

Testing The General Achievement Gap Reduction Hypothesis: A Ten Year View Of Ohio Proficiency Testing

Rich Hofmann
Miami University
V2.5

Abstract

Six hundred four public school districts in Ohio are grouped according to poverty into low (N=60), middle (N=484), and high (n=60) groups. Using effect size correlations the differential performances of these school districts on proficiency subtests (Writing, Reading, Math, Citizenship, and Science) for grades four, six, and nine are compared and contrasted by various year groupings representing the beginning and ending years (1990 through 1999) of Ohio Proficiency Testing.

Using effect size correlations large poverty related achievement gaps are identified at every grade level in almost every subject area for both beginning and ending year groupings. Comparisons of various composite effect size correlations suggest that the poverty related achievement gaps are being reduced across all fourth grade subtests, and three of four ninth grade subtests. The poverty related achievement gaps have not been reduced for the sixth grade subtests.

These implications for these findings are discussed from several different perspectives: positive and negative practitioner perspectives and positive and negative perspectives that might be expected from testing experts.

Paper presented at the annual meeting of the American Educational Research Association, New Orleans-April 2000.

Address Correspondence to:
Rich Hofmann
Educational Leadership
Miami University
Oxford, Oh 45056

Running Head: Testing Achievement Gap Reduction Hypothesis

Minimal Competency Testing

Minimal competency testing (MCT), by its very nature, is intended to assess the minimal competencies of students, generally in some or all of the traditional academic areas of Writing, Reading, Math, Science, and social studies. MCTs are given across grade levels that vary from state-to-state. Scores on the tests are usually reported as pass-fail. A large majority of those who do not pass MCTs tend to be students who are recognized as “low achievers”. Coleman’s work (Coleman 1966) called attention to the fact that parental demographics were better predictors of school achievement (school outcomes) than the more traditional school variables of class size, teacher training, and general school facilities (school inputs). Since Coleman’s study there have been many studies documenting the influence of family background on student achievement in general and standardized achievement tests in particular.

Popham (1981) as a proponent of MCT noted that ,”Early returns suggest that proportionally more minority students than others are failing minimum competency tests.” Early opponents of MTC in 1981 cited such findings as evidence of racial discrimination (see Thurston & House 1981). Popham (1981) asserted that accusations of racial discrimination “ignores the fact that minority students are currently receiving inadequate educations in far too many U.S. Classrooms.” Furthermore, he argued that by “systematically detecting and eliminating basic skill deficiencies” as identified by MCTs, U.S. public schooling will be improved with the greatest effects accruing for the low achieving students, a major argument of the MCT proponents. Twelve years later Airasian (1993) indicated that there is evidence that the proportion of students passing MCTs ,”has increased greatly from the first to the second or third year“ of testing. Nineteen years later Linn (2000) cautions us that experiences with MCT results suggest that we will see major gains in the first several years of MCT use and then a tapering off of the gains.

The MCT Promise: The Achievement Gap Reduction Hypothesis

A general achievement gap is represented by the difference in achievement performance between low scoring demographic groups and high scoring demographic groups. The promise of MCT is that those demographic groups represented by the failures in the early years of MCT will receive the benefits of attendant instruction in the deficit areas identified by the MCT and in subsequent years will show increased passing percentages at a faster rate than the rest of the population. Basically this means that the difference in passing rates between historically low scoring demographic groups and the non-low scoring groups, an achievement gap, is expected to become smaller. Thus, the general *achievement gap reduction hypothesis*.

Mandated minimum competency testing will, over time, reduce the achievement gap associated with differential "demographic grouping variables."

Discussions about increased passing rates or score gains do not address the issue of achievement gaps. If the achievement performances of two different demographic groups defining an achievement gap increase at the same rate then the achievement gap will not be reduced. In order for the gap to be reduced the lower scoring demographic group must show a more rapid improvement in achievement than the higher scoring demographic group.

Reference to differential demographic grouping is a term that encompasses the various groupings that researchers have associated with poor performance on MCTs. The language regarding differential demographic groupings is frequently confusing. Many researchers tend to confuse poverty and race. The low MCT achievers usually represent identifiable groups in school districts. Both proponents and opponents of MCT discussing these low scoring groups, since about 1975, refer to them as "students from minority and low-income backgrounds", "various racial and language minorities," or some variation of "racial or economic

minorities” , while others speak of the “discrepancy between ... performances of white and minority students.” In previous work with the Ohio proficiency tests (Hofmann 1997, 1999) it has been demonstrated that the racial make-up of a school districts predict a negligible (a small non-significant) proportion of school district achievement variation after the variation associated with poverty has been removed , see also Payne & Biddle (1999) and Berliner & Biddle (1998). It is reasonable to assume that the racial make-up of a community (in the form of percent non-white members of the community) is a proxy variable for community poverty, but if one has good indicators of community poverty these indicators will be better predictors of differential school district achievement than the racial make-up of the community.

This study will use aggregate data, percent pass for a whole school district. The school districts will be grouped into high, middle, and low poverty. This study will provide several related tests of a more specific achievement deficit reduction hypothesis.

Mandated minimum competency testing, in the form of Ohio Proficiency Testing will, over time, reduce the achievement gap defined by public school districts having differential community poverty.

The purpose of this study is two-fold, to describe poverty related achievement gaps defined by differential Ohio public school district performance on Ohio Proficiency Tests and to describe changes in the achievement gap from the inception of proficiency testing in 1990 to the present, 1999.

METHODS

Sampling Unit and Population

In this study the sampling unit is an Ohio public school district. The number of public school districts in Ohio vary from approximately 606 to 611 between the inclusive years of 1990 to 1999. For this

particular study the included school districts, 604, were those school districts that participated in all proficiency testing from 1990 to 1999, with the exclusion of the “Island Districts”. The Island Districts are atypical (no high school and very small class sizes) school districts associated with some of the Ohio islands in Lake Erie.

