesis of no residual variance between classrooms is retained, an alternative error term is available, and typically provides a more powerful F-test. The alternative is to pool the residual between-class variation with the residual within-class variation, yielding a dramatic increase in the degree of freedom associated with the error (Hopkins, 1982) (Rudman/Raudenbush, 1987, p. 9).

Determination of Variables

The variables chosen for inclusion in this research were selected on the following criteria.

1. A replication of the variables in the Rudman/ Raudenbush research.

2. They were variables that have been shown to be related to achievement in prior social science research and for which statistics were available to the researcher.

The variables--sex, ethnicity, family configuration, membership in Chapter I, and past academic achievement--were included in this study. Variables, such as these, have continued to be the focus of sociological research (Rehberg & Rosenthal, 1978, p. 4). In addition, it has been noted that children from families of low levels of education and different language and cultural styles do

not do well on formal tests of ability and aptitude (Brookover & Erickson, 1975, p. 104).

The demographic variables are defined in the following manner:

Sex is male or female.

Ethnicity is White (non-Hispanic) and other (largely Black, some Hispanic, Asian, Japanese, and Native Americans).

Family configuration included the categories of one-parent or two-parent family structures.

Membership in AFDC consists of students receiving Aid to Families of Dependent Children.

Membership in Chapter I is those children receiving supplemental help through the district Chapter I program. These students were those who achieved in the bottom quartile in reading and/or math on the Fall Iowa Test of Basic Skills.

The coding for the variables is contained in Table 3.8 and again in the data in the Appendix.

Directions for Administering the Test

The teachers who volunteered to participate in the study all received researcher prepared direction sheets. All of the instruction sheets were duplicates of the directions given in the manual for the Reading Comprehension and Reference Materials subtest of the Iowa

Test

alter

class

Appen

Time

exten

1986)

of th

Basic

Chapte

select

T

of the

the ti

(1986;

have pa

3.6 for

Time Ass
to Build

The

assigned

divided

blocking

ranking

Test of Basic Skills, except that the time limits were altered according to the experimental group to which the class belonged. These direction sheets are shown in Appendix A.

Time Alterations

It was felt reasonable to replicate the time limit extensions of the previous study (Rudman & Raudenbush, 1986) after comparing the Reading Comprehension subtest of the Standard Achievement Test and the Iowa Test of Basic Skills. A comparison of the two tests appears in Chapter IV. The tests are comparable on the criteria selected.

The time limits for the Reference Materials subtest of the Iowa Test of Basic Skills will be extended as were the time limits in the original Rudman/Raudenbush Study (1986; 1987). This is a different subtest and does not have parallel data from the earlier research. See Table 3.6 for testing times.

Time Assignments Related to Buildings

The classrooms within the blocks were randomly assigned to treatment groups. The treatments were divided between the schools as shown in Table 3.7. The blocking variables could be viewed as ordinal variables, ranking the classrooms on background variables related to

Table

Testi

Treatm

1

2

3

4

*

Skills

Table 3.6

Testing Time by Treatment and Subtest.

Treatment	Reading Time in Minutes	Reference Materials Time in Minutes
1	42*	25*
2	47	30
3	52	35
4	57	40

*Preliminary Technical Summary, Iowa Test of Basic Skills, Riverside Publishing Company, 1986, p. 4.

Table 3.7

Experimental Groups by Building Number of Classrooms per Site.

Time	School A	School B
Control (0)	1	2
5 minutes	1	2
10 minutes	2	1
15 minutes	2	1

the outcomes. The experimental assignments were as balanced as possible between buildings, considering that there were three blocks and two school sites (see Table 3.7).

Summary

This research was designed to look at the effect of increasing the time limits on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills. The samples were drawn from students in the sixth grade classrooms who volunteered to be part of the study. A total of twelve classrooms were divided into three blocks using ability, defined as prior results on the Reading Comprehension subtest of the Iowa Test of Basic Skills, as the primary blocking factor.

The increases in time limits were set in five-minute increments, a control group (the time limits set by the publisher), five additional minutes, ten additional minutes, and fifteen additional minutes. These paralleled the prior research by Rudman and Raudenbush.

The analytical design employed standard regression procedures and HLM statistical procedures to analyze the data for the three balanced blocks of four groups each. The HLM model is thought to be more accurate when there are small sample sizes and unequal block sizes as it

Table 3.8

Demographic Variables Codes

Variable	Code
Sex	
Male	0
Female	1
Ethnicity	
White	0
Non-White	1 or 2
Family Configuration	
One Parent	1
Two Parents	2
Chapter I	
No Chapter I	0
Receiving Chapter I	1

automatically takes these statistical problems into account as it analyzes the data.

The blocks were established to be as different as possible on achievement. The three blocks were ordered in ability from high to low. Assignment to experimental groups within blocks was done randomly.

The best covariate for the Reading Comprehension subtest of the Iowa Test of Basic Skills proved to be prior results on the Reading Comprehension subtest of the Iowa Test of Basic Skills. The best covariate for the Reference Materials subtest of the Iowa Test of Basic Skills proved to the prior achievement on the Reference Materials subtest of the Iowa Test of Basic Skills.

The standard F test of .05 level of confidence was used. This matches most of the social science research.

CHAPTER IV

ANALYSIS AND FINDINGS

This research study was designed to test the effect of increasing time on the achievement test results on the Reading Comprehension and Reference Materials subtests of the 1986 edition of the Iowa Test of Basic Skills. The possible relationship of selected covariates to the outcome was examined using a multiple-regression procedure. The best predictors of achievement on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills were determined to be past achievement on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills. The results were analyzed using adjusted and unadjusted means for the effect of the best covariate. Linear, quadratic, and cubic effects were analyzed to determine the nature of any relationship. A linear effect would establish a direct relationship between time and achievement scores, a quadratic effect would suggest that, if a relationship exists, it may diminish in time and give some indication of how much time is "optimal" or when the point of diminishing returns is reached. A cubic effect would

suggest a mixed relationship between increased time and achievement.

Outcome Data

The experimental sample took their tests on mark sense sheets that were machine scored. The results were transferred to the mainframe computer for analysis via a tap transfer to minimize error. The data were analyzed using a statistical program, SPSS-X (Norusis, 1988). A summary of the outcome data appears in Table 4.1.

Visually, there appeared to be a moderate scattered effect of increasing scores when test time is increased for most treatment groups on the Reading Comprehension subtest. The effect did not appear to be linear as there was not an increasing raw score for each increase in time. The group in the five-minute treatment in Block 3 did not appear to fit the general pattern of the results, and their score was quite depressed. The effects for the Reference Materials subtest showed moderate scatter in scores when additional time was given to take the test. The difference in general outcome for the five-minute experimental time increase in Block 3 was also present in the results for this subtest. This class also had depressed scores on the pretest, 19.43 (see Table 3.3). The results did not appear to be linear.

Table 4.1

Outcome Data Unadjusted Means Raw Score

Subtest/ Blocks	None	Five Minutes	Ten Minutes	Fifteen Minutes	Pooled
Unadjusted Classroom Means and Sample Sizes					
Reading Comprehension					
Block 1					
Pretest	33.17	38.84	34.17	34.78	35.29
Posttest	37.33	44.88	39.48	42.74	41.11
Block 2					
Pretest	27.83	27.04	30.20	28.92	28.49
Posttest	27.87	32.68	34.04	28.53	30.78
Block 3					
Pretest	26.30	19.43	26.73	25.40	24.95
Posttest	38.23	19.14	34.55	34.00	31.48
Posttest Pooled	34.25	33.38	35.34	34.52	34.37
Reference Materials					
Block 1	29.21	41.31	33.31	30.04	33.47
Block 2	25.26	30.96	32.12	25.40	28.44
Block 3	27.14	18.00	25.18	32.59	25.73
Pooled	27.21	30.42	30.24	28.96	29.21
Sample Size "Counts"					
Block 1	24.00	26.00	29.00	27.00	106.00
Block 2	23.00	22.00	25.00	25.00	95.00
Block 3	23.00	14.00	22.00	15.00	74.00
Pooled	70.00	62.00	76.00	67.00	275.00

Analysis

Choosing the Best Covariate

The following questions were posed concerning covariate issues:

1. Is prior achievement, defined as a student's score on a prior test of the Reading Comprehension subtest of the Iowa Test of Basic Skills, a good predictor of expected achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills?
2. Is gender a good predictor of expected achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills?
3. Is family configuration a good predictor of expected achievement of the Reading Comprehension subtest of the Iowa Test of Basic Skills?
4. Is assignment to Chapter I programming a good predictor of expected achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills?
5. Is membership in a minority ethnic group a good predictor of expected achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills?

An analysis of variance procedure was used to determine the correlation between the discrete variables and achievement on the Reading Comprehension subtest and the Reference Materials subtest of the Iowa Test of Basic Skills. Correlations were tested using variables in isolation and in combination with other variables. Previous research has shown that the variables selected for this research are correlated with an increase in achievement. They, singly or in combination, should explain all or most of variance in scores. Two- and three-way interactions were included. The statistical data appear in Tables 4.2 and 4.3.

The data confirmed that the best predictor of success on the Reading Comprehension subtest of the Iowa Test of Basic Skills is achievement, defined as prior performance on the Reading Comprehension subtest. This relationship showed significance to the .000 level of confidence. There is virtually no chance that the results on a prior subtest in Reading Comprehension would not be a good predictor of subsequent results. Prior results on the Reference Materials subtest also appeared to be a good predictor of success on the Reading Comprehension subtest, but not as strong as prior results on the Reading Comprehension subtest. The significance level was .039, within the .05 guideline usually used in social science research and also embraced in this

Table 4.2

Relationship Between Covariates and Achievement on the Reading Comprehension Subtest of the Iowa Test of Basic Skills Adjusted Score Data.

Source of Variation	Sum of Squares	DF	Mean Square	F	P of F
Covariates	16348.336	2	8174.168	88.945	.000
Preread	4624.378	1	4624.378	126.794	.000
Preference	397.612	1	397.612	4.327	.039
Main Effects	474.306	4	118.576	1.290	.274
Sex	15.906	1	15.906	.173	.678
Chapter I	414.813	1	414.813	4.514	.035
Race	14.904	1	14.904	.162	.687
Family	92.918	1	92.918	1.011	.316
2-Way Interactions	339.390	6	56.565	.615	.718
Sex/Chapter	46.251	1	46.251	.503	.479
Sex/Race	9.122	1	0.122	.099	.753
Sex/Family	222.005	1	222.005	2.416	.121
Chapter/Race	12.248	1	12.248	.133	.715
Chapter/Family	.015	1	.015	.000	.990
Race/Family	28.013	1	28.013	.305	.581
3-Way Interactions	897.381	4	224.345	2.441	.047
Sex/Chp/Race	2.998	1	2.998	.033	.857
Sex/Chp/Fam	268.884	1	268.884	2.926	.088
Sex/Race/Fam	46.209	1	46.209	.503	.479
Chp/Race/Fam	229.074	1	229.074	2.493	.116
Explained	18059.413	16	1128.713	12.282	.000
Residual	23618.645	257	91.901		
Total	41678.058	273	152.667		

Table 4.3

Relationship Between Covariates and Achievement on the Reference Materials Subtest of the Iowa Test of Basic Skills Adjusted Score Data.

