Educational Considerations

Volume 39
Number 1 *Class Size and Student Achievement*

Article 3

9-1-2011

Another Look at the Glass and Smith Study on Class Size

James L. Phelps

Recommended Citation

Phelps, James L. (2011) "Another Look at the Glass and Smith Study on Class Size," *Educational Considerations*: Vol. 39: No. 1. https://doi.org/10.4148/0146-9282.1100

Another Look at the Glass and Smith Study on Class Size

James L. Phelps

One of the most influential studies affecting educational policy is Glass and Smith's 1978 study, *Meta-Analysis of Research on the Relationship of Class-Size and Achievement*.[1] Since its publication, educational policymakers have referenced it frequently as the justification for reducing class size. While teachers and the public had long believed lowering class size would be advantageous, Glass and Smith gave the idea legitimacy. This article is a review and reanalysis of the Glass and Smith study. While this review maybe considered much too late, it does serve the purpose of re-evaluating a frequently cited study to either support or challenge various aspects of the original findings. To that end, the article is divided into six major parts. It begins with an overview of the Glass and Smith study for those who may not be familiar with the specifics. This is followed by a description of their findings and comments upon these by the author. The fifth section presents a reanalysis of their data. The article closes with observations and conclusions.

Overview

To capture the character of the original study, the summary from Glass and Smith is presented here in its entirety (pp. iv-vi):

> Research on the relationship between class-size and academic achievement is old, huge and widely believed to be inconclusive. Previous reviews of the evidence have been overly selective and insufficiently quantitative. Timid qualifications were offered where bold generalizations were possible. In the summer of 1978, the *New York Times* gave front-page coverage to a study published by Educational Research Services, Inc. (Porwell, 1978). This organization is funded jointly by the American Association of School Administrators, the Council of Chief State School Officers, and several other professional administration groups. The "Porwell Report" staggered visibly under the weight of the research data and eventually arrived at the following conclusion sad for teachers to behold:

James L. Phelps holds a Ph.D. from the University of Michigan in Educational Administration. He served as Special Assistant to Governor William Milliken of Michigan and Deputy Superintendent in the Michigan Department of Education. Active in the American Education Finance Association, he served on the Board of Directors and as President. Since retirement, he spends a great deal of time devoted to music, composing and arranging, playing string bass in orchestras and chamber groups, as well as singing in two choirs. He resides with his wife, Julie, in East Lansing, Michigan.

> (Quotation, continued)
>
> Research findings on class size to this point document repeatedly that the relationship between pupil achievement and class size is highly complex.
>
> There is general consensus that the research findings on the effects of class size on pupil achievement across all grades are contradictory and inconclusive.
>
> Existing research findings do not support the contention that smaller classes will of themselves result in greater academic achievement gains for pupils (Porwell 1978, 68-69).
>
> The research reported herein contradicts the conclusions of the Porwell Report. Indeed, it establishes clearly that reduced class-size can be expected to produce increased academic achievement. In pursuing this conclusion, we discovered many of the reasons why previous research reviewers lost their way in the forest of data and failed to find a defensible generalization.
>
> We collected nearly 80 studies on the relationship between class-size and achievement. These studies yielded over 700 comparisons of the achievement of smaller and larger classes; these comparisons rest on data accumulated from nearly 900,000 pupils of all ages and aptitudes studying in all manner of school subject. Using complex methods of regression analysis, the 700 comparisons were integrated into a single curve showing the relationship between class-size and achievement in general. This curve revealed a definite inverse relationship between class-size and pupil learning. Similar curves were derived for a variety of circumstances hypothesized to alter the relationship between achievement and class-size. Virtually none of the special circumstances altered the basic relationship; not grade level, nor subject taught, nor ability of pupils. Only one factor substantially affected the curve, viz., whether the original study controlled adequately (in the experimental sense) for initial differences among pupils and teachers in smaller and larger classes. The nearly 100 comparisons of achievement from the well-controlled studies thus form the basis of our conclusion about how class-size is related to academic achievement. This curve appears in the Figure below. As class-size increases, achievement decreases. A pupil, who would score at about the 83rd percentile on a national test when taught individually, would score at about the 50th percentile when taught in a class of 40 pupils. The difference in being taught in a class of 20 versus a class of 40 is an advantage of 6 percentile ranks. The major benefits from reduced class-size are obtained as size is reduced below 20 pupils.

As one looks at the representation of the relationship between achievement and class size, several immediate questions arise:

(1) Why are the relationships all above the 50th percentile?

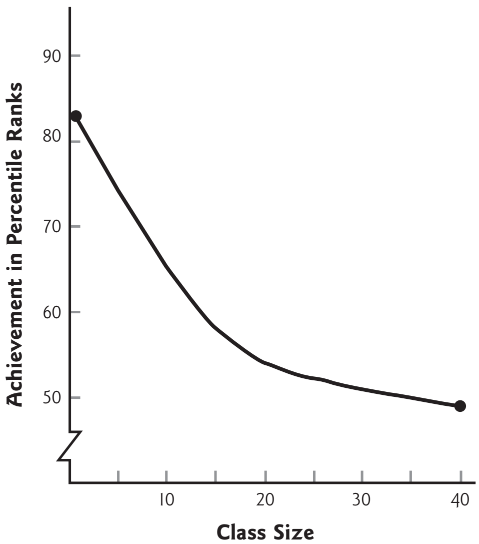
(2) Why is the relationship curved?

(3) Why are the relationships not reported for class sizes larger than 40?

(4) How many teachers are necessary to bring the class size down from 40 to 20, from 40 to 10, and from 40 to 1?