Ohio Proficiency Testing

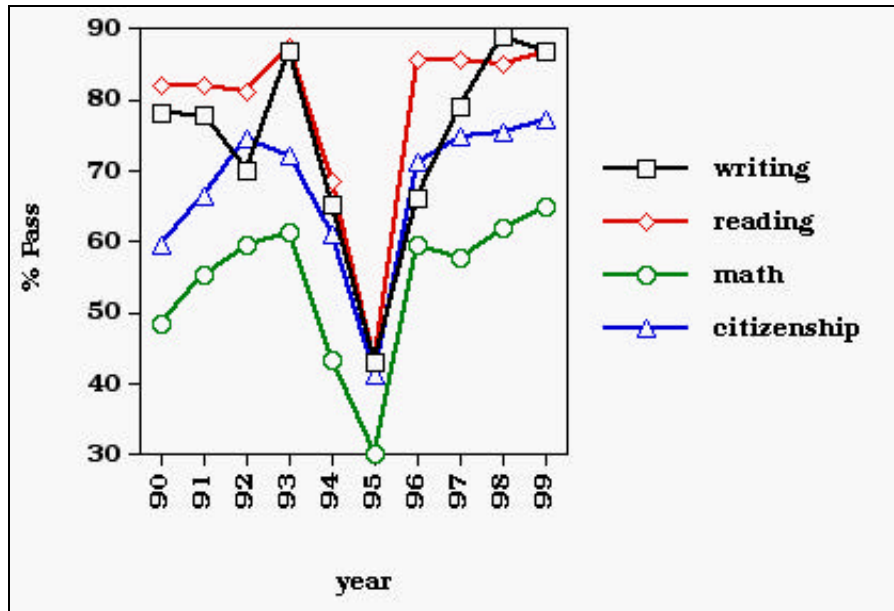
The state of Ohio mandated state-wide minimum competency testing for all public school ninth grade students in 1990, the Ohio Ninth Grade Proficiency Tests. In 1995 the state allowed eighth grade students to take the ninth grade proficiency tests and mandated minimum competency testing for all public school fourth grade students, the Ohio Fourth Grade Proficiency Tests. In 1996 the state mandated minimum competency testing for all public school sixth grade students, the Ohio Sixth Grade Proficiency Tests. During the past ten years Ohio has revised their proficiency tests on an annual basis. This study will use the fourth, sixth, and ninth grade proficiency tests¹.

Initially there were four subtests defining the ninth and fourth grade proficiency tests: Writing, Reading, Math, and Citizenship. In 1996 a fifth subtest was added, Science. District data are available from state data bases through the Ohio Department of Education website (http://www.ode.state.oh.us/rc_download.htm). Presently the “proficiency results“ are reported directly to the school districts and parents of the children in the school districts through the use of a school district report card issued by the Ohio Department of Education. Newspapers provide comparative summaries of the report cards for the school districts within the paper's readership area and usually obtain their data from the website.

Comparative Problems With The Ohio Ninth Grade Proficiency Tests.

Each public school district receives an aggregate percent pass score for each proficiency subtest at each grade level, grades 4, 6, and 9 . For purposes of clarifying certain comparative problems the mean aggregate %pass “scores” for the Ninth Grade Proficiency tests across all school districts by subtest within test year are represented in Figure 1.

Figure 1 Ohio Ninth Grade Proficiency subtest % pass means by year from 1990 to 1999.



Several problems associated with attempts to compare performance on the Ohio proficiency subtests (grades 4, 6, and 9) are represented in Figure 1. The subtests within grade level are not of equal difficulty within any testing year. Because of this inequality of subtest difficulty it is not reasonable to assume that the yearly relative strengths and weaknesses of a school district might be represented by the district's %pass profile across subtests (within grades 4, 6, and 9). Following the mean profile for any subtest across the years indicates that the difficulty of the subtests vary from one year to the next (within grades 4, 6, and 9). This suggests that school district improvement or lack of improvement should not be determined by comparing "this year's test results" with "last year's test results" (within grades 4, 6, and 9). Documenting beneficial school district change is especially problematic when the assessment instruments and their associated criteria for passing change on an annual basis, as they do with the Ohio Proficiency Tests.

The years from 1994 to 1996 were problematic years for the Ohio Ninth Grade Proficiency Tests. The 1994 proficiency tests results appear

to be quite low. This year was associated with major scoring changes and format changes in the proficiency subtests. The 1995 proficiency test year was a low scoring year for all subtests, the lowest in the ten year testing period. It was the first year that eighth grade students were allowed to take the test. For the ninth grade proficiency test the two middle years of proficiency testing, 1994 and 1995, provide atypical test results.

School District Poverty

Poverty as it is used in this study represents community poverty, but within the context of school achievement, community poverty is actually child poverty since it is the children of poverty who attend the schools and the effects of community poverty on the children are reflected by their collective school achievement. Elsewhere (Hofmann 1999) the index of community poverty ² used in this study was developed using indicators of poverty obtained from 1990 community census data reported from the *School District Data Book* (SDDB), % of single parent households and % of adults in the community without a high school education. An additional indicator of community poverty is % of children receiving aid to dependent children, obtained from the 1990 *Ohio Educational Management Information Systems* (EMIS) data base available from the Ohio Department of Education website (http://www.ode.ohio.gov/www/ims/extract_ode_data.html). All indicators were based on 1990 data.

Conditions of community poverty are associated with ever present **undesirable events** such as home evictions (for some families every two to three months), physical illness, criminal assault, teenage pregnancy, child neglect, child abuse and ever present **undesirable conditions** such as easy access to drugs, inadequate housing, poor health care, dangerous neighborhoods, and even environmental toxins (see McLoyd 1997, 1998.) The index of community poverty represents a measure of the extent to which the presence of the above noted undesirable events and conditions

exist within a community, and also the extent to which there is child poverty in the school district associated with the community. All school districts in Ohio have “pockets of poverty.”