Source of Variation	Sum of Squares	DF	Mean Square	F	P of F
Covariates	6664.956	2	3332.478	20.671	.00
Preread	336.244	1	336.244	2.086	.150
Preference	1480.832	1	1480.832	9.186	.003
Main Effects	753.835	4	183.959	1.141	.338
Sex	145.481	1	145.481	.902	.343
Chapter I	107.392	1	107.392	.666	.415
Race	84.828	1	84.828	.526	.469
Family	248.353	1	248.353	1.541	.216
2-Way Interactions	332.765	6	55.461	.344	.913
Sex/Chapter	54.492	1	54.492	.338	.561
Sex/Race	14.381	1	14.381	.089	.765
Sex/Family	22.449	1	22.449	.139	.709
Chapter/Race	71.288	1	71.288	.442	.507
Chapter/Family	.266	1	.266	.002	.968
Race/Family	145.825	1	145.825	.905	.342
3-Way Interactions	1045.475	4	261.269	1.621	.169
Sex/Chp/Race	65.415	1	65.415	.406	.525
Sex/Chp/Fam	199.242	1	199.242	1.236	.267
Sex/Race/Fam	81.458	1	81.458	.505	.478
Chp/Race/Fam	104.030	1	104.030	.645	.423
Explained	8779.031	16	548.689	3.404	.00
Residual	41431.674	257	161.213		
Total	50210.704	273	183.922		

research. This variable, however, was not as good as the pretest in Reading Comprehension. One other variable turned out to be significant: membership in Chapter I. It was significant to the .035 level of confidence. The raw score data shown in Table 4.1 shows that Block 3, the block containing the highest percentage of students in Chapter I (38% on average to 10.9% and 26.9% for Blocks 1 and 2, respectively) had a different achievement pattern over time than either Block 1 or Block 2. The group in the five-minute experimental time for Block 3 had very different results, as noted in Table 4.1, is entirely made up of Chapter I students. The rest of the covariates were not significant, either alone or in combination, with other covariates. The only other covariate that is even close to significant is a three-way interaction effect between sex, Chapter I membership, and family configuration at .088 level of significance.

Reference Materials

The same analysis of covariance was performed on the results for the Reference Materials subtest of the Iowa Test of Basic Skills. The results for this subtest are shown in Table 4.3.

The only good predictor of achievement on the Reference Materials subtest of the Iowa Test of Basic Skills proved to be the pretest test of the Reference

Materials subtest of the Iowa Test of Basic Skills, at the .003 level of significance. The rest of the covariates are not significant, either alone or in combination with other covariates. Membership in Chapter I and the three-way interaction between sex, membership in Chapter I, and family configuration were not significant or even close to significant as they were on the Reading Comprehension subtest.

After analyzing the data, it appeared that the best predictor of achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills is prior achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills; the best predictor of achievement on the Reference Materials subtest of the Iowa Test of Basic Skills is the pretest score on a previous Reference Materials subtest of the Iowa Test of Basic Skills. When controlled for the best predictor, the other covariates did not appear to be significant. The effect of these two covariates were used to adjust the scores in later analysis.

Reading Comprehension

The following research question was used when analyzing the Reading Comprehension subtest scores.

Will increasing the time limits on the Reading Comprehension subtest of the Iowa Test of Basic Skills by five, ten, or fifteen minutes significantly increase student achievement?

When the data for Reading Comprehension was analyzed by comparing the post test unadjusted means for each experimental group, there appeared to be a mixed relationship between time and achievement (Table 4.1). Blocks 1 and 2, the two highest ability groups, showed a small increase across all treatments. There was no clear pattern of a greater effect for any one treatment. Block 3, the group of lowest ability, showed a decrease in mean across all treatments. When the means were pooled across all blocks, the effect appeared to be fairly flat or no increase in average score (Table 4.4).

Table 4.4

Pooled Unadjusted Treatment Means (Raw Score) for Reading Comprehension Subtest Iowa Test of Basic Skills

Treatment Group	Treatment Mean
Control	34.25416
Treatment 1 (5 minutes)	33.38303
Treatment 2 (10 minutes)	35.34277
Treatment 3 (15 minutes)	34.52455

To determine if the moderate change in the raw score means on the Reading Comprehension subtest of the Iowa Test of Basic Skills, noted visually by comparing the treatment means, is significant, a standard ANOVA procedure was employed. The results of the ANOVA are shown in Table 4.5.

Using adjusted raw scores the relationship between increased time and Reading Comprehension subtest scores on the Iowa Test of Basic Skills did not appear to be significant. However, the treatment effect on the outcome measure does appear to be dependent upon block membership. Any effect, significant or not, of time on achievement may be dependent on ability (p=.05). Block membership in this experimental sample was determined by ability level. In addition, there was a significant treatment by block interaction effect (p=.00). It may be that effect of increasing time on the achievement results of student, although not significant, may be different for students of low ability than students of average or better ability on the Reading Comprehension subtest of the Iowa Test of Basic Skills.

The adjusted means Table 4.6 for the results of the Reading Comprehension subtest were graphed by block to help explain the treatment by block interaction effect. It is possible that what we are seeing is a teacher or a tracking effect and not directly attributable to

Table 4.5

Analysis of Variance Using Unique Sum of Squares, Raw Scores, Unadjusted Reading Comprehension.

	SS	DF	MS	F	P of F
Within Cells	22414.39	262	85.55		
Covariate					
Pretest-read	9192.37	1	9192.37	107.45	.000
5 Minutes	25.31	1	255.31	.30	.587
10 Minutes	.05	1	.05	.00	.982
15 Minutes	102.040	1	102.40	1.20	.275
Block	942.32	2	471.16	5.51	.005
Treat/Block	2571.40	6	428.57	5.01	.000

increasing the time. Figure 4.1 illustrates a treatment by block effect. The depressed scores for the 5 minute treatment in Block 3 is markedly different from all other treatment scores. All members of this group were Chapter I students and the pretest score was also depressed. This could be the result of prior placement of students by ability, low expectations of a teacher or teachers due to the tracking or poor test administration practices.

Reference Materials

The next research question was concerned with adding time to the Reference Materials subtest of the Iowa Test of Basic Skills.

> Will increasing the time limits on the Reference Materials subtest of the Iowa Test off Basic Skills by five, ten, or fifteen minutes significantly increase student achievement?

When the unadjusted treatment means for the Reference Materials subtest were viewed, it appeared that there was a mixed effect of time on achievement. Blocks 1 and 2 showed a small increase in mean across all treatments. Block 3 showed a decrease for treatment 1 (a five-minute increase in time) and treatment 2 (a ten-minute increase in time); however, the mean was slightly higher for treatment 3 (a fifteen-minute increase in time). When the means were pooled, the increase in means appeared fairly flat across all treatments (see Tables 4.1 and 4.7).

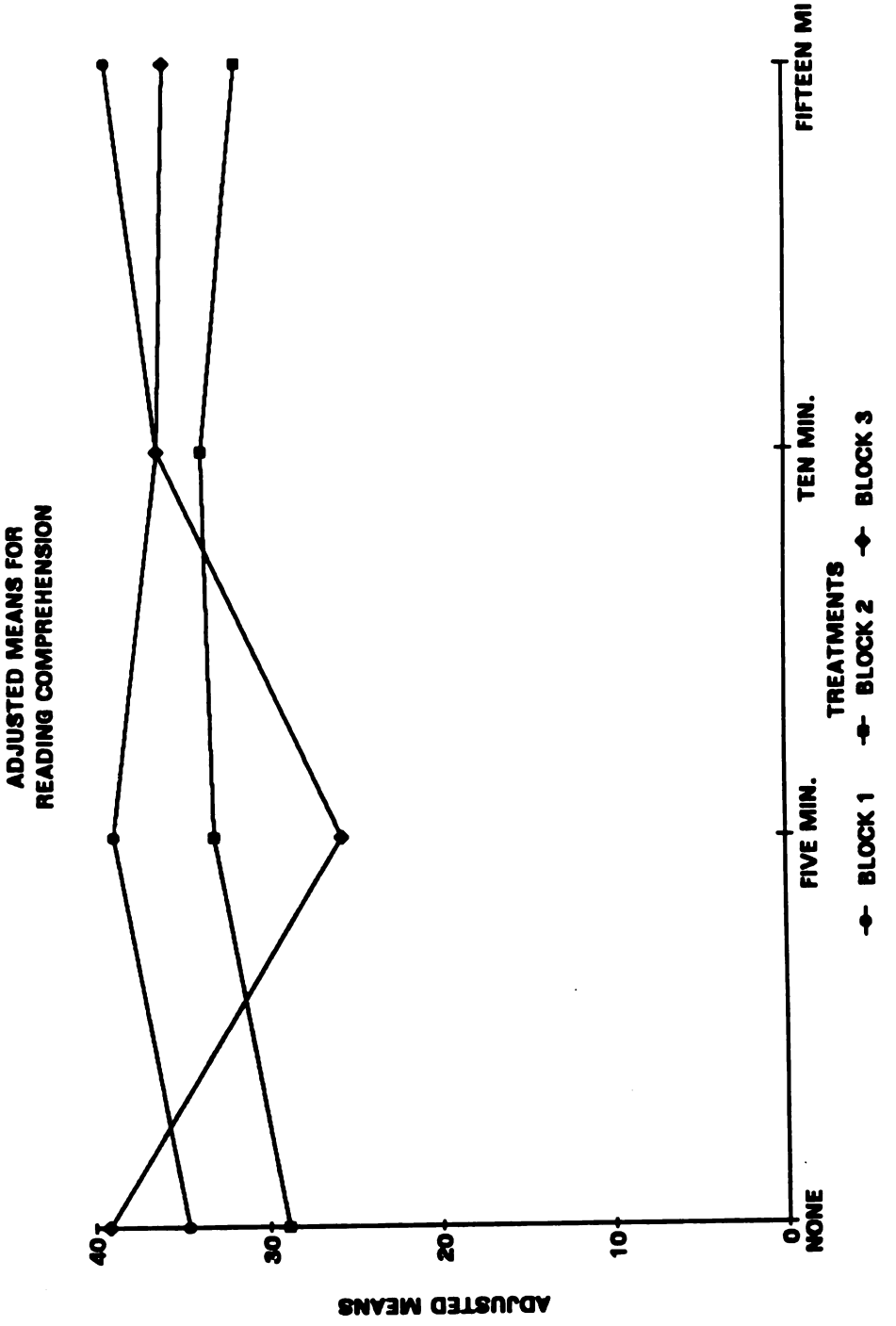


Figure 4.1. Treatment Effect by Block. Adjusted Means.