Published by New Prairie Press, 2017

Figure 1
Curve Derived by Glass and Smith from 100 Comparisons from Well Controlled Studies



Source: Gene V. Glass and Mary Lee Smith, *Meta-Analysis of Research on the Relationship of Class-Size and Achievement* (San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978), vi, Figure 1.

Research Method

Glass and Smith described their research method, meta-analysis, in detail.[2] They took comparisons between achievement and class size from many studies, formed a new data set, and then conducted a regression analysis using this data set. The following subsections summarize each of the topics addressed.

Defining the class size field (p. 9). Glass and Smith selected the number of pupils within a class with one teacher as the measure rather than a measure of "staff adequacy," the number of teachers per 100 pupils. While there was a mathematical transformation equating the two notions, there was a substantial difference in their policy implications, to be discussed later.

Coding characteristics of studies (pp. 10-13). Glass and Smith collected data for the following fields, although data from some studies were not available and not all fields were completely filled: ID number of study; year of study (1900-1979); source of data (whether from journal, book, thesis, or unpublished source); subject taught (reading, mathematics, language, psychology, natural/physical science, social science and history, and "all others"); duration of instruction, in hours and in weeks; number of pupils, instructional groups, and teachers; pupil/instructor ratios for small and large classes; assignment of pupils and teachers; subject of achievement measure; and achievement measure (the difference in achievement between the small and large classes). Other data items were collected but are not included in this listing because they were not incorporated into their analyses.

Quantifying outcomes (pp. 13-14). For each of the comparisons from each of the studies a single statistic was required. Glass and Smith stated:

> No matter how many class-sizes are compared, the data can be reduced to some number of paired comparisons, a smaller class against a larger class... The most obvious differences involve the actual sizes of the "smaller" and "larger" classes and the scaled properties of the achievement measure... The measurement scale properties can be handled by standardizing all mean differences in achievement by dividing by the within group standard deviation (a method that is complete and discards no information at all under the assumption of normal distributions).

The achievement measure was standardized across all studies through the use of standard or Z-scores. The achievement measure in Z-scores was notated by Glass and Smith (pp. 13-14) as:

$$\Delta_{(S-L)} = (\bar{X}_{(S)} - \bar{X}_{(L)} / \hat{\sigma}$$

where

S represents the small class;

L represents the large class;

$\bar{X}$ represents the achievement mean;

and $\hat{\sigma}$ represents the standard deviation.

Calculating the achievement measure $\Delta_{(S-L)}$. Because many of the studies from which the data were taken did not include basic descriptive statistics, alternative methods to calculate the achievement variable had to be developed. Glass and Smith described their methods on pages 14-15.

Describing the class size and achievement relationship. Glass and Smith considered several alternative statistical techniques to describe the aggregated findings. The selected alternative is quoted below (pp. 15-19):

> Finally, regression equations could be constructed in which $\Delta_{(S-L)}$ is partitioned into a weighted linear combination of S and L and function thereof and error... But the regression of $\Delta_{(S-L)}$ into only S and L requires three dimensions to be depicted. Anything more complex than a simple two-dimensional curve relating achievement to the size of class was considered undesirably complicated and beyond the easy reach of most audiences who hold a stake in the results.
>
> The desire to depict the aggregate relationship as a single line curve is confounded with the problem of essential inconsistencies in the design and results of the various studies. A single study of class-size and achievement may yield several values of $\Delta_{(S-L)}$... This set of Δ's from a single study will form a consistent set of values in that they can be joined to form a single connected graph depicting the curve of achievements as a function of class-size. However, various values of $\Delta_{(S-L)}$ arising from difference studies can show confusing inconsistencies. For example, suppose that Study #1 gave Δ (10-15), Δ (10-20), and Δ (15-20) and Study #2 gave Δ (15-30), Δ (15-40), and Δ (30-40). A few moments reflection will reveal that *there is no obvious or simple way to connect these values into a single connected curve* [emphasis added].[3]
>
> The eventual solution to these problems proceeded as follows: $\Delta_{(S-L)}$ was regressed onto a quadratic function of S and L by means of the least-squares criterion: then that set of values of $\hat{\Delta}$ that could be expressed as a single, connected curve was found.

https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

The regression model selected accounted for variations in $\Delta_{(S-L)}$ by means of S, S^2 and L. *Obviously, something more than a simple linear function of S and L was needed, otherwise a unit increase in class-size would have a constant effect regardless of the starting class-size S; and the S^2 term seemed as capable of filling the need as any other. The size differential between the larger and smaller class, L-S was used in place of L for convenience* [emphasis added]. Thus, the $\Delta_{(S-L)}$ values were used to fit the following model:[4]

$$\Delta_{(S-L)} = \beta 0 + \beta_1 S + \beta_2 S^2 + \beta_3 (L-S) + \varepsilon ... \quad (1)$$

The problem now is to find the set of $\hat{\Delta}$'s in this surface that can be depicted as a single curved-line relationship in a plane.

It is important at this point to determine the dimensions of the equation. Obviously, achievement is the first dimension. Class size (the S and S^2 terms forming a parabola) is the second because for any value of S a value for achievement can be calculated. The uncertainty pertains to a possible third dimension. L would be a third dimension if it were a data variable entered into the regression equation and a value for achievement could be calculated for each value of L. However, L was not a data variable entered into the regression; rather, (L-S) was the variable. This point is critical: (L-S) can produce a value for achievement if, and only if, L is fixed and S varies. Therefore, (L-S) is not an independent third dimension; instead, it is a line within the class size dimension.