Child poverty at a community level is a long term condition which carries with it immense ramifications for the emotional and physical health of the school children coming from poverty conditions. From the perspective of a school district, an index of community poverty, and therefore child poverty, is generally indicative of the extent to which the children of the community will have difficulty with measures of school achievement. The higher the index of community poverty the lower the aggregate achievement of the school district.

The poverty index ranges from approximately zero to one with a mean of 50 and a standard deviation of 8.55. This index has excellent consistency as a long term predictor of school district proficiency performance with median proficiency subtest correlations of approximately .60 from 1990 to 1999 (see Hofmann, 1999).

For the purposes of this study the public school districts in Ohio were subdivided into three poverty groups: those school districts representing the highest 10 percent of the poverty index schools define high poverty school districts, the **high poverty group**; those representing the lowest 10 percent of the poverty index schools define the low poverty school districts, the **low poverty group**; the remaining 80 percent of the schools define the middle poverty school districts, the **middle poverty group**. In the discussion of the ten years of proficiency testing to follow the membership in these three groups remains the same over the full ten years. The means and standard deviations for poverty by poverty group are reported in Table 1.

Table 1. Means and standard deviations for poverty by poverty group

poverty group	mean	standard dev.	N
low	38.2	2.7	60
middle	49.2	4.8	484
high	68.5	7.6	60

These three groups differ significantly ($F(2,601)=591.62, p<.01$) with the low poverty group mean significantly lower than the middle poverty group and high poverty group means and the high poverty group mean significantly higher than the middle and low poverty group means. The mean of the middle poverty group is just below the poverty mean for all public schools in the state of Ohio, .50. The low group is the most homogeneous and the high group is the most heterogeneous of the three groups with regard to community poverty. In the discussion to follow these three poverty groups serve as the grouping variable for all comparisons. An achievement gap will be defined as the difference between the average %pass for two different poverty groups. The achievement gaps will be defined on the basis of proficiency subtest differences between these three groups of schools – three differences or contrasts: differences between low and high poverty groups, between low and middle poverty groups, and between middle and high poverty groups.

Making Comparisons Over Time Across Tests Between Grade Levels: A Use Of Effect Size Correlations

The use of effect size circumvents the problems associated with making comparisons between subtest within and across years. Effect size may be used to summarize and compare poverty group differences on proficiency subtests within years, across years, within grade level and across grade levels (see Cohen 1988; Cohen 1994; Rosenthal 1991; Rozeboom 1997; Schmidt and Hunter 1997).

Use of effect size will allow a mean proficiency test difference between two poverty groups on a specific proficiency subtest at a specific grade level to be standardized such that the standardized mean difference from one poverty group comparison may be compared to the standardized mean difference involving a different poverty group comparison, on the same subtest or a different subtest, in the same year or a different year, at the same grade level or a different grade level. The standardized

difference as used in this study is referred to as an *effect size correlation*³, henceforth ESC. ESCs can also be manipulated statistically⁴.

The interpretation of an effect size correlation, ESC, is unambiguous. An ESC may be thought of as indicating the size of the achievement gap represented by the associated proficiency subtest mean difference between the two poverty groups associated with the difference. The greater the ESC the greater the achievement gap. While it may range from plus to minus one, in this study it is, with one exception, always be positive. Following Cohen (1988, p.25) and Rosenthal (1991, p. 20) the magnitude of ESCs may be qualified by ranges. ESCs less than 0.10 but greater than or equal to 0.0 are **very small** and probably not meaningful. ESCs greater than or equal to 0.10 and less than 0.24 are **small** and are probably associated with small meaningful mean differences, a meaningful effect of poverty group membership on proficiency subtest performance. ESCs of 0.24 and greater but less than .37 are **medium** ESCs and are associated with meaningful medium mean differences. ESCs of .37 or larger are **large** effect sizes. and are associated with large, substantive and meaningful mean differences. Although it is possible to have an ESC of 1.00, it would be unusual, to have very large effect sizes greater than $r=.70$ and $.80$ ⁵.

For the purposes of this study the proficiency test results are considered as two groups: the *beginning years* and the *final years*. For the ninth grade tests the *beginning years* are represented by the years from 1990 to 1992 and the *final years* by the years from 1997 to 1999. The 1991 to 1995 years as previously noted were not reasonable years for analysis. The years 1996 and 1997 are assumed to represent the outer boundaries of the middle years and are not included in the comparative analyses of the ninth grade to follow. For the fourth grade proficiency tests the beginning years are 1995 and 1996 and the final years are assumed to be 1998 and 1999. For the fourth grade proficiency tests 1997 is considered to be a middle year and is not included in the fourth grade analyses. The sixth grade proficiency tests are the most recently

implemented proficiency tests. The beginning years for the sixth grade tests are 1996 and 1997 and the final years are 1998 and 1999. For the comparisons to follow all three grade levels share in common the same ending years. The mean differences for each poverty group comparison on each proficiency subtest, at each grade level, for each year from 1990 to 1999 were determined and converted to ESCs. These ESCs are reported in Table 2. In the comparative analyses based on Table 2 the years from 1993 to 1994 are not included for any grade level.

Table 2.