Table 4.6

Adjusted Posttest Means Reading Comprehension

Blocks	None	Five Minute	Ten Minutes	Fifteen Minutes
Block 1	34.67	38.77	36.15	39.01
Block 2	28.91	33.02	33.66	31.55
Block 3	39.18	25.68	36.18	35.70

Table 4.7

Pooled Unadjusted Treatment Means for Reference Skills Iowa Test of Basic Skills Raw Scores.

Treatment Group	Treatment Mean
Control	27.21199
Treatment 1 (5 minutes)	30.42072
Treatment 2 (10 minutes)	30.24109
Treatment 3 (15 minutes)	28.96399

An ANOVA was run on the treatment effect of time on achievement on the Reference Materials subtest of the Iowa Test of Basic Skills. The ANOVA showed that there was no significant effect. Block membership was a significant factor at the .017 level of significance, as it was in the Reading Comprehension subtest ($F=5.51$, $p=.005$) (see Table 4.8).

The study does not support any significant increase in test scores when the time is increased on the Reference Materials subtest of the Iowa Test of Basic Skills when the data source is unadjusted raw score. The effect was dependent upon block membership.

Additional Research Questions

Although no overall significant effect of increasing time on test scores was found, the researcher asked some additional questions about the important issue of added time on achievement. Those questions are:

Question 1: Are the results affected by student ability?

Results indicated that the effect of excess time on achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills may vary, depending on block membership. There was a similar finding for the Reference Materials subtest. Further analysis of the results by treatment by block showed some interesting patterns of results. The groups within each block were

Table 4.8

Analysis of Variance Using Unique Sum of Squares, Raw Scores, Adjusted Reference Skills.

	SS	DF	MS	F	P of F
Within Cells	40416.12	261	154.85		
Covariate					
Pretest-read	2313.36	1	2313.36	14.94	.000
5 Minutes	84.46	1	84.46	.55	.461
10 Minutes	329.24	1	329.243	2.13	.146
15 Minutes	17.16	1	17.16	.11	.740
Block	1285.66	2	642.83	4.15	.017
Treat/Block	1858.10	6	309.68	2.00	.066

established to be as alike as possible on the previous falls' Reading Comprehension subtest scores, used as an achievement variable. The blocks were ordered in ability on this variable as noted in Chapter III. A comparison of the means broken down by block by treatment is shown in Table 4.9.

The means for the experimental groups in Blocks 1 and 2 suggested that increasing time does have at least a minimal effect on achievement; however, in Block 3 the effect appeared to be negative.

The results for Treatment 1 (5 minutes) in Block 3 appears to be suspect. The severe drop in the score suggests that some other factor, such as a teacher's effect or an effect of tracking, may be interfering with that result. An analysis of this group did show that all of the students in this class are Chapter I students. Chapter I students in the middle-sized city school district are those students who score at or below the 35.8 NCE on the Spring Iowa Test of Basic Skills. In practice, this usually means the bottom reading group at each grade level in each school. This class was ability grouped so that all the students in it were Chapter I students. This group also had a severely lower score on the pretest. The possible teacher effect was discussed earlier.

Table 4.9

Unadjusted Means for the Reading Comprehension Subtest of the Iowa Test of Basic Skills by Block by Treatment.

Treatment Group	Raw Score Mean
Block 1	
Control	37.3333
Treatment 1 (5 minutes)	44.8846
Treatment 2 (10 minutes)	39.4828
Treatment 3 (15 minutes)	42.7407
Block 2	
Control	27.8696
Treatment 1 (5 minutes)	34.0000
Treatment 2 (10 minutes)	34.0400
Treatment 3 (15 minutes)	32.6800
Block 3	
Control	38.2272
Treatment 1 (5 minutes)	19.1429
Treatment 2 (10 minutes)	34.5455
Treatment 3 (15 minutes)	28.5333

The strongest relationship appeared to be in Blocks 1 and 2, or the most able classes in the sample. The least linear pattern is in Block 3, the students who are the least able in the sample. This supports the evidence noted earlier that any possible effect, although not significant, of time on achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills may be different depending on student ability.

The same general patterns do seem to exist for the Reference Materials subtest. There is no clear pattern. The Reference Materials means by block by treatment are seen in Table 4.10.

Question 2: Are the Reading Comprehension subtest of the Iowa Test of Basic Skills and the Reading Comprehension subtest of the Stanford Achievement Test comparable?

The results of the Rudman and Raudenbush research indicated that a relationship does exist between time and achievement on the Reading Comprehension subtest of the Stanford Achievement Test (1986 and 1987). The fact that this relation was found to be significant and confirmed by additional testing, but not supported by the findings in this research using the Reading Comprehension subtest of the Iowa Test of Basic Skills suggests an additional research question. Does the Reading Comprehension subtest of the Stanford Achievement Test and the Reading Comprehension subtest of the Iowa Test of Basic Skills

Table 4.10

Unadjusted Means for the Reference Skills Subtest of the Iowa Test of Basic Skills by Block by Treatment.

Treatment Group	Raw Score Mean
Block 1	
Control	29.2083
Treatment 1 (5 minutes)	41.3077
Treatment 2 (10 minutes)	33.3103
Treatment 3 (15 minutes)	30.0370
Block 2	
Control	25.2609
Treatment 1 (5 minutes)	32.5909
Treatment 2 (10 minutes)	32.1200
Treatment 3 (15 minutes)	30.9600
Block 3	
Control	27.1363
Treatment 1 (5 minutes)	18.0000
Treatment 2 (10 minutes)	25.1818
Treatment 3 (15 minutes)	25.4000

test the same thing? Or, are the two subtests substantially different in some way? The previous research determined that not all subtests of the Stanford Achievement Test were sensitive to extra time. This researcher may have simply found two more subtests of a standardized achievement test that are not sensitive to extra time. A comparison of the two Reading Comprehension subtests follows.

When looking at the comparability of these two tests, the researcher considered objectives that form the basis for the development of the tests, the internal design of each test, and standardization procedures need to be considered. These factors give a clear idea of the focus of each test, the nature of the material used to test that focus, and the standardization samples and process.

Objectives Used for Test Design

The objectives of the Iowa Test of Basic Skills Reading Comprehension subtest are:

Facts: To recognize and understand stated factual details and relationships (literal meaning).

- F1 Description: To understand factual details relating to description of people, places, objects, and events.
- F2 Categorization: To understand factual details relating to classification.
- F3 Relationships: To understand functional relationships, time, and sequence.

F4 Contextual Meaning: To deduce the meanings of words or phrases from context.

Inferences: To infer underlying relationships (interpretative meaning).

I1 Cause and Effect: To understand cause, effect, concomitance, and interaction.

I2 Draw Conclusions: To draw conclusions from information and relationships.

I3 Traits and Feeling: To infer traits, feelings, and emotions of characters.

I4 Motives: To infer the motives and reasons for the actions of characters.

To develop generalizations from a selection (evaluative meaning.

G1 Main idea: To recognize the main idea or topic of a paragraph or selection.
G2 Organization: To understand the organization of a paragraph or selection.

G3 Application: To apply information through generalization or prediction.

G4 Purpose: To recognize the author's purpose, motive, or intention.

G5 Viewpoint: To recognize the author's viewpoint, attitude, or bias.

G6 Figurative language: To interpret figurative language.

G7 Mood: To recognize the mood or tone of a selection.

G8 Style: To recognize qualities of style or structure (Hieronymus, 1986, p. 80).

These objectives form the basis for the development of the test questions. They reflect very specific, discrete reading skills that are parts of the process of reading comprehension. They are not as concerned with

testing various kinds of reading in "life" as they are with the academic makeup of reading. The number of questions devoted to each objective and skill are listed in Table 4.11 (Hieronymus, 1986c, pp. 62-63).

In contrast, the Stanford Achievement Test Reading Comprehension subtest is based on broader objectives that are designed to look at comprehension for different kinds of reading. These objectives do not focus on the many discrete subskills of reading comprehension, but rather, they focus on the gestalt of comprehension as encountered in real life situations both in school and out of school. This is a major difference in emphasis. The objectives for the Reading Comprehension subtest of the Stanford Achievement Test are:

Textual Reading: Demonstrate the ability to read and comprehend passages that are typical of material found in grade-appropriate textbooks.

Functional Reading: Demonstrate the ability to read and comprehend printed material that is typically encountered in every day life.

Recreational Reading: Demonstrate the ability to read and comprehend printed material typically read for pleasure.

Literal Comprehension: Demonstrate the ability to comprehend explicitly stated meanings and details by answering questions about a variety of reading passages.

Table 4.11

Number of Questions Per Objective Iowa Test of Basic Skills.

Objective	Number of Questions
Facts	21
F1	6
F2	5
F3	6
F4	4
Inferences	16
I1	4
I2	4
I3	2
I4	6
Generalizations	19
G1	5
G2	2
G3	3
G4	1
G5	2
G6	3
G7	2
G8	1

Inferential Comprehension: Demonstrate the ability to draw conclusions and make inferences and generalizations from explicitly and implicitly stated meanings by answering questions about a variety of reading passages (Gardner, 1987, pp. 51-53).

The number of questions devoted to each objective are contained in Table 4.12 (Gardner, 1987, pp. 51-53).

The Reading Comprehension subtest of the Stanford Achievement Test is designed to measure broader

Table 4.12

Number of Questions Per Objective.

Objective	Number of Questions
Textual Reading	20
Functional Reading	20
Recreational Reading	20
Literal Comprehension	30
Inferential Comprehension	30

objectives than the more focused objectives of the Reading Comprehension subtest of the Iowa Test of Basic Skills. Both tests have literal and inferential comprehension as stated objectives. The Stanford also

stated objectives related to textual reading, functional reading, and recreational reading and these are designed to overlap with inferential and literal comprehension in the design of the test. The Iowa does not list these considerations as objectives. Based on this analysis, the focus of the Reading Comprehension subtest of the Iowa Test of Basic Skills is more limited to discrete academic skills and literal and inferential comprehension as they relate to usual reading in school. The Stanford has a more realistic presentation of real life reading, especially out of school topics and kinds of reading material. It should be noted that the authors of the Iowa Test of Basic Skills do state that they include materials within the reading passages from "all types of material encountered by . . . pupils in their every reading . . . as possible" (Hieronymus, 1986, p. 78). But, this material is really material students ordinarily encounter in school.