Next, Glass and Smith described a "consistency property." The relevant section of their study (pp. 17-19) has been included here because of its importance to the commentary in the fourth section of this article:

The property that must hold for a set of $\hat{\Delta}$'s before they can be depicted as a connected graph in a plane is what might be called the consistency property [emphasis in the original]:

$$\Delta_{n1-n2} + \Delta_{n2-n3} = \Delta_{n1-n3}$$

for n1 < n2 < n3. If this property is not satisfied, then one is in the strange situation of claiming that the differential achievement between class size 10 and 20 is not the sum of the differential achievement from 10 to 15 and then from 15 to 20.

When the consistency property is imposed on the regression equation, it follows that:

$$\hat{\beta}_0 + \hat{\beta}_1 n_1 + \hat{\beta}_2 n_1^2 + \hat{\beta}_3 (n_2-n_1) + \hat{\beta}_0 + \hat{\beta}_1 n_2 + \hat{\beta}_2 n2^2 + \hat{\beta}_3 (n_3-n_2) =$$
$$\hat{\beta}_0 + \hat{\beta}_1 n_1 + \hat{\beta}_2 n_1^2 + \hat{\beta}_3 (n_3-n_1) \quad (3)$$

Simple algebraic reduction produces the following:

$$\hat{\beta}_0 + \hat{\beta}_1 n_2 + \hat{\beta}_2 n_2^2 = 0 \quad (4)$$

The two solutions to the quadratic equation...are points n_2 such that if $\hat{\Delta}_{(S-L)}$ is measured with n_2 as either the larger L, or smaller, S, class-size then the resulting set of $\hat{\Delta}$'s will lie on the four dimensional regression curve...but can be depicted as a single line curve in a plane. Since n_2 becomes the point around which values of n_1 and n_3 are selected, it will be called the pivot point [emphasis in the original]. That there are two solutions for n_2 is perplexing; fortunately in the analyses to be reported the two corresponding curves were virtually parallel in practice.

A single line curve in a plane can be constructed by solving for one or the other values of n_2 in (4) and constructing a set of Δ values. These values will give the standardized mean differences in achievement between n_2 and any other class-size. The curve that connects these Δs has no non-arbitrary starting point. One can assume for convenience sake that the achievement curve (z), instead of the differential achievement curve (Δ) is centered around an arbitrary class-size, e.g., something like the national average in the low 20s. Finally, for descriptive purposes, the metric of percentile ranks was chosen over the metric of z-scores; thus, the curve z was transformed into a curve of percentile ranks by assuming a normal distribution of achievement.[5]

Comment on Statistical Inference [Underline in original]

In the analyses that follow, ordinary matters of statistical inference have been ignored. The application of usual interval estimation procedures or statistical tests makes little sense for two reasons. The data base is laced with a complicated structure of interdependent observations; several comparisons arise from a single study when more than two class-sizes are compared, and there is no sensible way to reduce each study to one observation... Secondly, randomization is absent from the data set in any form that would make probabilistic models based on it applicable.

Findings

According to Glass and Smith (p. 20), "The report of findings falls into two broad categories: (1) description of the data base and (2) regression analyses relating to achievement and class-size." I begin here with a quotation from their description of the data base (p. 20):

In all, 77 different studies were read, coded, and analyzed. These studies yielded a total of 725 Δ's. The comparisons are based on data from a total of nearly 900,000 pupils spanning 70 years research in more than a dozen countries. (The entire set of data is reproduced in the appendix to this report.)

Table 1
Glass and Smith Regression Equation Results

Class Size	Delta	Interval	Difference
1	0.5859	1 to 65.81	0.00001
10	0.2895	1 to 10	0.2964
20	0.0723	10 to 20	0.2172
25.84	0.0000	20 to 25.84	0.0723
30	-0.0269	20 to 30	0.0269
33.41	-0.0338	30 to 33.41	0.0068
40	-0.0081	30 to 40	-0.0256
40.97	0.0000	40 to 40.97	-0.0081
50	0.1287	40 to 50	-0.1287
60	0.3835	50 to 60	-0.2548
65.81	0.5857	60 to 65.81	-0.2022
Sum			0.0003

Source: Glass and Smith (1978).

Several tables were presented in the study showing the frequency distributions of the data characteristics (Tables 1-5, pp. 20-26). These are not summarized here. However, in the data set, small class size ranged between 1 and 70. Large class-size ranged between 2 and 146. These values come into consideration when parameters are set in the regression equations.

In their regression analyses section (p. 29), Glass and Smith presented the statistical properties of the dependent variable $\Delta_{(S-L)}$. Most interesting, 40% of the values for $\Delta_{(S-L)}$ were negative, and 60% were positive. The large percentage of negative values for Δ raises an interesting situation. For any value of S, if the sum of the Δ's is positive, the slope of the line will be positive; however, if the sum of the Δ's is negative, the slope of the line will be negative. This circumstance raises the possibility that the curve representing the full range of class sizes will be comprised of both positive and negative slopes.

The result of the regression analysis for the entire data set was (p. 33):

$$\Delta_{(S-L)} = .57072 - .03860\ S + .00059\ S^2 + .00082\ (L-S)$$

At this point, Glass and Smith provided a table with a range of small and large class-sizes with the Δ as calculated from the regression results above (p. 34). The small class size (S) is only up to 30, and the large class size is (L) up to 40, even though these values are substantially higher in the data set. This table, in an expanded form, is provided below. (See Table 1.) In order to calculate the regression results, a value for the large class size must be set, in this case a class size of 65.81, for a reason to become clear later. Calculations have also been included to test the consistency property: If intervals A to B + B to C = A to C.

Glass and Smith concluded:

> These data show that the difference in achievement between class-size 1... and class-size of 40 is more than one-half standard deviation. The difference between class-size 20 and 40 is only about five hundredths standard deviation. Class-size differences at the low end of the scale have quite important effects on achievement; differences at the high end have little effect (p. 34).