ESC by proficiency subtests within test year for Fourth, Sixth, and Ninth grades, noting the final and beginning years pooled by grade level to test achievement gap reduction hypothesis

test year	beginning and ending years used in pooling			subject	L ¹ - H ³ differences grades			M ² - H differences grades			L-M differences grades		
	4	6	9		4	6	9	4	6	9	4	6	9
99	final	final	final	Write	.63	.53	.59	.46	.42	.48	.25	.17*	.22*
99	final	final	final	Read	.59	.67	.53	.48	.51	.38	.18	.30	.19
99	final	final	final	Math	.57	.65	.62	.45	.48	.45	.21	.29	.29
99	final	final	final	Citizen.	.67	.70	.57	.57	.56	.43	.21	.29	.20*
99	final	final	final	Science	.57	.72	.71	.54	.57	.55	.11*	.31	.34
98	final	final	final	Write	.55	.56	.51	.45	.46	.42	.15*	.20	.12*
98	final	final	final	Read	.66	.70	.66	.56	.61	.53	.23	.22	.26
98	final	final	final	Math	.64	.63	.71	.49	.52	.59	.25	.23	.26
98	final	final	final	Citizen.	.67	.72	.62	.58	.61	.51	.25	.26	.23
98	final	final	final	Science	.64	.68	.71	.56	.59	.58	.20	.21	.26
97		begin	final	Write	.58	.47	.44	.47	.42	.36	.23	.10*	.08*
97		begin	final	Read	.65	.59	.68	.50	.51	.56	.28	.14*	.24
97		begin	final	Math	.61	.62	.64	.49	.50	.49	.24	.22	.26
97		begin	final	Citizen.	.68	.68	.60	.55	.56	.44	.27	.21	.22
97		begin	final	Science	.68	.63	.70	.56	.50	.55	.24	.22	.27
96	begin	begin		Write	.68	.60	.43	.53	.48	.35	.28	.23	.10*
96	begin	begin		Read	.71	.66	.59	.62	.60	.42	.24	.16*	.22
96	begin	begin		Math	.69	.68	.67	.61	.56	.54	.21	.27	.22
96	begin	begin		Citizen.	.68	.70	.62	.57	.58	.42	.23	.23	.27
96	begin	begin		Science	.71	.69	.03*	.61	.58	.07*	.25	.24	-.02*
95	begin			Write	.61	.	.29	.51	.	.19*	.19	.	.13*
95	begin			Read	.69	.	.38	.59	.	.28	.20	.	.16
95	begin			Math	.70	.	.58	.58	.	.43	.23	.	.23
95	begin			Citizen.	.71	.	.37	.63	.	.34	.23	.	.06*
94				Write	.	.	.48	.	.	.36	.	.	.14*
94				Read	.	.	.52	.	.	.39	.	.	.18
94				Math	.	.	.66	.	.	.50	.	.	.27
94				Citizen.	.	.	.54	.	.	.43	.	.	.14
93				Write	.	.	.43	.	.	.29	.	.	.18
93				Read	.	.	.67	.	.	.52	.	.	.29
93				Math	.	.	.72	.	.	.53	.	.	.40
93				Citizen.	.	.	.65	.	.	.50	.	.	.23
92			begin	Write	.	.	.39	.	.	.29	.	.	.13*
92			begin	Read	.	.	.69	.	.	.51	.	.	.35
92			begin	Math	.	.	.76	.	.	.54	.	.	.44
92			begin	Citizen.	.	.	.76	.	.	.47	.	.	.55
91			begin	Write	.	.	.47	.	.	.39	.	.	.14*
91			begin	Read	.	.	.66	.	.	.48	.	.	.36
91			begin	Math	.	.	.73	.	.	.58	.	.	.38
91			begin	Citizen.	.	.	.60	.	.	.47	.	.	.24
90			begin	Write	.	.	.48	.	.	.34	.	.	.20
90			begin	Read	.	.	.61	.	.	.46	.	.	.28
90			begin	Math	.	.	.71	.	.	.55	.	.	.36
90			begin	Citizen.	.	.	.67	.	.	.52	.	.	.26

¹low poverty group (N= 60); ²middle poverty group (N= 484); ³high poverty group (N= 60)

* The mean comparison associated with this ESC is not significant (P>.05)⁶.

The ESCs in all tables to follow are determined from the ESCs in Table 2. There are a number of trends in Table 2. The ESCs associated with the proficiency subtests differences between low and high poverty groups tend to be the largest in the table. It is also apparent that the ESCs associated with all high poverty group comparisons are quite large. The tables to follow systematically summarize trends that are associated with Table 2.

Table 3

Summary of poverty group average⁷ effect size correlation across subtest within years by grade level, a composite ESC

differences	Grade 4*		Grade 6*		Grade 9*	
	begin 95-96	end 98-99	begin 96-97	end 98-99	begin 90-92	end 97-99
composite ⁸	.52	.46	.47	.50	.49	.46

Composite Effect Size Correlations (Achievement Gaps) By Grade And Testing Period.

Table 3 represents general grade level ESC summaries, *composites*, for the beginning and ending years of the three grade levels as noted in Table 2. Each value entered in Table 3 is an ESC, for example the first left entry of .52, represents the extent to which there are overall poverty group achievement gaps associated with the fourth grade proficiency tests in the beginning test years, 1995 to 1996. As expected, based on the discussion of Table 2, the ESCs are extremely large for the beginning and ending years, greater than .35, suggesting substantial achievement gaps associated with the poverty group differences at each of the three grade levels. Most important in Table 3 are the beginning and ending ESCs for the fourth and ninth grade proficiency tests. For these two grade levels the ending ESCs are smaller than the beginning ESCs, demonstrating a reduction in achievement gap from the beginning to ending years for both grade levels.

Grade four had the greatest achievement gap for the beginning years. However, grades four and nine have identical ESCs for the ending years. Grade six has the smallest beginning years' ESC and, unlike the fourth and ninth grades, the largest ending years' ESC, suggesting a

growth in the sixth grade achievement gap from the beginning to ending years.

Composite Effect Size Correlations (Achievement Gaps) By Poverty Group Differences Within Grade Level And Testing Period. Table 4 represents general summaries, composites, of the ESCs in Table 2 by poverty group contrast. Each value entered in Table 4 is an ESC, for example the first upper-left entry of .69, represents the extent to which there is an achievement gap between the low and high poverty groups on all fourth grade subtests administered in the years of 1995 and 1996.