A review of the passages chosen for inclusion in the Iowa Test of Basic Skills revealed that all were the kinds of reading a student would expect to find in school. There were three articles that dealt with science, two that dealt with social studies material, two poems, and one general reading article. For the most part, the paragraphs are short and numbered. The writing is also set in green boxes to direct the attention of the

test taker to the reading material. The paragraph numbers are referred to in some of the questions in the test. A listing of the articles, number of paragraphs, and number of questions related to each appears in Table 4.13. The focus of the Reading Comprehension subtest of the Iowa Test of Basic Skills is clearly academic.

The passages chosen for the Stanford Achievement Test were more universal in nature, that is, they include material that students would not ordinarily encounter in school. There were three general information articles, one story, a section from the T.V. Guide, a poem, a friendly letter, and some directions for a science experiment. The paragraphs were not numbered and they were not set off in a different colored box. In some cases (i.e., the sales slip, the memo, and the plant tag), the articles were set off in boxes. A listing of the articles, number of paragraphs, and number of questions related to each appears in Table 4.17.

Internal Design Characteristics

An analysis of the internal characteristics of the tests considered the factors of speed, length of passages, and reading difficulty. The Iowa Test of Basic Skills is not as speeded as the Stanford Achievement Test based on either the measures of words per minute or questions per minute. The Iowa requires a student who

Table 4.13

Comparison of the Content of the Reading Comprehension Subtests of Stanford Achievement Test and Iowa Test of Basic Skills.

	Number of Paragraphs	Number of Questions
Stanford Achievement Test		
Type of Writing:		
Article (science)	2	6
Story	1	7
T.V. Guide	4	7
Directions (science)	5	5
Article (social studies)	3	7
Friendly Letter	3	8
Article (social Studies)	3	9
Story	1	6
Poem	1	5
Iowa Test of Basic Skills		
Type of Writing:		
Article (science)	5	8
Article (social studies)	10	10
Poem	1	5
Article (science)	5	8
Article (science)	3	8
Article (social studies)	6	7
Poem	1	4
Story	1	6

finishes the test to read an average of 42.5 words per minute compared to 46.93 for the Stanford. This is not as big a difference as the number of questions per minute. Students taking the Iowa must answer an average of 1.3 questions per minute compared with 2.0 questions per minute for the Stanford, almost twice as many. It is important to remember that this comparison is between the Reading Comprehension subtest for the fifth grade (Stanford) and sixth grade (Iowa). The Iowa test has considerably longer sentences and more difficult words. These factors attest to the topic selections for the Iowa Test of Basic Skills which included mainly the kinds of reading children do in school.

The researcher used standard procedures to determine reading levels. All passages of each test were included in the analysis. The figures shown in Tables 4.14 to 4.18 include the readability factors considered important by the authors of the major readability tests. When corrected for grade level, the readability of the two appears to be comparable.

Standardization

The standardization program used for both the Stanford Achievement Test and the Iowa Test of Basic Skills employed school enrollment, geographic region, socioeconomic status, and public vs. nonpublic schools as

Table 4.14

Length and Number of Passages Comparison of the Reading Comprehension Subtest of the Iowa Test of Basic Skills, Level 12 and the Reading Comprehension Subtest of the Stanford Achievement Test, Intermediate 1.

Passage Number	Iowa	Stanford
	Number of Words	Number of Words
1	164	154
2	240	105
3	187	150
4	249	183
5	114	244
6	532	188
7	299	232
8	--	102
9	--	50
Total Words	1785	1408
Total Testing Time	42	30
Words/Min of Test	42.5	46.93

Table 4.15

Questions Per Passage Comparison of the Reading Comprehension Subtest of the Iowa Test of Basic Skills, Level 12 and the Reading Comprehension Subtest of the Stanford Achievement Test, Intermediate 1.

Passage Number	Iowa	Stanford
	Count	Count
1	7	6
2	8	7
3	8	7
4	8	5
5	8	7
6	11	8
7	6	9
8	--	6
9	--	5
Total	56	60
Total Testing Time	42 minutes	30
Questions per Minute	1.3	2.0

Table 4.16

Comparison of the Reading Comprehension Subtest of the Iowa Test of Basic Skills, Level 12 and the Reading Comprehension Subtest of the Stanford Achievement Test, Intermediate 1 on Readability Factors used in the Spache, Dale-Chall, Fry, Rayor, Flesch, and Gunning-Fog Readability Tests.

Readability Factor	Passage Number						
	1	2	3	4	5	6	7
Iowa Test of Basic Skills							
Number of Sentences	14	19	14	23	2	29	18
Number of Words	164	240	187	249	114	532	299
Number of Syllables	239	341	254	345	150	854	373
Words of 6 or more letters	47	56	53	64	33	183	60
3 or more syllable words	21	20	9	20	7	90	12
% of 3 or more syllable words	12.8	8.3	4.8	8	6.1	16.9	4
Average sentence length	11.7	12.6	13.4	10.8	57.0	18.3	16.6
Average letters per word	4.4	4.4	4.5	4.3	4.6	4.7	4.1
Average syllables per word	1.5	1.4	1.4	1.4	1.3	1.6	1.2

Table 4.17

Comparison of the Reading Comprehension Subtest of the Iowa Test of Basic Skills, Level 12 and the Reading Comprehension Subtest of the Stanford Achievement Test, Intermediate 1 on Readability Factors used in the Spache, Dale-Chall, Fry, Rayor, Flesch, and Gunning-Fog Readability Tests.

Readability Factor	Passage Number								
	1	2	3	4	5	6	7	8	9
Stanford Achievement Test									
Number of Sentences	9	7	9	11	15	14	21	17	6
Number of Words	154	105	150	183	244	188	232	102	50
Number of Syllables	186	151	211	216	367	229	310	129	75
Words of 6 or more letters	27	29	36	24	67	27	46	25	16
3 or more syllable words	4	12	13	6	39	6	17	2	2
% of 3 or more syllable words	2.6	11.4	8.7	3.3	11.9	3.2	7.3	2.0	4.0
Average sentence length	17.1	15.0	16.7	16.6	16.3	13.4	11.0	6.0	8.3
Average letters per word	3.9	4.5	4.2	3.8	4.4	3.7	4.2	3.9	5.0
Average syllables per word	1.2	1.4	1.4	1.2	1.5	1.2	1.3	1.3	1.5

Table 4.18

Comparison of the Reading Comprehension Subtest of the Iowa Test of Basic Skills, Level 12 and the Reading Comprehension Subtest of the Stanford Achievement Test, Intermediate 1 on Readability Factors used in the Spache, Dale-Chall, Fry, Rayor, Flesch, and Gunning-Fog Readability Tests.

Readability Factor Means Per Passage	Means	
	Mean of the Readability Factor	
	Iowa Level 12	Stanford Intermed 1
Number of Sentences	17	12.11
Number of Words	255	156.44
Number of Syllables	365.14	208.22
Words of 6 or more letters	70.85	33
3 or more syllable words	25.57	11.22
% of 3 or more syllable words	8.7	6.04
Average sentence length	20.05	13.37
Average letters per word	4.42	4.17
Average syllables per word	1.4	1.33
Grade level tested	6	5
Testing Time	42 minutes	30 minutes

criteria. Further both publishers used their companion ability test in the process (Wellens, p. 7; Heironoymus, p. 62). Both tests were given in the Spring and Fall and used similar selection criteria. There is no reason to believe that any major differences exist between the two tests as far as method of standardization is concerned.

Summary

There was no significant effect of time on achievement results for the Reading Comprehension or the Reference Materials subtests on the Iowa Test of Basic Skills, when testing time is increased by five, ten, or fifteen minutes. Any effect that does exist appeared to be dependent on block membership for both subtests and a treatment by block effect for Reading Comprehension which in this research represents general ability levels for both subtests and a treatment by block effect of Reading Comprehension. The treatment by block effect for Reading Comprehension is believed to be explained as a random effect of the teacher or tracking, and not an effect of the treatment. A comparison between the Reading Comprehension subtest of the Iowa Test of Basic Skills and the Reading Comprehension subtest of the Stanford Achievement Test revealed major differences in the goals that were used to establish the tests and that the Reading Comprehension subtest of Iowa Test of Basic

Reading Skills was not as speeded as the Reading Comprehension subtest of the Stanford Achievement Test. These differences strongly suggest that the two subtests are not measuring the same thing, and the researcher may have simply found two more subtests that are not sensitive to extra time. In addition, the established time limits for the Reading Comprehension subtest of the Iowa Test of Basic Skills appeared to be more generous that those specified for the Reading Comprehension subtest of the Stanford Achievement Test. It may be that more time than is optimal has been given to the Reading Comprehension subtest of the Iowa Test of Basic Skills by the publisher.

Question 3: Was the optimal time limit exceeded when the publisher set the recommended time limits on both the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills?

The fact that no linear, quadratic, or cubic effect was found when time was increased by five, ten, or fifteen minutes on the Reading Comprehension and Reference Material subtests of the Iowa Test of Basic Skills could mean that to determine if a relationship exists between time and achievement a reduction in time must be employed. The 'Rudman/Raudenbush research suggested such a possibility for the Stanford Achievement Test (Rudman & Raudenbush, 1985).

Summary

This research found no significant relationship between time and achievement on the Reading Comprehension or Reference Materials subtests of the Iowa Test of Basic Skills. It did find a block effect for both Reading Comprehension and Reference Materials and a treatment by block effect for Reading Comprehension. Additional questions were raised concerning a possible differential effect on time on achievement for students of different abilities and the comparability of the Reading Comprehension subtests of the Iowa Test of Basic Skills and the Stanford Achievement Test. There is no significant data to support a difference in effect on time on achievement for different students of different abilities.

A comparison of the Reading Comprehension subtest of the Iowa Test of Basic Skills and the Stanford Achievement Test reveals some major differences in the internal design of the tests and the pacing of questions, but no substantial difference in the standardization process. The Stanford Achievement Test is designed to test the kinds of reading in every day life. The Iowa Test of Basic Skills is more closely tied to academic reading.

CHAPTER V

SUMMARY

Purpose

The purpose of this study was to determine if the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills are sensitive to increases in time. They were not found to be sensitive to increases in time. This question is important due to the need for accurate testing results when test results are used to make high stake decisions. Achievement tests are time/power tests; they are designed to provide a measure of how much a student knows under power conditions (the absence of time limits), but set in a comfortable framework of time that answers the administrative need for efficient use of available student instructional time. Prior research by Rudman and Raudenbush showed that the Reading Comprehension subtest of the Stanford Achievement Test was sensitive to alterations in time (Rudman & Raudenbush, 1986). If other subtest of other standardized tests are found to be time sensitive, then teachers or schools could control, to some extent, the outcome data by altering the testing time.