It should be noted in Table 1 that the predicted achievement for a class size of 40 is marginally better than that for a class-size of 30; and achievement continues to increase to a class-size 65.81 where achievement is virtually the same as for a class-size of 1.

Most interestingly, when the consistency property is tested using the data from the table, the sum of the intervals of class size from 1 to 10 and 10 to 20 equals the interval from 1 to 20, and all other intervals as well.[6] As will be demonstrated later, the data from Table 1 can be graphed in two dimensions.

Three questions arise: (1) Why does the regression equation predict almost the same achievement level for a class-size of 1 and 65.81; (2) Why are there two predicted achievement values of 0; and (3) What is the consequence of setting the value of L?

Utilizing the consistency property, Glass and Smith (p. 34) observed: "The curved regression surface can be reduced to a single line curve in a plane by imposing the consistency condition and solving for the pivot points. The two pivot points are the solutions to .57072 - .03860 (P) + .00059 (P^2) = 0." They calculated the pivot points to be approximately 43 and 23. Because a parabola was selected as the curve for the regression analysis, it comes as

Table 2
Glass and Smith Results Including the Consistency Property Transformation

Small Class Size	*Large Class Size*	*Standardized Differential Achievement, Δ_{S-L}*
1	23	0.551
10	23	0.254
20	23	0.037
23	23	-0.005
23	30	0.001
23	40	0.009
23	50	0.017
23	60	0.025
23	65.81	0.030

Source: Glass and Smith (1978, 35).

no surprise that the curve on its downward path intersected the zero plane of the Z-axis, continued downward to a minimum point, about 33.4, and then moved upward, again intersecting the zero plane of the Z-axis as it continued upward.[7] The pivot points are the intersections of the parabola with the Z-axis. As part of the results of their study, Glass and Smith (p. 35) presented a table showing the results of the consistency property transformation, although no calculations were presented. This statement preceded and followed the table, which has been expanded here as Table 2 to show class-sizes larger than 40:

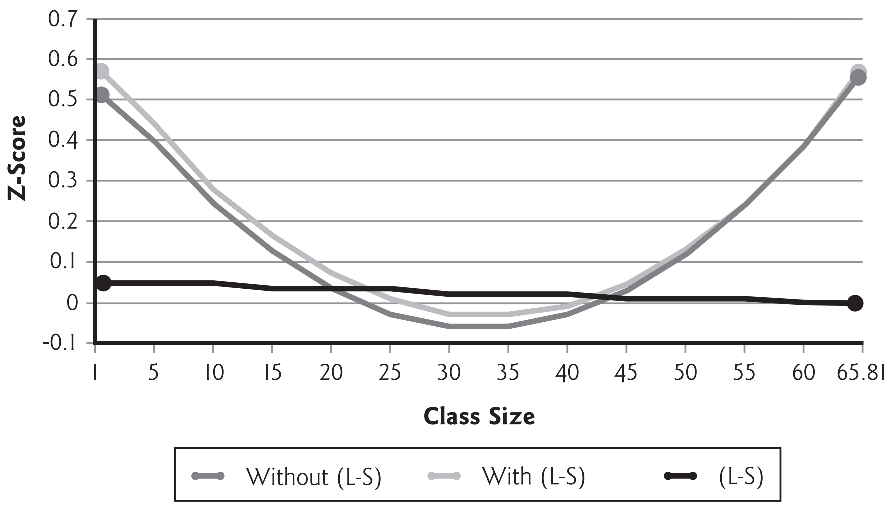
> The lower value, 23 was selected as the pivot point around which to construct the connected curve; the choice was arbitrary and calculations not reported here revealed it to be largely immaterial. The values are for $\Delta_{(S-P)}$ and $\Delta_{(P-L)}$ are as follows for P = 23:
>
> $\Delta_{1\text{-}23} = .551$
> $\Delta_{2\text{-}23} = .513$
> $\Delta_{5\text{-}23} = .407$
> $\Delta_{10\text{-}23} = .254$
> $\Delta_{20\text{-}23} = .037$
> $\Delta_{23\text{-}30} = .001$
> $\Delta_{23\text{-}40} = .009$
>
> Hence, on this curve the difference between achievement in class-sizes 1 and 40 is .551 + .009 = .560... The ordinate is represented by a standard score metric; the zero point (of the graph) is arbitrarily fixed at a class-size of 30 (p. 35).

The reader is urged to pay particular attention to the shift in the calculations due to introduction of the condition: $\Delta_{(S-P)}$ and $\Delta_{(P-L)}$ where P = 23. In the first case, P is substituted for L, and, in the second, P is substituted for S. Therefore, up to S = 23, the variable S changes, and L is fixed; above 23, S is fixed, and L changes. Below S = 23, the relationship is curved (parabolic) while above S= 23 the relationship is linear. In essence, at S = 23, the regression equation changes.

Glass and Smith presented finding for subsets of the data, including "elementary vs. secondary grades" and "well-controlled studies vs. poorly-controlled studies" (pp. 38-42). Several graphs were pre-

https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

Figure 2
Glass and Smith Regression Equation



sented to support their findings, based on the consistency property transformation, not on the derived regression equations. The graphs depict predicted achievement in terms of Z-scores. Finally, because the Z-axis was measured in Z-scores, the final presentation is easily converted into percentiles. Because of the similarities, there is no reason to present the individual analyses; however, the regression coefficients for the subsets of the data set are found in Appendix B of this article.

Glass and Smith closed with this statement: "Taking all findings of the meta-analysis into account, it is safe to say that between class-sizes of 40 and one pupil lie more than 30 percentile ranks of achievement... There is little doubt that, other things equal, more is learned in smaller classes" (pp. 45-46).