Table 4.
Summary of poverty group contrasts average⁹ ESCs across subtest within year groupings and grades, composite poverty group contrast ESCs.

contrasts	Grade 4*		Grade 6*		Grade 9*	
	begin 95-96	end 98-99	begin 96-97	end 98-99	begin 90-92	end 97-99
L-H	.69 ¹⁰	.63	.64	.67	.64	.61
M-H	.59	.52	.53	.54	.47	.47
L-M	.23	.22	.20	.25	.31	.22

*Science tests are not included¹¹

In Table 4 the largest set of achievement gaps, regardless of time period, are associated with the ESCs between the low and high poverty groups. The second largest set of achievement gaps, regardless of time periods, are associated with the ESCs between the middle and high poverty groups and are substantially less than the previously noted largest set. All ESCs associated with the middle-low poverty group contrasts are small, by both the pre-established size criteria and relative to the ESCs associated with high poverty group contrasts.

A second trend in Table 4 is associated with the beginning and ending fourth and ninth grade ESCs for all poverty group contrasts. All ending ESCs for both grades are smaller than their associated beginning ESCs, suggesting, in general, that the achievement gap discrepancies for all three poverty group comparisons have been reduced over the testing period for these two grade levels. However, the composite ESCs for the

beginning and ending sixth grade testing periods suggest that the achievement gaps for all three poverty group comparisons are generally larger in the ending years than in the beginning years.

A third trend associated with the contrasts within grade levels shows that the achievement gaps between low and high poverty groups have been reduced in both the fourth and ninth grade. For the fourth grade both achievement gaps associated with the high poverty group, L-H and M-H, were reduced, to a greater degree than for the middle-low group. But the ninth grade achievement gap associated with the middle - high group contrast has not changed from the beginning to ending test years, yet the achievement gap associated with the middle-low group contrast has gotten smaller. This suggests that generally, with the exception of the sixth grades, the achievement gaps associated with the difference between low and high poverty groups are being closed more rapidly than the achievement gaps associated with the other poverty group comparisons.

A fourth, somewhat curious trend, is associated with the sixth grade testing. The ESCs associated with each set of sixth grade poverty group contrasts, all suggest that the achievement gaps are getting larger on the sixth grade proficiency tests. Reflecting the same trend noted in Table 3.

Composite Effect Size Correlations (Achievement Gaps) By Proficiency Subtest Within Grade And Testing Period. Table 5 represents general summaries, composites, of the ESCs in Table 2 by subtest. The first upper left entry of .48 represents the extent to which there is an achievement gap between the poverty groups on the fourth grade Writing subtests administered in the years of 1995 and 1996.

Table 5.
Summary of poverty group ESCs within subtests, year groupings, and grade level, composite subtest ESCs.

	Grade 4	Grade 6	Grade 9
--	---------	---------	---------

subtest	begin 95-96	end 98-99	begin 96-97	end 98-99	begin 90-92	end 97-99
Writing	.48	.43	.40	.40	.32	.37
Reading	.54	.47	.47	.53	.50	.46
Math	.53	.45	.49	.49	.58	.50
Citizen.	.53	.51	.52	.55	.52	.35
Science	.55*	.47	.48	.53		.54

*Science was not a subtest until 1996

As suggested by the previous tables there is an achievement gap reduction for every fourth grade subtest from the beginning to the ending years. Writing with the smallest fourth grade ESCs is associated with the smallest achievement gaps, but still sizable gaps.

For the ninth grade Writing subtest the ending years' ESC is greater than the beginning years', suggesting an achievement gap growth. For the remaining ninth grade proficiency subtests, with the exclusion of Science which is not included in the beginning years, the achievement gap is reduced between the beginning to the ending years.

For the sixth grade, already established as a problematic grade level with regard to achievement gap reduction, there is no evidence of an achievement gap reduction for any subtest. There has been an achievement gap increase for Reading, Citizenship, and Science, but for Writing and Math there has been no change in the achievement gap from the beginning to the ending years.

Composite Effect Size Correlations (Achievement Gaps) By Poverty Group Contrast Within Grade Level Proficiency Subtest And Testing Period. Table 6 represents general summaries, composites, of the ESCs in Table 2 by poverty group comparison within subtest. Like the previous tables each value entered in Table 6 is an ESC. The first upper left entry of .65 represents the extent to which there is an achievement gap between the low and high poverty groups on the fourth grade Writing subtests administered in the years of 1995 and 1996.

As expected, based on the discussion of prior tables, the ESCs in Table 6 are extremely large, suggesting substantial achievement gaps associated with the poverty groups contrasts within subtest.

As suggested by Table 3 and verified also in Table 4, there is a trend that the largest achievement gaps are associated with the contrasts between the high and low poverty groups and the smallest achievement gaps are associated with the contrasts between the low and middle poverty groups, regardless of subtest and testing year.

With the exception of the low-middle poverty group contrasts for Math and Citizenship, the beginning years' achievement gap associated with each poverty group contrast for each fourth grade proficiency subtest is reduced by the ending years. It is reasonable to conclude that every fourth grade proficiency subtest achievement gap associated with the high poverty group has been reduced by the ending years.

Table 6.
Composite Subtest ESC: Summary of poverty group ESCs within subtests and years with global ESC for subtests.

Group Differences	subtest	Grade 4		Grade 6		Grade 9	
		begin 95-96	end 98-99	begin 96-97	end 98-99	begin 90-92	end 97-99
L-H	Writing	.65	.60	.54	.55	.45	.52
	Reading	.70	.62	.63	.69	.66	.63
	Math	.69	.61	.66	.64	.74	.66
	Citizen.	.69	.67	.68	.71	.68	.60
	Science	.71*	.61	.66	.70	*	.70
M-H	Writing	.52	.46	.45	.44	.34	.42
	Reading	.61	.52	.56	.57	.49	.49
	Math	.53	.47	.53	.51	.54	.51
	Citizen.	.60	.58	.57	.59	.49	.46
	Science	.61*	.55	.53	.58	*	.56
L-M	Writing	.24	.20	.17	.18	.16	.14
	Reading	.22	.20	.14	.26	.33	.23
	Math	.22	.23	.24	.26	.39	.27
	Citizen.	.23	.23	.22	.27	.36	.21
	Science	.25*	.19	.23	.26	*	.29

*Science was not a subtest until 1996

For the ninth grade proficiency tests with the exception of the middle-high contrast in Reading, which is unchanged, all poverty group contrasts associated with Reading, Math, and Citizenship show a reduction in ESCs from the beginning to ending years, a reduction in the achievement gaps for all three poverty group comparisons.