High Stakes

Tests are becoming the preferred quality control devise in education (Salganik, 1985). The nation has become increasingly concerned about the state of American education. As the world shrinks and competition between countries for economic superiority increases in intensity, American educators have been charged with making America No. 1 in all aspects of education, but literacy, math, and science have been labeled as most crucial. At the 1991 Governor's Conference in Charlottesville, Virginia, six National Educational Goals were established which, when achieved by the year 2000, will ensure that future generations in the United States will possess the skills necessary to compete in the world of tomorrow. President Bush has embraced these goals for education. The national agenda includes the use of standardized tests to help demonstrate that these goals are being met by the nation's students.

Individual states, Florida, Texas, Tennessee, and Michigan, among others, have also established goals for education and plans to achieve those goals. Michigan's P.A. 25 calls for the measurement of achievement by standardized tests as well as other measures. Failure to met the criteria established by the government and educators will result in reduction of funding or even possible dismissal of educational staff. The plan calls

for the state to take over schools that do not meet criteria for success. For these reasons, it is imperative that standardized achievement tests be absolutely representative of the national curriculum and goals, and not be subject to measurement error based upon relatively minor changes in testing time.

Standardized achievement tests are normed on the basis of being given by classroom teachers in regular sized classrooms. This increases the chances that test irregularities will occur. As the number of testers increases, so does the likelihood of some one accidentally or on purpose adding testing time to a given subtest. If you add the dimension of purposeful cheating, or unethical conduct brought on by high stakes decision making, the possibility increases even more dramatically. If we are truly to have accurate measures of educational achievement, our tests must not be sensitive to minor increases in time. If selected subtests are sensitive, they must be discovered and alternative ways found to improve the testing limits.

Design of the Research

This research looked at the effect of increasing time limits on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills. The study used twelve sixth grade classrooms from a middle-

sized, Midwestern school system that volunteered to be part of the study. The sample was divided into three blocks of four classrooms each using ability as the primary blocking variable, defined as achievement on the prior Fall's Reading Comprehension subtest of the Iowa Test of Basic Skills, and proportion of children receiving AFDC as a secondary blocking variable. Within each block classrooms were then randomly assigned to treatments, the addition of five, ten, or fifteen minutes of testing time. The covariates closely matched those of the replicated study were achievement, ethnicity, sex, economic status, defined as receiving AFDC, and family status, defined as number of parents. The best covariate for Reading Comprehension was found to be prior achievement on the Reading Comprehension subtest of the Iowa Test of Basic Skills and the best covariate for the Reference Materials subtest was found to prior achievement on the Reference Materials subtest of the Iowa Test of Basic Skills. Reference materials and membership in Chapter I also were significant covariates for the Reading Comprehension subtest. The results were analyzed using an analysis of variance and multiple regression. Linear, quadratic, and cubic trends were analyzed. The research determined that increasing the test time on the Reading Comprehension and Reference Materials subtests of the Iowa Test of Basic Skills does

not significantly effect the scores, however, there is a block effect and for Reading Comprehension a treatment by block effect.

Conclusions

1. There is no effect of increasing achievement when the testing time is increased five, ten, or fifteen minutes on the Reading Comprehension subtest of the Iowa Test of Basic Skills.

2. There is a block effect and a treatment by block interaction when the testing time is increased by five, ten, or fifteen minutes on the Reading Comprehension subtest of the Iowa Test of Basic Skills. However, there is reason to believe that this was a random effect.

3. There is no effect of increasing achievement when the testing time is increased five, ten, or fifteen minutes on the Reference Materials subtest of the Iowa Test of Basic Skills.

4. There is a block effect when the testing time on the Reference Materials subtest of the Iowa Test of Basic Skills is increased by five, ten, or fifteen minutes.

Discussion

There was no significant relationship found between time and achievement on either the Reading Comprehension

or the Reference Materials subtest of the Iowa Test of Basic Skills. A block effect was significant for both Reading Comprehension and Reference Materials suggesting that any effect was dependent on ability. In addition, a treatment by block interaction effect was found for Reading Comprehension, suggesting that treatment was significant for students in at least one of the groups. The treatment by block effect is thought to be a random effect. Only one group adjusted mean was very different from the general pattern of scores. It is possible that this effect would have occurred, regardless of what treatment the group had randomly been assigned. The effect is thought to be either a teacher effect or an effect of tracking as this group was made up entirely of Chapter I students. Rudman and Raudenbush found a significant relationship between increasing time and achievement on the Reading Comprehension subtest of the Stanford Achievement Test (1986; 1987). This study did not find that this sensitivity to time could be duplicated for the Reading Comprehension subtest of the Iowa Test of Basic Skills. But are the two subtests really comparable, or are they sufficiently different as to constitute two different subtests?

An analysis of the two subtests showed that the Reading Comprehension subtest of the Stanford Achievement Test was more speeded, 2.0 questions per minute, about

65% faster than the Reading Comprehension subtest of the Iowa Test of Basic Skills, 1.3 questions per minute, (see Table 4.17). The slower pace established by the authors of the Iowa Test of Basic Skills appears to reduce the chance that additional time will significantly effect the outcome. However, it may be that time limits established for the Reading Comprehension subtest of the Iowa Test of Basic Skills is too generous. The lack of any relationship between time and achievement suggests that maybe less time could be given to this subtest without effecting the results. There are considerable differences in speededness. It would be interesting for another researcher to reduce the time given for the Reading Comprehension subtest of the Iowa Test of Basic Skills to see if increasing the speededness would show an effect on achievement, and where the time limit is beyond which no significant increase in achievement occurs.

Another research question asked if the two tests, the Reading Comprehension subtest of the Stanford Achievement Test and the Reading Comprehensive subtest of the Iowa Test of Basic Skills actually measure the same thing. An analysis of the two subtests showed that the Stanford Achievement Test was oriented more toward the kind of reading found in every day life, "real world reading," and the Iowa Test of Basic Skills more toward the kind of reading students encounter in textbooks,

"academic reading." It is possible, then, that the two subtests do not actually measure the same kind of reading, and therefore, the reason that results of the Rudman/Raudenbush study were not duplicated is that the two subtests, in fact, measure different kinds of reading. Therefore, they may be different kinds of subtests and we have just discovered another subtest that is not sensitive to time.

The results for the Reference Material subtest of the Iowa Test of Basic Skills also showed no effect of increasing testing time on achievement.

A final suggestion of this research was to suggest that any possible effect of adding time to the Reading Comprehension subtest of the Iowa Test of Basic Skills was dependent on group membership, that is, achievement level. It appeared that average and above-average students were more positively affected by increasing testing time than below average students.

Implications for Future Research

Content Match

Are there different kinds of speed for different kinds of tasks (Lord, 1956)? One of the research questions asked by this research was:

> Are the reading comprehension subtests of the Iowa Test of Basic Skills and the Stanford Achievement Test equivalent, that is, do they appear to test the same things?

An analysis of the two tests showed differences in goals, topics, mixture of reading purposes, and pacing of questions and reading speed. The two subtests really are not very parallel, even though they purport to measure the same thing: reading comprehension. It is possible that reading comprehension is so wide a topic that the multitude of tasks involved are sampled different by different authors. These different tasks, represented by different goals, may respond differently to variations in time. A future researcher may wish to compare the Reading Comprehension subtest of various standardized tests to see how closely they match and then chose a test that more closely approximates the Stanford Achievement Test and repeat this replication study to see if it is sensitive to increased time. Additional matchings may be done between other achievement tests to see if they show similar sensitivity. The relationship between the achievement subtest task and the parameters for establishing time limits is an interesting one. It calls into question the standard use of the 90% rule, a process that seems to be used more by convention than on the basis of solid research. It may be that looking at the trends of various time limits may be a more accurate way to establish efficient time limits.

Setting Time Limits

Rudman and Raudenbush suggested that the 90% rule for setting time limits for subtests of standardized achievement tests may no longer be precise enough, due to the high stake decisions that are based on outcomes of testing. Although this study did not affirm these findings, additional research may still need to be done to see if the "questions per minute" variable is significantly related to sensitivity to time. There may be a standardized test that comes closer to the same speed per question as the Reading Comprehension subtest of the Stanford Achievement Test; if so, this study should be pursued to see if the Rudman/Raudenbush (1986; 1987) can be replicated. If the results parallel the findings of this research, some other characteristic of test construction should be studied.

The manner in which time limits are set for standardized achievement tests should be questioned. The time limits for the Iowa Test of Basic Skills and the Stanford Achievement Test are both set using variations of the 90% rule. One test was sensitive to additional time; one was not, although results appear to be related to ability. Something is happening in the internal makeup of the subtests, themselves, to cause this difference. Until that factor is isolated, greater care must be taken in making high stake decisions on the basis of test scores.

Ability vs. Time

The results of this study in the Reading Comprehension subtest of the Iowa Test of Basic Skills suggest that any possible effect of increasing time may be different for students of different abilities. This was also suggested by Lord in 1956, Durso and Rubin in 1982, and Boag and Neid in 1962. The group effect in Reading Comprehension was significant to the .005 level of confidence and treatment by block interactive effect was significant to the .000 level of confidence (see Table 4. 5). This interaction effect is thought to be a random effect, possibly the result of tracking. This research could be replicated by changing the design so that each treatment be given to students in each group, and thereby eliminating the tracking or teacher effect found in this research.

Additional research needs to be done to see if alterations in testing time effects the outcome of students of different abilities in different ways. Does time to second guess or analyze an answer help average or better students more than poorer students? Do average students do better when they are allowed to slow down? Are less capable students more or less likely to benefit from more time? This study does not give a definitive answer, but does suggest further research.

APPENDICES

APPENDIX A

ALTERED DIRECTIONS FOR EXPERIMENTAL GROUPS

APPENDIX A

ALTERED DIRECTIONS FOR EXPERIMENTAL GROUPS

Control Group

ALTERNATE DIRECTIONS FOR TEST R:

READING COMPREHENSION

This test may be administered during the same session as the vocabulary test or in a later session. It requires 42 minutes of actual working time.

1. After the pupils are ready to begin work, say:

   "Now we are going to take a reading test. Find the section for Test R: Reading on your answer sheet. Open your test booklet to page 11. Read the directions silently while I read them aloud. Then say:

   "This test consists of several reading selections. After each selection there are some exercises. Read each selection quickly, and then answer the exercises. Four answers are given for each exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as

the exercise. Fill in the answer space for the best answer.

The sample exercise below shows you how to mark your answers on the answer sheet."