Commentary Regarding the Glass and Smith Study

Recall the reason for including the parabola (S^2) and the (L-S) term in the regression equation was presented by Glass and Smith (p. 17) as follows:

> The regression model selected accounted for variations in Δ (S-L) by means of S, S^2 and L. Obviously, something more than a simple linear function of S and L was needed, otherwise a unit increase in class-size would have a constant effect regardless of the starting class-size S; and the S^2 term seemed as capable of filling the need as any other. The size differential between the larger and smaller class, L-S was used in place of L for convenience.

The reason for the consistency property was presented as:

> The problem now is to find the set of $\hat{\Delta}$'s in this surface that can be depicted as a single curved-line relationship in a plane. The property that must hold for a set of Δ's before they can be depicted as a connected graph in a plane is what might be called the consistency property [underline in original]:

$$\Delta_{n1\text{-}n2} + \Delta_{n2\text{-}n3} = \Delta_{n1\text{-}n3} \ldots$$

This section reviews whether the terms S^2 and (L-S) were appropriate choices; whether there are unintended consequences of these choices; and whether the inclusion of the consistency property transformation was warranted.

The Glass and Smith regression equation can be graphed as a two dimensional curve when L is set to a fixed value.[8] The U-shaped curve (parabola) is depicted in Figure 2 with and without the (L-S) term, and the (L-S) term is depicted separately. The value for the large class size is set at 65.81, so (L-S) will equal 0 at the right-hand portion of the graph.

From the presentation of the Glass and Smith results and the graph above, six questions or inconsistencies emerge:

(1) What is the interpretation of the relationship between achievement and class-size? The interpretation of the class size from the graph above seems obvious: As class size changes so does the level of predicted achievement, measured in Z-scores. Of note, the class sizes of 1 and 65.81 predict the same achievement level, with the lowest achievement predicted for a class size of about 33. This is because the S^2 term in the regression equation forms a U-shaped parabola. This representation does not correspond to the conclusion reached by Glass and Smith who report the regression results only to a class-size of 30.

(2) What was the reason for introducing the parabolic curve into the regression equation? Glass and Smith assumed that the relationship between achievement and class-size was nonlinear, and "... the S^2 term seemed as capable of filling the need as any other" (p. 17). No other rationale was provided. The reanalysis section of this paper will explore other options.

(3) What is the interpretation of the relationship between achievement and the (L-S) term? The achievement variable is related to the interval between the large and small class size (L-S). For example, if L = 65.81 and S = 1, then (L-S) = 64.81, with the coefficient of .00082, achievement is predicted to be an additional .053. The (L-S) term adds the most achievement when the class-size is 1 and gradually reduces as class size moves to 65.81, where no achievement is added. In other words, for every pupil added to the classroom, achievement decreases by .00082.[9] In order to make the two dimensional calculations, L must be a fixed value.

(4) What happens if the large class size (L) is set to another value? The value of L determines the relationship of the (L-S) line to the Z-axis (Z-score of 0). If L is set to a lower class size, the (L-S) line shifts lower and, as a consequence, the parabolic curve also shifts lower. In Figure 1, the (L-S) line intersects the Z-axis at a class size of 65.81 because of the value set for L was set at 65.81. Setting a different value to L does not change the basic relationship, only the magnitude of the Z-score; and because the coefficient is small, the magnitude of change is small. The value of L would be important, however, if the regression equation was linear (no S^2 term). In that case, L should be set to the average class-size where the achievement value would also be at the average—a Z-score of 0.

(5) What is the consistency property, and is it necessary? The consistency property transformation is offered for two reasons. Reason one is that the whole must equal the sum of the parts, or the sum of the intervals A to B and B to C must equal the interval A to C. Glass and Smith provided no illustration or example of why the condition was not met in the regression equation and, therefore, the necessity for a transformation. The conditions of the consistency property are met in the presentation of the regression results. (See Table 1.) Moreover, there is no necessity to apply the consistency property transformation to any linear or parabolic relationship. The line and the parabola are in a mathematical class called polynomials, which are continuous functions within the closed interval of the data points; the consistency property is inherent. In all circumstances, the Z-value for the intervals S_1 to S_2 plus the Z-value for the interval S_2 to S_3 equals the Z-value for the interval S_1 to S_3. Therefore, no transformation was necessary. (See also, Appendix A.)

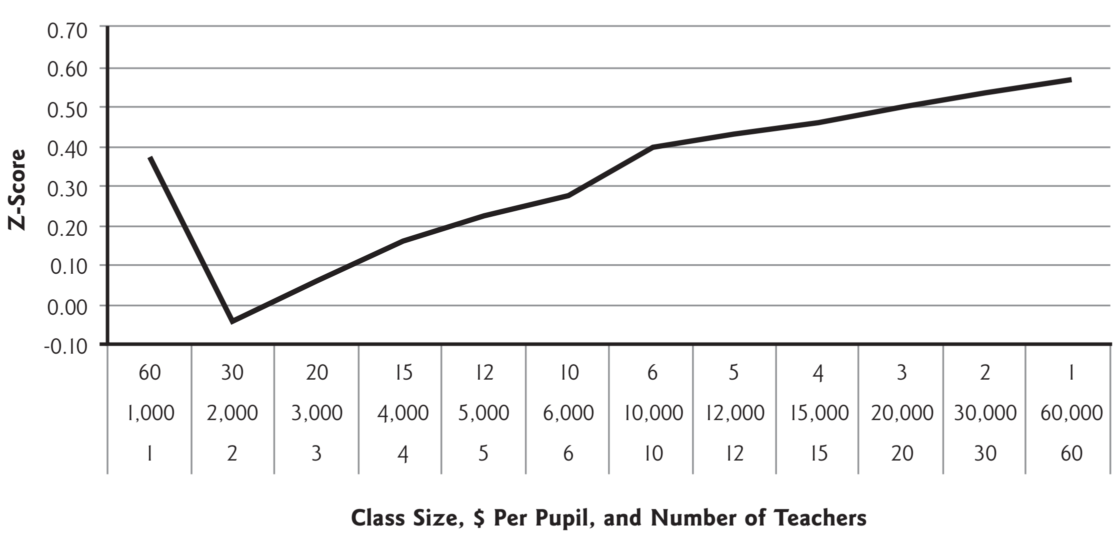
Glass and Smith (p. 18) proposed that the second reason for the consistency property transformation was to produce a "single line curve in a plane" from a three-dimensional surface. Apparently, they assumed the consistency property was related to the (L-S) term and considered it a third dimension. The transformation via the consistency property was not necessary to change a three-dimensional surface into a two-dimensional plane. The change is accomplished by setting L to a fixed value; indeed, setting a value for L is the only way to establish the connected curve in a plane.