For the sixth grade, as suggested by the other analyses, all poverty group achievement gaps have increased from the beginning to the ending test periods in Reading, Citizenship, and Science. In Math the ending years' middle -high poverty group ESC is smaller than the beginning years' ESC, indicative of a reduction of the ninth grade Math achievement gap between the middle and high poverty groups. The other two contrasts in Math, middle - low and high-low, show an increase in ESCs, indicating an increase in their associated achievement gaps from the beginning to ending years. The Writing subtest is relatively unchanged across the three poverty group contrasts from beginning to ending years. There is not much positive that can be said about improvement in the sixth grade achievement gaps from the beginning to ending years of testing.

Summary of Analyses

- (1) With the exception of the Writing subtest the effect size correlations for all remaining subtests of the Ohio Proficiency Tests, at all three grade levels, define very large poverty related achievement gaps. Those achievement gaps associated with the high poverty group are especially large.
- (2) There is a clear and definite reduction of the poverty related achievement gaps associated with the Fourth Grade Proficiency Tests and evidence of a reduction of the poverty related achievement gap associated with the Ninth Grade Proficiency Tests. When comparing the yearly composite effect size correlations defined by the final years' achievement gap for the fourth and ninth grades they are identical.
- (3) As indicated by the yearly composite effect size correlations for sixth grade the poverty related achievement gap associated with the Sixth Grade Ohio Proficiency Tests became larger in the final years, somewhat unexpected given the reductions for the fourth and ninth grades.

Conversations with the Ohio Department Of Education regarding this finding were quite informative¹². There is some concern that by attempting remediation in fifth grade, the problems identified by poor performance on the Fourth Grade Proficiency Tests, the schools are effectively “short-changing” the fifth grade curriculum that is a pre-requisite for the sixth grade curriculum, with the end result being poor performance on the Sixth Grade Proficiency Tests.

(4) The effect size correlation changes, reduction of achievement gaps, appear to be small, but they are associated with all public schools in Ohio and even a small change represents a substantial change when considering the total number of students benefiting from the change.

(5) With the exception of the Writing subtest the effect size correlations for all remaining subtests of the Ohio Proficiency Tests, at the fourth and ninth grade level, show a reduction from the beginning to the ending years.

(5) In summary, the promise of minimum competency testing has been confirmed for two of the three grade levels associated with the Ohio Proficiency Tests.

Discussion

The discussion of the analyses presented is challenging. The concept of a poverty related achievement gap is easy to understand. As indicated by the tables and discussion in the preceding section the poverty related achievement gap associated with the fourth and ninth grade Ohio proficiency tests has been reduced, demonstrating that it is possible to reduce the gap. As suggested by the debates in the early 80's minimum competency testing can lead to reductions in achievement gaps defined by achievement performance differences between certain demographic groupings. It would be nice to have a sense of how this reduction was actually accomplished. Unfortunately, how the high poverty school districts actually accomplished the feat of raising the passing rates of their districts more rapidly than the low poverty districts must remain a mystery at this point in time..

The challenge of this discussion is primarily one of speculating on “the effects of proficiency testing” on differential poverty group proficiency performance. After considerable interaction with teachers and administrators, Reading, and “hand-wringing”. I have concluded that there are at least four distinct discussions that might be offered: a pessimistic commentary from those in the field (teachers and administrators), an optimistic commentary from those in the field, a philosophical commentary from the perspective of the “testing experts” who are proponents, and a philosophical commentary from the perspective of “testing experts” who are opponents of minimum competency testing.

Pessimistic Commentary By Teachers And Administrators

There are teachers and administrators that agree that a poverty related achievement gap has been reduced, “but if you think this is good news go out to a high poverty school, come to my school, and see what it’s really like. This change is simply not large enough and not fast enough”

A recent conversation with a superintendent was especially pessimistic. He noted that while his school district had received almost two million dollars in grants for in-service, workshops, after school tutoring and other activities intended to enhance performance on state proficiency tests, the percentiles associated with the school district’s percent pass rate were becoming lower rather than higher. He was genuinely concerned about staff morale, administrator despondency, and the learning of the children in his school district. He was the district’s superintendent before proficiency testing and noted that he and his staff were always proud of their district’s academic achievements, but now all they have are academic disappointments.

Teacher criticism tends to focus on curriculum and those who do not benefit from the curriculum change, “True we’ve reduced an achievement gap between high poverty schools and the rest of the schools in Ohio, but at the expense of our excellent students who are no longer being challenged. Many of our resources are being re-channeled into

programs to increase our pass rates on proficiencies. We are now spending far more money on our non-college-bound students than on our college-bound students.”

“Our curriculum has been transformed into a minimum competency curriculum. We no longer have time to address the individual needs in our classrooms. We must continually drill on concrete proficiency competencies – at the expense of the “hands-on” things that our children enjoyed so much.”

In summary, the reactions of the pessimists are generally statements of frustration, noting what has been lost, and what things have not been accomplished because of proficiency testing. While there is acceptance that there may be benefits to be accrued from proficiency testing, the pessimists perceive their schools as being disadvantaged by proficiency testing.