2. Allow pupils time to study the sample exercise. Then say:

"You will have 42 minutes for this test. If you finish early, you may go back over your work, but do not look at any of the other tests in the booklet. Now find your place to begin." (If only one level is being tested, say, "Turn to page 7 and begin with exercise 45.) (Pause) "Does everyone have the right place? (Pause) "Ready, go."

Observe the time and write it down.

_____ START

_____ STOP

3. Make a systematic check to make sure pupils are marking the right section of the answer sheet, that they are marking the right row of answer spaces, and that they are marking heavy black marks. Keep checking continually throughout the period. Give any necessary suggestion to pupils who do not seem to understand the directions, but, of course, do not help them answer any of the questions. During the latter part of the period, watch particularly that no one looks at any test other than Test R.

4. If considered desirable, especially in the lower grades, at the end of 25 minutes allow a brief stand-up period. However, make sure that the total time allowed for the test is exactly 42 minutes.
5. At the end of 42 minutes say:
   "Stop. Close your test booklet and place your answer sheet under the cover."
6. Collect the test booklets and answer sheets.

Experiment Group--Five Minutes

ALTERNATE DIRECTIONS FOR TEST R:

READING COMPREHENSION

This test may be administered during the same session as the vocabulary test or in a later session. It requires 47 minutes of actual working time.

1. After the pupils are ready to begin work, say:
   "Now we are going to take a reading test. Find the section for Test R: Reading on your answer sheet. Open your text booklet to page 11. Read the directions silently while I read them aloud. Then say:
   "This test consists of several reading selections. After each selection, there are some exercises.
   Read each selection quickly, and then answer the exercises. Four answers are given for each

exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise. Fill in the answer space for the best answer.

The sample exercise below shows you how to mark your answers on the answer sheet."

2. Allow pupils time to study the sample exercise. Then say:

"You will have 47 minutes for this test. If you finish early, you may go back over your work, but do not look at any of the other tests in the booklet. Now find your place to begin." (If only one level is being tested, say, "Turn to page 7, and begin with exercise 45.) (Pause) "Does everyone have the right place? (Pause) Ready, go."

Observe the time and write it down.

_____ START

_____ STOP

3. Make a systematic check to make sure pupils are marking the right section of the answer sheet, that they are marking the right row of answer spaces, and that they are marking heavy black marks. Keep checking continually throughout the period. Give any necessary suggestion to pupils who do not seem

to understand the directions, but, of course, do not help them answer any of the questions. During the latter part of the period, watch particularly that no one looks at any test other than Test R.

4. If considered desirable, especially in the lower grades, at the end of 25 minutes allow a brief stand-up period. However, make sure that the total time allowed for the test is exactly 47 minutes.
5. At the end of 47 minutes say:
   "Stop. Close your test booklet and place your answer sheet under the cover."
6. Collect the test booklets and answer sheets.

Experiment Group--Ten Minutes

ALTERNATE DIRECTIONS FOR TEST R:

READING COMPREHENSION

This test may be administered during the same session as the vocabulary test or in a later session. It requires 52 minutes of actual working time.

1. After the pupils are ready to begin work, say:
   "Now we are going to take a reading test. Find the section for Test R: Reading on your answer sheet. Open your text booklet to page 11. Read the directions silently while I read them aloud. Then say:

"This test consists of several reading selections. After each selection, there are some exercises. Read each selection quickly, and then answer the exercises. Four answers are given for each exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise. Fill in the answer space for the best answer.

The sample exercise below shows you how to mark your answers on the answer sheet."

2. Allow pupils time to study the sample exercise. Then say:

"You will have 52 minutes for this test. If you finish early, you may go back over your work, but do not look at any of the other tests in the booklet. Now find your place to begin." (If only one level is being tested, say, "Turn to page 7, and begin with exercise 45.) (Pause) "Does everyone have the right place? (Pause) Ready, go."

Observe the time and write it down.

_____ START

_____ STOP

3. Make a systematic check to make sure pupils are marking the right section of the answer sheet, that

they are marking the right row of answer spaces, and that they are marking heavy black marks. Keep checking continually throughout the period. Give any necessary suggestion to pupils who do not seem to understand the directions, but, of course, do not help them answer any of the questions. During the latter part of the period, watch particularly that no one looks at any test other than Test R.

4. If considered desirable, especially in the lower grades, at the end of 25 minutes allow a brief stand-up period. However, make sure that the total time allowed for the test is exactly 52 minutes.
5. At the end of 52 minutes say:
   "Stop. Close your test booklet and place your answer sheet under the cover."
6. Collect the test booklets and answer sheets.

Experiment Group--Fifteen Minutes

ALTERNATE DIRECTIONS FOR TEST R:

READING COMPREHENSION

This test may be administered during the same session as the vocabulary test or in a later session. It requires 57 minutes of actual working time.

1. After the pupils are ready to begin work, say:

"Now we are going to take a reading test. Find the section for Test R: Reading on your answer sheet. Open your text booklet to page 11. Read the directions silently while I read them aloud. Then say:

"This test consists of several reading selections. After each selection, there are some exercises. Read each selection quickly, and then answer the exercises. Four answers are given for each exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise. Fill in the answer space for the best answer.

The sample exercise below shows you how to mark your answers on the answer sheet."

2. Allow pupils time to study the sample exercise. Then say:

"You will have 57 minutes for this test. If you finish early, you may go back over your work, but do not look at any of the other tests in the booklet. Now find your place to begin." (If only one level is being tested, say, "Turn to page 7, and begin with exercise 45.) (Pause) "Does everyone have the right place? (Pause) Ready, go."

Observe the time and write it down.

_____ START

_____ STOP

3. Make a systematic check to make sure pupils are marking the right section of the answer sheet, that they are marking the right row of answer spaces, and that they are marking heavy black marks. Keep checking continually throughout the period. Give any necessary suggestion to pupils who do not seem to understand the directions, but, of course, do not help them answer any of the questions. During the latter part of the period, watch particularly that no one looks at any test other than Test R.

4. If considered desirable, especially in the lower grades, at the end of 25 minutes allow a brief stand-up period. However, make sure that the total time allowed for the test is exactly 57 minutes.

5. At the end of 57 minutes say:

   "Stop. Close your test booklet and place your answer sheet under the cover."

6. Collect the test booklets and answer sheets.

Control Group

TEST W-2 REFERENCE MATERIALS

(25 MINUTES)

1. As soon as pupils have been brought to attention, say:

"Now we are ready to take the test. Open your test booklet to page 75. (PAUSE) Find the section of your answer sheet for Test W-2: Reference Materials. (PAUSE) Read the directions on page 44 silently while I read them aloud. Then say:

"This is a test of study skills, such as looking up words in a dictionary, alphabetizing, using an index, and locating information.

Read the directions for each part carefully, and then mark your answers to the exercises on your answer sheet.

You will have 25 minutes for this test. Be sure that you mark your answers on the proper sections of your answer sheet. Now find your place to begin."

(If only one level is being tested, say: Turn to page 80, and being with exercise 34.) (PAUSE)

"Does everyone have the correct place? (PAUSE) Ready, go."

Observe the time and write it down.

START _____.

STOP _____.

2. Circulate among the pupils to make sure that they are marking their answer sheets properly.
3. At the end of 25 minutes, say:
   "Stop. Close your test booklet, and place your answer sheet under the cover."
4. Collect the test booklets and answer sheets.

Experimental Group--Five Minutes

TEST W-2 REFERENCE MATERIALS

(30 MINUTES)

1. As soon as pupils have been brought to attention, say:
   "Now we are ready to take the test. Open your test booklet to page 75. (PAUSE) Find the section of your answer sheet for Test W-2: Reference Materials. (PAUSE) Read the directions on page 44 silently while I read them aloud. Then say:
   "This is a test of study skills, such as looking up words in a dictionary, alphabetizing, using an index, and locating information.
   Read the directions for each part carefully, and then mark your answers to the exercises on your answer sheet.
   You will have 30 minutes for this test. Be sure that you mark your answers on the proper sections of your answer sheet. Now find your place to begin."

(If only one level is being tested, say: Turn to page 80, and being with exercise 34.) (PAUSE) "Does everyone have the correct place? (PAUSE) Ready, go."

Observe the time and write it down.

START _____.

STOP _____.

2. Circulate among the pupils to make sure that they are marking their answer sheets properly.
3. At the end of 30 minutes, say:

   "Stop. Close your test booklet, and place your answer sheet under the cover."
4. Collect the test booklets and answer sheets.

Experimental Group--Ten Minutes

TEST W-2 REFERENCE MATERIALS

(35 MINUTES)

1. As soon as pupils have been brought to attention, say:

   "Now we are ready to take the test. Open your test booklet to page 75. (PAUSE) Find the section of your answer sheet for Test W-2: Reference Materials. (PAUSE) Read the directions on page 44 silently while I read them aloud. Then say:

   "This is a test of study skills, such as looking up words in a dictionary, alphabetizing, using an index, and locating information.

Read the directions for each part carefully, and then mark your answers to the exercises on your answer sheet.

You will have 35 minutes for this test. Be sure that you mark your answers on the proper sections of your answer sheet. Now find your place to begin." (If only one level is being tested, say: Turn to page 80, and being with exercise 34.) (PAUSE) "Does everyone have the correct place? (PAUSE) Ready, go."

Observe the time and write it down.

START _____.

STOP _____.

2. Circulate among the pupils to make sure that they are marking their answer sheets properly.
3. At the end of 35 minutes, say:

   "Stop. Close your test booklet, and place your answer sheet under the cover."
4. Collect the test booklets and answer sheets.

Experimental Group--Fifteen Minutes

TEST W-2 REFERENCE MATERIALS

(40 MINUTES)

1. As soon as pupils have been brought to attention, say:

"Now we are ready to take the test. Open your test booklet to page 75. (PAUSE) Find the section of your answer sheet for Test W-2: Reference Materials. (PAUSE) Read the directions on page 44 silently while I read them aloud. Then say:

"This is a test of study skills, such as looking up words in a dictionary, alphabetizing, using an index, and locating information.

Read the directions for each part carefully, and then mark your answers to the exercises on your answer sheet.

You will have 40 minutes for this test. Be sure that you mark your answers on the proper sections of your answer sheet. Now find your place to begin."

(If only one level is being tested, say: Turn to page 80, and being with exercise 34.) (PAUSE)

"Does everyone have the correct place? (PAUSE) Ready, go."

Observe the time and write it down.

START _____.

STOP _____.

2. Circulate among the pupils to make sure that they are marking their answer sheets properly.
3. At the end of 40 minutes, say:
   "Stop. Close your test booklet, and place your answer sheet under the cover."
4. Collect the test booklets and answer sheets.