The transformation via the consistency property made a fundamental change in the relationship between predicted achievement and class size. Up to P= 23, S was a variable; L was fixed; and the transformation was not applied. Above P = 23, the transformation was applied; S was fixed; and L was the variable. In essence, the transformation was only for values above S = 23 (P= 23). If the whole equals the sum of the parts below 23, then the whole equals the sum of the parts above S = 23, and the transformation is not necessary. If the relationship between achievement and class size is two dimensional below a class size of 23 (by setting the value of L), then it is two-dimensional above 23 (by setting the value of L). The value of L is immaterial to the number of dimensions. Glass and Smith's reasoning is not compelling; their logic is mathematically suspect, i.e., interchanging the character of S and L between fixed and variable.

The parabolic curve was an acceptable solution for class sizes between 1 and 23 because it was consistent with generally held perceptions. Because the parabolic curve was not consistent with perceptions for class-sizes above 33 (the low point), a method was employed that maintained the perceptions and modified the equation, hence the consistency property transformation. It appears that the consistency property transformation was invoked to reconcile the fact that the regression curve moves upward from the minimum and continues upward for all values of small class size, which extend well beyond 66. The value 65.81 is the class size where achievement is virtually the same as a class-size of 1. Essentially, it appears that the consistency property transformation was invoked to avoid this dilemma. If the S^2, the (L-S) terms, and the consistency property were not included in the methodology, there would be no dilemma.[10]

(6) Why are nearly all the value of the Z-scores above zero, when one would expect about half the values to be below the standard score mean of zero? The predicted Z-score values are mostly always above zeros because of the parabola and the consistency property;

Figure 3
Cost Implications of Reducing Class Size: Glass and Smith Regression Equation



https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

in other words, decisions by Glass and Smith. Under normal circumstances, one-half of the observations will be negative; or half will be below average. The reanalysis section of this article will address this issue more specifically.

Class Size or Staff Adequacy?

In the methods section, Glass and Smith discussed the difference between class size and staff adequacy, and provided their reasons for choosing the first for the analysis. No discussion was entertained regarding the potential value of teacher aides, specialized teachers, or administrators as alternatives to increasing the number of classroom teachers. Perhaps there is a way to determine a cost-effective mix of these various educational roles (Phelps 2008).

The staff adequacy measure highlights the number of teachers required to achieve a particular class size, thus shedding light on the potential cost of reducing class size in relationship to increased achievement. For 66 students, it would take only one additional teacher to reduce class size from 66 to 33, for a total of 2 teachers, but it would take an additional 64 teachers to bring class-size to 1. Clearly class size and staff adequacy are on different measurement scales. It is possible to convert the class size ratio, the number of students (S) in a class with one teacher (T), or 1/S, to a measure of staff adequacy (the number of teachers (T) for a given number of students (NS), or T/NS), or 1/S = T/NS). For example, if the class size is 4 (1/4), and the number of students was set at 60 (NS), then 1/4 = T/60 = 15 /60; that is, it would require 15 teachers to have a class size of 4 for 60 students.

When the Glass and Smith regression curve is converted to the staff adequacy measure based on 60 pupils, the cost implications become clear. As class size is reduced, there is an increased cost per pupil (based on $60,000 per teacher) because of the increased number of teachers. As class size is reduced, the predicted achievement does increase (above a class-size of 30), but only up to a point, at which it levels off. Notice the different increments of teachers presented in Figure 3. Initially, class size is reduced dramatically with the addition of 1 teacher. After 10, the number of teachers must increase substantially to reduce class size; the last increment requires 30 additional teachers.

Observations Regarding Glass and Smith

Several initial questions were raised upon looking at the Glass and Smith regression curve. To follow are four observations based on the commentary above.

(1) Why are the relationships all above the 50th percentile? Glass and Smith made a reasonable decision to establish the 50th percentile as the reference point absent any other persuasive point. However, for any distribution only half of the observations can be above the 50th percentile. Their decision creates a strange world where every class size predicts above average achievement. It is logically inconsistent. Is there another way to interpret the situation? The reanalysis in the fifth section of this article addresses this issue.

(2) Why is the relationship curved? Glass and Smith included a squared term in the regression equation because they assumed the relationship between achievement and class size was curved, and the parabolic curve "seemed as capable of filling the need as any other" (p. 17). What is illustrated in the Glass and Smith figure is essentially the left side of the parabolic curve. The right-hand side was modified via the consistency property transformation. Is it possible that the relationship between achievement and class size is not parabolic? The purpose of the reanalysis will be to determine the natural shape of the curve.