Optimistic Commentary By Teachers And Administrators

Our conversations with practitioners indicate that there are far fewer optimists than pessimists with regard to proficiency testing. The optimists tend to be associated with schools that are doing well or at least better than expected. They too realize that there are schools that have been successful with regard to improving school district performance on proficiency tests and they want to learn more about these successful procedures. As one Appalachian superintendent (of a successful high poverty school) noted, “We need to make a real effort to identify the high poverty school districts that are having success. We need to study these schools to better understand what is taking place so that we can perpetuate it in other similar schools.”

A senior teacher noted, “Clearly, we have made competency gains with our teaching in the lower grades (*fourth grade*). We will continue this progress into the sixth grade and should soon see the results of these efforts.”

One assistant superintendent (from a low poverty school district) noted that he liked the idea of the proficiency tests, “opening the door to

the school so that people could see that there are certain students not being served by the school.” It is especially gratifying for him to see a reduction in the achievement gap. He views this as evidence that “the problem of differential school district achievement associated with poverty differences is a problem that can be addressed.”

A theme that echos from teachers associated with schools that are doing “well” on proficiency tests is represented by the following, “Now that we have had the proficiency tests for a few years we have an effective curriculum in place. We know what we are doing and I like that. We’re doing well and we feel good about it.”

In summary, the optimist are delighted that the high poverty schools, and many of the other schools with marginalized students are addressing the needs of the lower achieving students. They see the proficiency tests as providing a context within which they can construct a meaningful curriculum that will evolve with the tests. The desire to study successful schools suggests that there may be important curricular innovation taking place and the new innovative curriculum may prove to be beneficial for other schools once they learn about it.

Proponent Commentary - Test Experts

Proponents of competency testing such as Popham (1981, 1987) would argue that reporting proficiency test results by school district has sensitized the public as well as the school officials to the fact that there are certain groups of children not being “served” by public schools. The public distribution of school district proficiency test performance by state report cards in Ohio as well as by local newspapers brings public pressure on the school district administrators and teachers by making them publicly accountable for educating the children of the school district. The school districts have responded to this pressure and, as evidenced by the tables in this study, have “learned how” to better accommodate the pedagogical needs of these students.

Indeed the achievement gap reduction hypothesis has been confirmed. With regard to schools that are doing poorly on the Ohio Proficiency Tests, “Would you kill the messenger that brings bad news?”

Opponent Commentary – Test Experts

Similar to the community of practitioners the community of testing experts who are opponents of mandated competency testing outnumber the proponents. There are a substantial number of publications in the 90’s representing opponent positions to minimum competency testing. Even proponents for minimum competency express some concerns for the effects of minimum competency testing on curriculum (see Popham 1998). In the brief commentary to follow, Ohio proficiency testing is discussed within Foucault’s (1977, pp. 187-194) framework of examinations. His framework is brief, but encompassing with regard to the contemporary concerns and criticisms. Although no single opponent of MCT describes testing quite like Foucault and no single opponent of MCT in a single article raises all of the issues briefly noted in the passage to follow [Madaus (1993) come close], the passage organizes and summarizes most of the criticism of MCT that might be focused on the Ohio Proficiency Tests by testing experts.

Foucault (1977, pp187- 194) discusses testing as disciplinary power. The power of minimum competency testing is that it exposes those who are associated with performance on the test, “a principle of compulsory visibility.” The students, the teachers, the building administrators and the district administrators are hierarchically exposed by their proficiency performance, - a collective test performance which is represented by a building and, ultimately, school district performance. The qualitatively complex and frequently abstract aspects of learning that are the essential pedagogical embodiment of a school district, and hierarchically those within the school district, are represented by several numerical scores derived from district performance on the Ohio Proficiency Tests (Foucault refers to this as objectification). These are the scores that are publicized

by the newspapers and the school district report cards, a public exposure of school district performance on proficiency tests.

The public holds and the newspapers perpetuate a perspective that these few scores are predictors of the future educational and life choices of their children. In turn, the students, teachers, and administrators perceive the proficiency tests as being important. As a consequence of this perceived importance the Ohio Department of Education , ODE, amasses a greater power over the schools. This power is granted to ODE by the teachers who "teach to the proficiency tests" and the administrators who tolerate such teaching. The net result is that the traditional curriculum is abandoned and ODE establishes the curriculum by proficiency subtest content – and now the ODE owns the curriculum through the proficiency tests and can exercise its power over the school district (administrators, teachers and, of course, the curriculum) with the proficiency tests.

Following Foucault (p189) to his conclusion there is an inversion of power. By making the school districts "objectively visible", through a public display of the proficiency tests % pass as a school's scores, the school districts lose their control of the curriculum, lose their pedagogical autonomy, and no longer provide an education leading to higher order thinking.

In conclusion, with regard to the reduction of the achievement gap, "You can't discount the probability that the reduction of the achievement gap has occurred because the teachers have taught to the test and are not yet providing a quality education in the high poverty schools – perhaps the quality of the education in these high poverty schools is even worse now than it was before proficiency testing."

End Note

Notes

¹ Unlike the fourth and sixth grade proficiency tests, the ninth grade proficiency test must be passed by students by the completion of twelfth grade in order to receive a high school diploma. Thus, there are students who take the ninth grade proficiency more than a single time. These students are not included in the aggregate data used in this study.

² This index was derived from factor scores based on a confirmatory factor analysis. There is substantial evidence attesting to the validity of this index (see Hofmann note 1 and Hofmann 1999).

³ Initial effect size in this study follows Cohen's (1988) pooled variance approach for his *d*. Specifically: $d = (\text{Mean}_1 - \text{Mean}_2) / ((\text{Variance}_1 + \text{Variance}_2) / 2)^{1/2}$ where in this study group 1 will refer to the comparison group with the lowest poverty and group 2 will be the comparison group with the highest poverty. So when making comparisons the low poverty group will always be represented by group 1 and the high poverty group will be represented by group 2. In this study such comparisons will almost always result in a positive difference. Rosenthal's (1991) effect size correlation, represents the correlation between the two poverty groups, a dichotomous variable, since the effect size is computed from the comparison of the two poverty group means, and a continuous measure represented by the mean proficiency score associated with each school district in the two poverty groups associated with the mean difference. This correlation is computed as, $r = d / [d^2 + 4]^{1/2}$ (Rosenthal 1991).