APPENDIX B

DEMOGRAPHICS AND POSTTEST STATISTICS BY HOMEROOM

Table B.1

Demographic Variables Codes

Variable	Code
Sex	
Male	0
Female	1
Ethnicity	
White	0
Non-White	1 or 2
Family Configuration	
One Parent	1
Two Parents	2
Chapter I	
No Chapter I	0
Receiving Chapter I	1

Table B.2

Raw Data Group 1: Five Minute Treatment, Block 1.

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
10	H-01	1	0	0	2	39	43	34
35	H-01	1	0	0	2	43	47	39
60	H-01	0	0	0	2	49	54	40
65	H-01	1	0	0	2	35	41	35
99	H-01	1	2	0	2	38	41	34
100	H-01	1	0	0	2	45	41	34
105	H-01	1	0	0	2	47	49	--
124	H-01	1	0	0	2	44	50	37
128	H-01	0	0	0	2	48	52	38
161	H-01	0	0	0	2	45	48	38
167	H-01	0	0	0	2	39	47	37
180	H-01	0	0	0	2	44	48	41
184	H-01	0	0	0	2	46	48	38
202	H-01	1	0	0	2	44	42	--
203	H-01	1	0	0	2	42	50	37
204	H-01	0	0	0	2	36	46	36
205	H-01	1	0	0	2	48	47	37
206	H-01	1	0	0	2	40	51	35
207	H-01	1	0	0	2	25	36	32
208	H-01	0	0	0	2	25	46	39
209	H-01	0	0	0	2	25	31	29
210	H-01	0	0	0	2	29	41	--
211	H-01	0	0	0	2	31	36	35
238	H-01	0	0	0	2	31	44	39
243	H-01	0	2	0	2	25	39	34
246	H-01	0	0	0	2	52	53	40

Table B.3

Raw Data Group 2: Fifteen Mintue Treatmetn Block 1.

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
6	H-02	1	2	0	2	36	50	35
27	H-02	1	1	0	1	27	39	26
61	H-02	1	0	0	2	41	44	35
68	H-02	1	0	0	2	42	45	30
80	H-02	1	0	0	2	36	45	27
84	H-02	1	0	0	2	48	50	28
85	H-02	1	0	0	2	30	43	31
91	H-02	1	0	0	2	39	42	32
94	H-02	1	0	0	1	44	47	32
101	H-02	1	2	0	1	38	33	24
102	H-02	1	1	0	2	20	37	30
108	H-02	1	1	1	2	18	36	21
109	H-02	1	0	0	2	33	39	33
110	H-02	1	0	0	2	38	45	38
115	H-02	1	0	1	1	15	42	8
118	H-02	1	0	0	2	42	49	38
123	H-02	1	0	0	2	36	47	32
143	H-02	0	0	0	2	32	34	34
146	H-02	0	0	0	1	36	44	31
150	H-02	0	0	0	2	42	47	30
160	H-02	0	0	0	2	34	35	36
164	H-02	0	0	0	2	38	48	25
177	H-02	0	0	0	2	39	52	26
188	H-02	0	0	0	2	40	43	37
192	H-02	0	0	0	2	26	35	37
232	H-02	1	1	1	2	28	42	16
255	H-02	1	2	0	1	41	41	39

Table B.4

Raw Data Group 3: Control, Block 1

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
17	F-03	1	0	0	2	41	45	37
28	F-03	1	0	0	2	36	28	18
42	F-03	1	0	0	2	37	45	28
59	F-03	0	2	0	2	16	26	19
70	F-03	1	0	0	2	43	47	37
74	F-03	1	1	0	2	11	15	11
78	F-03	1	0	0	2	25	32	24
95	F-03	1	2	0	2	33	35	33
104	F-03	1	0	0	2	51	50	39
122	F-03	1	1	0	2	21	22	21
139	F-03	0	0	0	2	28	41	32
141	F-03	0	0	0	2	41	49	35
148	F-03	0	0	0	2	41	42	36
149	F-03	0	0	0	2	41	47	35
152	F-03	0	0	0	2	32	34	30
155	F-03	0	0	0	2	39	37	29
163	F-03	0	1	0	2	21	24	28
182	F-03	0	0	0	2	45	50	38
185	F-03	0	0	0	2	33	38	32
219	F-03	0	0	0	2	30	41	30
228	F-03	1	0	0	2	17	13	13
237	F-03	1	0	0	2	45	50	32
245	F-03	1	0	0	2	35	50	36
261	F-03	1	0	0	1	34	35	28

Table B.5

Raw Data Group 4: Ten Minute Treatment, Block 1

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
31	F-04	1	0	0	2	25	23	8
34	F-04	1	0	0	2	20	30	21
44	F-04	0	0	0	2	50	53	40
54	F-04	0	1	0	1	36	40	31
63	F-04	1	0	0	2	37	44	30
82	F-04	1	0	0	2	34	28	30
106	F-04	1	0	0	2	33	43	33
113	F-04	1	0	0	2	33	38	37
114	F-04	1	0	0	2	39	42	32
119	F-04	1	2	0	2	33	40	33
120	F-04	1	0	0	2	32	40	17
125	F-04	0	0	0	2	29	35	31
136	F-04	0	0	0	2	44	41	30
138	F-04	0	1	0	2	26	39	28
153	F-04	0	0	0	2	31	49	38
169	F-04	0	0	0	1	35	44	25
178	F-04	0	0	0	2	33	34	33
183	F-04	0	0	0	2	28	32	29
187	F-04	0	0	0	2	39	44	34
191	F-04	0	0	0	2	33	47	34
194	F-04	0	0	0	2	28	33	27
217	F-04	0	0	0	2	37	34	36
2240	F-04	0	0	0	2	46	49	35
241	F-04	1	1	0	1	34	40	22
249	F-04	0	1	0	2	29	35	25
253	F-04	1	0	0	2	46	43	37
254	F-04	0	0	0	1	29	36	27
264	F-04	1	1	0	1	36	44	33
266	F-04	1	0	0	2	36	45	39

Table B.6

Raw Data Group 5: Ten Minute Treatment, Block 2.

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
29	H-05	1	2	0	2	25	29	32
57	H-05	0	0	1	1	14	30	14
62	H-05	1	0	0	2	27	26	37
66	H-05	1	0	0	2	31	35	25
69	H-05	1	0	0	1	41	46	37
76	H-05	1	0	0	2	44	49	33
77	H-05	1	0	0	2	40	34	35
112	H-05	1	2	0	2	26	24	30
116	H-05	1	2	0	2	35	39	33
131	H-05	0	0	0	1	50	51	36
134	H-05	0	0	0	2	45	43	33
151	H-05	0	0	1	1	18	21	--
165	H-05	0	0	1	2	16	24	27
172	H-05	0	1	0	2	30	31	29
173	H-05	0	2	0	2	27	28	25
189	H-05	0	1	0	1	44	49	32
190	H-05	0	0	0	2	32	32	27
196	H-05	0	0	0	2	39	32	37
215	H-05	1	0	1	2	20	24	25
222	H-05	1	0	0	1	21	36	24
224	H-05	1	0	0	2	33	39	28
262	H-05	1	0	0	2	29	40	27
265	H-05	1	0	0	1	34	35	33
271	H-05	0	1	1	1	17	34	29
274	H-05	0	0	1	1	17	20	16

Table B.7

Raw Data Group 6: Control, Block 2

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
2	F-06	1	0	0	2	34	33	14
14	F-06	1	0	0	1	43	19	28
18	F-06	1	0	0	2	31	35	31
23	F-06	1	1	1	1	24	22	18
48	F-06	0	0	1	2	26	13	21
71	F-06	1	0	0	2	37	33	30
87	F-06	1	0	0	2	24	22	30
107	F-06	1	0	0	2	36	30	29
111	F-06	1	2	0	2	29	33	33
142	F-06	0	0	0	1	28	23	16
145	F-06	0	2	1	1	14	26	21
156	F-06	0	1	0	1	30	32	28
157	F-06	0	0	0	2	32	28	16
162	F-06	0	0	0	1	26	39	28
186	F-06	0	1	0	1	23	22	33
213	F-06	1	0	0	2	28	36	28
220	F-06	1	0	0	1	29	38	30
227	F-06	0	0	1	1	19	21	23
235	F-06	0	1	0	2	29	31	24
239	F-06	0	0	0	1	25	31	24
250	F-06	0	0	0	1	21	19	24
256	F-06	0	0	0	2	29	30	23
269	F-06	0	2	0	1	23	25	29

Table B.8

Raw Data Group 7: Fifteen Minute Treatment, Block 2.

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
7	H-07	1	0	1	2	20	33	21
11	H-07	1	0	1	2	13	17	24
40	H-07	1	0	1	1	11	11	18
43	H-07	0	1	0	2	38	49	41
51	H-07	0	0	1	2	21	24	13
52	H-07	0	0	1	1	41	46	33
53	H-07	0	0	1	2	11	23	17
72	H-07	1	2	0	2	24	22	23
88	H-07	1	0	1	2	16	14	15
93	H-07	1	1	1	2	24	24	--
96	H-07	1	0	0	1	19	22	33
97	H-07	1	2	0	2	28	35	30
129	H-07	0	0	1	2	42	48	38
133	H-07	0	1	1	1	23	28	28
144	H-07	0	0	1	1	32	43	33
170	H-07	0	0	0	2	37	32	30
198	H-07	0	0	0	2	48	42	39
229	H-07	1	2	1	1	37	41	24
233	H-07	1	0	0	2	37	42	27
242	H-07	0	0	0	2	45	41	34
247	H-07	0	1	1	1	17	23	18
257	H-07	0	0	0	2	47	48	34
272	H-07	0	0	0	1	23	32	35
273	H-07	0	0	0	2	40	38	37

Table B.9

Raw Data Group 8: Five Minute Treatment, Block 2.

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
12	F-06	1	0	0	2	30	29	--
13	F-06	1	0	0	2	25	30	10
30	F-06	1	0	0	2	28	32	28
32	F-06	1	0	0	2	50	50	34
33	F-06	1	1	0	1	30	22	16
49	F-06	0	2	1	1	17	27	29
73	F-06	1	0	1	2	18	25	27
86	F-06	1	0	0	1	39	36	34
89	F-06	1	1	1	2	21	21	11
127	F-06	0	0	0	2	19	12	--
168	F-06	0	0	0	2	21	24	30
193	F-06	0	1	0	1	29	21	27
197	F-06	0	1	0	2	44	47	36
212	F-06	1	0	0	2	48	49	34
218	F-06	1	0	0	2	25	45	28
221	F-06	0	1	1	1	14	--	10
225	F-06	0	0	0	1	24	33	33
236	F-06	1	0	0	2	30	43	29
244	F-06	0	0	0	1	26	34	28
252	F-06	0	0	0	1	8	18	20
260	F-06	0	0	0	2	28	42	30
267	F-06	1	0	0	1	32	29	34
270	F-06	1	0	0	2	16	22	21

Table B.10

Raw Data Group 9: Fifteen Minute Treatment, Block 3.