(3) Why are the relationships for class sizes above 40 not reported? Glass and Smith used a consistency property to reformulate the original regression equation. The effect of the reformulation was to change the right side of the parabolic curve to avoid the dilemma of having large class sizes predict achievement at the same level as small class sizes. The purpose of the reanalysis will be to account for the full range of data and to address this dilemma.

(4) How many teachers are necessary to reduce the class size from 60 to 1? Figure 2 provides a general idea. Importantly, the measurement scale used in representing the Glass and Smith findings is not an equal interval scale with respect to the number of teacher required to achieve the respective class sizes. The number of teachers and the associated cost of reducing class size increase geometrically. For what class size range might it be cost effective to make the investment? The reanalysis will consider this issue.

Reanalysis of Glass and Smith

When commenting on the Glass and Smith study, two of their methods were questioned: (1) the inclusion of the S^2 and (L-S) terms in the regression equation; and (2) the application of the consistency property transformation.

When discussing the possible analytical methods, given the data available from different studies, Glass and Smith (p. 17) stated: "A few moments reflection will reveal that there is no obvious or simple way to connect these values into a single connected curve." This section tests this statement by proposing another way to connect the data values into a single connected curve. If the results from this other way and Glass and Smith methodology are essentially the same, then their findings will be confirmed. If, however, the results are not the same, then the reader will have to judge the validity of the two approaches and the plausibility of the different results. The purpose of this reanalysis is to identify the relationship between achievement and class size without relying on the questioned methods.

Mathematical Analysis

Glass and Smith provided three critical pieces of data for the reanalysis: (1) the difference in achievement between the smaller and larger classes, measured in Z-scores ($\Delta Z_{(S-L)}$); (2) the small class size (S); and (3) the large class size (L). If the smaller class size has the larger Z-score, the value of the outcome measure is positive, and vice versa. However, $\Delta Z_{(S-L)}$ is not the desired achievement variable for the analysis: Z_S is the desired variable. From these data, the object of this reanalysis is to find a function other than the one presented by Glass and Smith predicting the value of Z for the entire data range of class sizes:

$$Zcs = f\ (CS)$$

The strategy of this reanalysis is to convert each observation from the data set into points on a line segment defined by Z and each class size between S and L. Where the class size points on the line segments are in common, the Z's are averaged. The averages for each class size point are then joined over the full range of class-sizes forming a data-driven curve.[11]

Figure 4
The Relationship Between Achievement and Class Size Based on Reanalysis: Data-Driven Curve

In the section, "Describing the Class-size and Achievement Relationship," Glass and Smith concluded (p. 17), "...various values of $\Delta_{(S-L)}$ arising from different studies can show confusing inconsistencies." This is because various $\Delta_{(S-L)}$ span different ranges of class size. When $\Delta_{(S-L)}$ is divided by (L-S), the inconsistencies disappear. With this value ($\Delta_{(S-L)}$ / (L-S)), a separate value can be calculated for each class size within the range. For example, instead of a single observation for $\Delta_{(10-20)}$, there can be 11 observations—one each for class size, starting with 10 and continuing through 20. With this shift in the paradigm, changing the achievement variable to a Z-score, the necessity for (L-S), S^2, and the consistency property all disappear. This paradigm seems obvious and is clearly less complex.

We start with the definition of the measure of achievement outcome:

$$\Delta Z_{(S-L)} = Z_S - Z_L$$

The achievement measure $\Delta Z_{(S-L)}$ is divided by the difference in the class sizes, CS_L - CS_S, to obtain the slope (M):

$$Z_S - Z_L / CS_L - CS_S = M$$

Therefore, the line segment between SS - SL is:

$$Z_{CS} = M\ CS_{(S-L)} + B$$

where B is the Z-axis intercept. The interpretation of this function is straightforward: For any give value of class size (CS), there is a corresponding value of Z_{CS}, measured as an achievement Z-score. If achievement levels decrease as class size increases, the slope is negative. Conversely, if achievement levels increase as class size increases, the slope is positive. Therefore, the sign of the achievement variable in this context is the opposite of the sign of the achievement variable in Glass and Smith.

With this slope-intercept line function, a new analytical paradigm emerges. The slope for each observation is calculated and a Z-score recorded for each class-size within the line segment. These Z-scores are averaged rather than summarized by a least-squared method because there is no intent to make statistical inferences. By joining these Z-scores into a line, a representation of the relationship between achievement and class-size is obtained directly, independent of any predetermined decisions of the researcher. In contrast, Glass and Smith relied upon the predetermined parabolic function, the (L-S) term, and a consistency property.

The relationship between achievement and class size with the method proposed in this reanalysis can take on any shape—linear, curved, or a combination—and accommodates positive and negative slopes. Using the above interpretation, 40% of the observations in Glass and Smith's data set had positive slopes. If these observations were clustered together in one region of class sizes, there would be a corresponding upswing in the curve. This method also allows for an inspection of the relationship between achievement and class size to determine if it is nonlinear in some ranges and what might be the appropriate curve to fit via future regression analysis.