⁴ When considering the statistical manipulation of effect size correlations, computing average Zs, each correlation should be converted to a normal curve Z score. This is accomplished using the Fisher (1970, p 212) *r* to Z transformation where:

$$Z = 1/2[\log_e(1+r) - \log_e(1-r)]$$

and when the average z has been determined it is transformed back to an *r*:

$$r = (e^{2z} - 1) / (e^{2z} + 1)$$

⁵ Because effect size correlations are computed directly from normal curve z-scores we can use the z-scores associated with the extreme 5% and 1% of the distribution to determine the magnitude of very large effect size correlations – those associated with the extreme 5% and 1% of a normal distribution. An effect size correlation of .70 would be associated with the extreme 5% and an effect size correlation of .79 would be associated with the extreme 1%.

⁶ For all mean comparisons the mean associated with the poverty group signified by the first letter (L or M) was significantly ($p < .01$) greater than the mean associated with the poverty group signified by the second letter **expe**ct when the ESC is followed by *. ANOVA followed by a Fisher PLSD *post hoc* was carried out for each subtest within each year. With two exceptions all ANOVA F-ratios were significant ($P < .001$) – the two exceptions were ninth grade Science in 1996 which defined a non-significant F-ratio ($P > .05$) and 1995 ninth grade Writing which defined an F-ratio that was significant at a value greater than .001, ($p < .003$).

⁷ All averages, referred to as composites, in all tables are computed using the normal curve Z representation of an ESC with the resulting average transformed from Z to an ESC and then placed in the table. As a consequence of this transformation process a composite will not be the same as the average of the column or row ESCs.

⁸ The composite value of .52 is the average ESC (using Zs to determine the average, for all contrasts on all fourth grade proficiency subtests for the 1995 and 1996 years.

⁹ All averages in all tables are computed using the normal curve Z representation of an ESC with the resulting average transformed from Z to an ESC and then placed in the table. As a consequence of this transformation process the composite will not be the same as the average of the column ESC.

¹⁰ The ESC value of .69 is the averaged ESC for all low –high poverty group contrasts on all fourth grade proficiency subtests for the 1995 and 1996 years, the beginning years of the Ohio Fourth Grade Proficiency Test.

¹¹ Because the Science subtests were not available for the beginning years for sixth and ninth grade they were not included in the analyses providing the ESCs in Table 3.

¹² I am indebted to individuals from the Ohio Department of Education who spent time helping me to understand the sixth grade anomaly and other aspects of Ohio proficiency testing that were problematic for me. This specific topic as well as other topics were discussed during an informal meeting (April 3, 2000) with individuals associated with Ohio Department of Education's Ohio Proficiency Tests, R. Trent (Director of Assessment & Evaluation), M. Cohen (Director of Policy, Research, and Analysis), and J. Crandell (Assistant Director of Assessment and Evaluation) – they are indeed, patient optimists and gentle proponents, of course.

References

- Airasian, P. W. (1987). State Mandated Testing and Educational Reform: Context and Consequences. *American Journal of Education*, 95, 393-412.
- Airasian, P. W. (1988). Symbolic Validation: The Case of State-Mandated, High-Stakes Testing. *Educational Evaluation and Policy Analysis*, 10, 301-313.
- Berliner, D., C. & Biddle, B.J. (1998). The lamentable alliance between the media and school critics. *School Administrator*, 55(8), pp. 12-16, 18
- Cohen, J. (1988). *Statistical power analysis for the behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A. M., Weinfeld, F.D. & York, R.L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office
- Fisher, R.A. (1970). *Statistical methods for research workers*. New York. Hafner
- Foucault, M. (1977). *Discipline and punish, the birth of the prison*. New York: Viking
- Hofmann, R. (1999). *Indices of community SES and poverty: Predictors of school district achievement and definers of the difficulty of the educational task (the dot index)*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada - April 1999
- Hofmann, R. (1997). *School district effectiveness, difficulty of the educational task, and Ohio ninth grade proficiency tests: a seven year study*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 1997.
- Linn, R.L. (2000) Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

- Meehl, P. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and qualify accuracy of risky numerical predictions. In L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393 - 425). Mahwah, NJ: Erlbaum.
- Payne, K.J. & Biddle, B.J. (1999) Poor school funding, child poverty, and Mathematics achievement. *Educational Researcher*, 28(6), 4-13.
- McLoyd, V. C. (1997). The impact of poverty and low socioeconomic status on the socio-emotional functioning of African-American children and adolescents: Mediating effects. *Social and emotional adjustment and family relations in ethnic minority families*, 7-34.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*. 53(2): 185-204.
- Popham, W. J. (1981). The Case for Minimum Competency Testing. *Phi Delta Kappan* 63(2): 89-91.
- Popham, W. J. (1987). The Merits of Measurement-Driven Instruction. *Phi Delta Kappan*. 68(9): 679-82.
- Popham, W. J. (1998). Farewell, Curriculum: Confessions of an Assessment Convert. *Phi Delta Kappan*. 79(5): 380-84.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Rozeboom, W. (1997). Good Science is abductive, Not Hypothetico-Deductive. In L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.) *What if there were no significance tests?* (pp. 335 - 392). Mahwah, NJ: Erlbaum.
- Schmidt, F. L. and Hunter, J. E.(1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37 - 65). Mahwah, NJ: Erlbaum.
- Thurston, P. & House, E.R. (1981). The NIE Adversary Hearing On Minimum Competency Testing. *Phi Delta Kappan*. 63 (2) 87-89.