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
1	F-09	1	0	0	2	16	24	21
20	F-09	1	1	0	1	23	21	26
24	F-09	1	0	0	2	17	16	14
26	F-09	1	1	0	2	29	30	16
37	F-09	1	0	0	2	15	20	13
38	F-09	1	1	0	1	22	30	24
50	F-09	0	0	0	2	16	20	26
55	F-09	0	0	0	1	29	34	34
67	F-09	1	0	0	2	21	16	28
92	F-09	1	1	0	2	36	41	27
137	F-09	0	0	0	1	34	42	32
154	F-09	0	0	0	2	30	33	31
195	F-09	0	1	0	2	47	47	36
263	F-09	0	0	0	2	18	21	21
268	F-09	1	0	0	2	28	33	32

Table B.11

Raw Data Group 10: Five Minute Treatmetn, Block 3

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
3	F-10	1	0	1	2	12	18	13
15	F-10	1	0	1	1	20	19	15
16	F-10	1	0	1	2	12	17	9
25	F-10	1	1	1	1	34	11	10
36	F-10	1	0	0	2	26	25	12
41	F-10	1	1	1	2	12	12	8
64	F-10	1	0	1	2	8	18	13
75	F-10	1	0	1	2	20	32	27
98	F-10	1	0	1	2	23	21	35
103	F-10	1	1	1	2	15	13	30
126	F-10	0	1	0	2	21	23	22
147	F-10	0	0	1	1	25	26	14
174	F-10	0	0	0	2	21	12	11
176	F-10	0	0	0	2	23	21	33

Table B.12

Raw Data Group 11: Ten Minute Treatment, Block 3

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
8	H-11	0	0	1	1	28	25	21
9	H-11	0	1	1	1	18	18	14
19	H-11	1	1	1	1	20	39	24
21	H-11	1	0	1	2	33	39	26
22	H-11	1	0	1	2	33	34	24
39	H-11	1	0	1	2	20	30	20
45	H-11	0	0	0	2	47	53	36
46	H-11	0	0	1	1	27	32	32
79	H-11	1	0	1	2	8	20	16
83	H-11	1	0	0	1	33	40	35
121	H-11	1	0	0	2	12	39	30
140	H-11	0	0	0	2	37	37	12
158	H-11	0	0	0	2	21	36	31
175	H-11	0	0	0	1	35	44	34
179	H-11	0	0	0	1	28	39	28
181	H-11	0	0	0	1	35	40	25
201	H-11	0	0	0	1	30	34	29
223	H-11	0	0	1	1	25	37	18
230	H-11	0	0	0	1	33	38	27
231	H-11	0	0	1	1	24	32	24
251	H-11	1	1	1	1	20	27	17
259	H-11	0	1	1	1	21	27	31

Table B.13

Group Date Group 12: Control, Block 3

St No.	Group Code	Sex	Race	Chi	Fam	Pre Test Read	Post Read	Post Ref
4	H-12	1	0	0	1	26	37	31
5	H-12	1	0	1	2	32	--	24
47	H-12	0	1	0	1	23	25	33
56	H-12	0	0	0	2	37	35	29
58	H-12	0	1	1	1	19	22	24
81	H-12	1	0	0	2	42	49	34
90	H-12	1	1	1	2	8	16	15
117	H-12	1	0	0	2	34	39	33
130	H-12	0	0	1	2	18	38	25
132	H-12	0	0	1	2	17	25	28
135	H-12	0	0	0	2	34	37	29
159	H-12	0	0	0	1	23	30	21
166	H-12	0	0	0	2	25	28	34
171	H-12	0	1	1	1	22	29	26
199	H-12	0	1	0	1	31	38	28
200	H-12	0	0	0	2	26	28	24
214	H-12	1	0	0	2	30	31	32
216	H-12	1	0	1	2	21	--	13
226	H-12	1	0	1	2	21	22	16
234	H-12	0	0	0	1	27	33	29
248	H-12	1	0	0	1	42	39	33
258	H-12	0	0	0	1	11	15	10
275	H-12	0	0	0	1	36	43	35

LIST OF REFERENCES

LIST OF REFERENCES

Anderson, B., & Odden, A. (1986, April). State initiatives can foster school improvement. Phi Delta Kappan, pp. 578-581.

Anderson, B., & Pipho, C. (1984, November). State-mandated testing and the fate of local control. Phi Delta Kappan, pp. 209-212.

Better Education for Michigan Citizens, A Blueprint for Action. (1984). Lansing: State of Michigan.

Boag, A. K., & Neild, M. (1962). The influence of the time factor on the scores of the Triggs Diagnostic Reading Test as reflected in the performance of secondary school pupils grouped according to ability. The Journal of Educational Research, 55, 181-183.

Bracey, G. W. (1987, May). Measurement-driven instruction: Catchy phrase, dangerous practice. Phi Delat Kappan, pp. 683-686.

Brookover, W. B., & Erickson, E. L. (1975). Sociology of education. Homewood, IL: The Dorsey Press.

Cronbach, L. J., & Warrington, W. G. (1951). Time-limit tests: Their reliability and degree of speeding. Psychometrika, 16, 167-188.

Daly, J. L., & Stahmann, R. F. (1968). The effect of time limits on a university placement test. Journal fo Educarional Research, 62, 103-104.

Doyle, D. P., & Hartle, T. W. (1985, September). Leadership in education: Governors, legislators, and teachers. Phi Delta Kappan, pp. 21-27.

Duke, D. L. (1985, June). What is the nature of educational excellence and should we try to measure it? Phi Delta Kappan, pp. 671-674.

Evans, F. R., & Reilly, R. R. (L1972). A study of speededness as a source of test bias. Journal of Educational Measurement, 9, 123-131.

Evans, M. C. (1985, March). Is state legislation a catalyst for changgee? Phi Delta Kappan, pp. 501-503.

Fuhrman, S., Huddle, E., & Armstrong, J. (1986, April). Improving schools: The state role. Phi Delta Kappan, pp. 594-596.

Gable, R., & Rogers, V.B. (1987, May). Taking the terror out of research. Phi Delta Kappan, pp. 690-695.

Goodlad, J. I. (1984). A place called school. New York: McGraw Hill Book Company.

Hawley, W. D. (1985, November). False premises, false promises: The mythical character of public discourse about education. Phi Delta Kappan, pp. 183-187.

Hieronymus, A. N., & Hoover, H. D. (1986a). Manual for school administrators, levels 5-14, ITBS forms G/H. Chicago: The Riverside Publishing Company.

Hieronymus, A. N., & Hoover, H. D. (1986b). Preliminary technical summary, the New Riverside Basic Skills Assessment Program K-12. Chicago: The Riverside Publishing Company.

Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. American Educational Reserach Journal, 19, 5-18.

Hunter, M. (1978). Improved instruction. El Segundo, CA: Tip Publications.

Kaplan, G. R. (1985, September). Shining lights in high places: Education's top four leaders and their heirs. Phi Delta Kappan, pp. 7-16.

Kendall, L. M. (L1964). The effects of varying time limits on test validity. Educational and Psychological Measurement, 24, pp. 789-800.

Kirst, M. W. (1986, January). Sustaining the momentum of state education reform: The link between assessment and financial support. Phi Delta Kappan, pp. 341-345.

Lieberman, M. (1985, October). Educational specialty boards: A way out of the merit pay morass. Phi Delta Kappan, pp. 103-107.

Lord, F. M. (1956). A study of speed factors in tests and academic grades. Psychometrika, 21, pp. 31-50.

Madaus, G. F. (1985, May). Test scores as adminsitrative mechanisms in educational policy. Phi Delta Kappan, pp. 103-107.

Mueller, D. J., & Wasser, V. (1977). Implications of changing answers of objective test items. Journal of Educational Measurement, 14, pp. 9-14.

Myers, C. T. (1952). The factorial composition and validity of differently speeded tests. Psychometrika, 17, pp. 347-352.

Nunnally, J. (1978). Psychometric theory. New York: McGraw Hill Book Company.

Odden, A., & Anderson, B. (1986, April). How successful state education improvement programs work. Phil Delta Kappan, pp. 582-585.

Pipho, C. (1986, December). State move reform closer to reality. Phi Delta Kappan, pp. 1-8.

Popham, W. J. (1987, May). The merits of measurement-driven instruction. Phi Delta Kappan, pp. 1-8.

Popham, W. J. (1987, May). Muddle-minded emotionalism. Phi Delta Kappan, pp. 687-689.

Popham, W. J., Cruse, K. L., Rankin, S. C., Sandifer, P. D., & Williams, P. L. (1985, May). Measurement-driven instruction: It's on the road. Phi Delta Kappan, pp. 628-634.

Rehberg, R. A., & Rosenthal, E. R. (1978). Class and merit in the American high school. An assessment of the Revisionist and Meritocratic Arguments. New York: Longman.

Resnick, D. P., & Resniccik, L. B. (1983, November). Improving education standards in American schools. Phi Delta Kappan, pp. 178-180.

Rosenthal, R., & Rubin, D. B . (1982). A simple, general purpose display of magnitude of experimental effect. Journal of Educational Psychology, 7, pp. 166-169.

Rudman, H. C., & Raudenbush, S. W. (1986). The effect of exceeding prescribed time limits in the administration of standardized achievement tests: An abstract. Unpublished manuscript. East Lansing, MI: Michigan State University.

Rudman, H. C., & Raudenbush, S. W. (1987). The effect of exceeding prescribed time limits in the administration of standardized test of reading comprehension applications. Unpublished manuscript. East Lansing, MI: Michigan State University.

Salganik, L. H. (1985, May). Why testing reforms are so popular and how they are changing education. Phi Delta Kappan, pp. 607-610.

Schlechty, P. C., & Ingwerson, D. W. (1987, April). A proposed incenitve system for Jefferson County tachers. Phi Delta Kappan, pp. 585-590.

Stedman, L. C., & Kaestle, C. F. (1985, November). The tet scooreer decline is over: Now what? Phi Delta Kappan, pp. 204-210.

Stellman, L. C., & Powell, B. (L1985, May). Appraising the implications of the SAT for educational policy. Phi Delta Kappan, pp. 603-606.

Sutherland, JJ. D. (1934). The speed factor in intelligent reactions. British Journal of Psychology, 24, pp. 276-294.

Wild, C. L., Durso, R., & Rubin, D. B. (1982). Effect of increased test-taking time on test score by ethnic group, years out of school, and sex. Journal of Educational Measurement, 19, pp. 19-28.