Data Set for Reanalysis

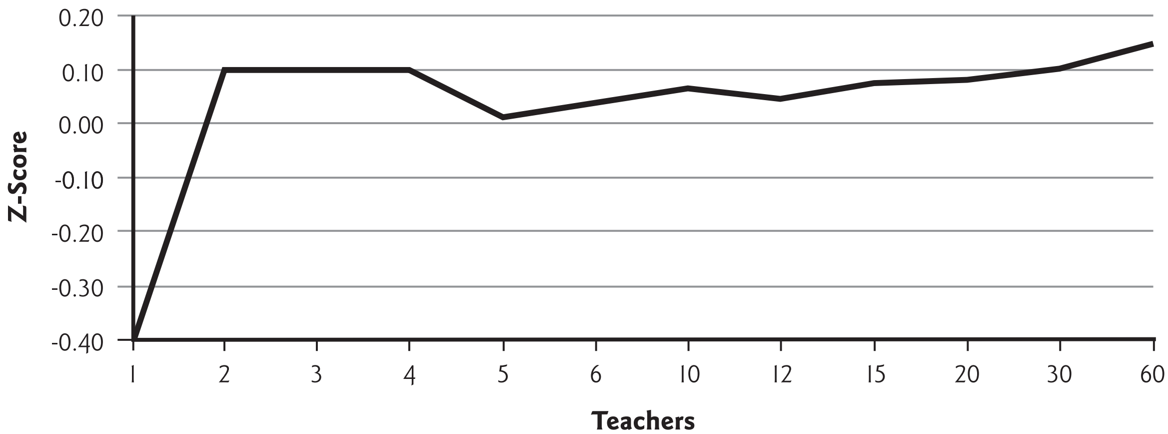
Although Glass and Smith's raw data were listed in an appendix to their study, it is not available in a current electronic format.[12] As a result, the data for this reanalysis were entered by hand from the appendix, but not all data were included. Only data for the categories of elementary school classes (all subjects combined), reading, mathematics, and language were transcribed while the data for the categories of psychology, natural/physical sciences, social sciences and history, and "all others" were excluded. This decision was made for two reasons: First, transcribing was labor intensive; and, second, the categories of elementary school classes, reading, mathematics, and language were considered to be the more relevant subjects in reviewing public school achievement.

Glass and Smith included 725 comparisons taken from 77 studies, including 343 observations for elementary school, 39 in reading, 84 in mathematics, and 144 in language. For the reanalysis data set, there were 309 observations for elementary school, 21 in reading, 84 in mathematics, and 50 in language.

While entering the data, some discrepancies were observed. There were data for the number of pupils and the number or teachers for most of the observations as well as an entry for the ratio of the number of pupils per teacher, but they did not always align. For example, the first data entry for the smaller class size showed 60 students for 10 teachers but with a ratio of 1 instead of 6. There was no way to know the reason for the inconsistency, but because the actual numbers were available, it seemed logical to enter the newly calculated figure rather than the suspicious ratio. This principle was applied to other similar observations. In addition, there was a series of entries with the number of pupils but no entry for the number of teachers. At the same time, the ratio was always

https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

Figure 5
Cost Implications of Reducing Class Size: Data-Driven Curve



1. These observations were not included in the reanalysis because there were a substantial number of observations with a small class size of 1 that could be used.

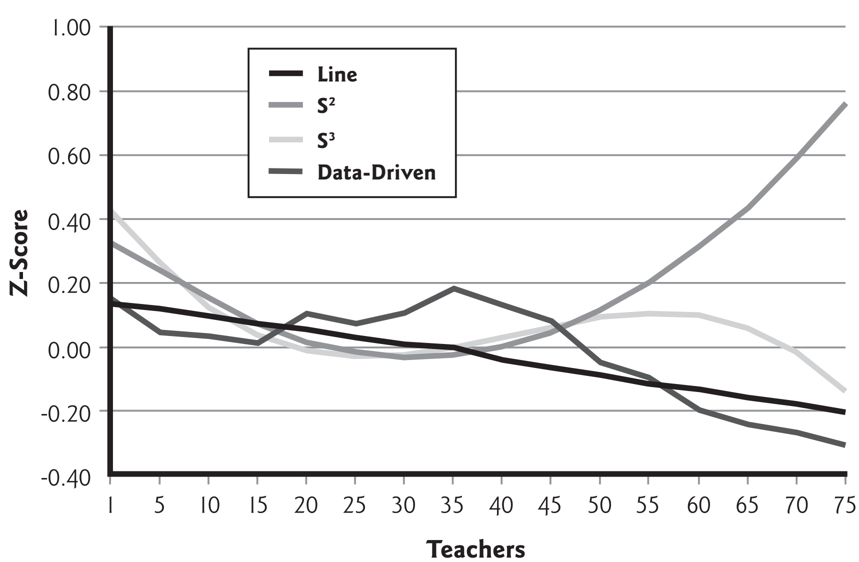
From the reanalysis data set, the slope for each of the observations was calculated and inspected. Four observations had slopes substantially higher or lower when compared to the rest of the data set. These four inconsistent observations were considered extreme outliers and eliminated from the reanalysis. As a result of these decisions, a total 463 observations comprised the reanalysis data set.

What was left was to decide was the value of B, the Z-score intercept. Because the achievement variable was measured in Z-scores, with the midpoint or average at zero, B could be set so that the average class size would correspond to a Z-score of zero. This method of estimating B is not perfect, but it gives some indication of the relative contribution class size makes to achievement over the full range of class sizes. It also avoids the dilemma of having all class sizes predicting above average achievement. The result of the reanalysis is portrayed in Figure 4.

The representation of the data-driven curve presents a more complicated picture of the relationship between achievement and class size than that of the Glass and Smith regression curve. The data-driven curve is essentially U-shaped between 1 and 33, then consistently downward to 75. The predicted achievement level at a class size of about 33 is higher—almost double—than the achievement level at a class size of 1. However, the similarity of predicted achievement between class sizes of 1 and 65.81 is not present, as was the case with Glass and Smith. The substantial number of positive slope observations concentrated between class sizes 15 and 33 explains the upward curve.

From a class size of about 33 upward, there was a continuous and consistent reduction in predicted achievement. The anomalies in the curve at a class size of 54 and between 56 and 60 were due to slopes that are substantially different from the corresponding studies.[13] Removing these observations from the data set would smooth out the descending line.

Figure 6
Comparison of Four Relationships Between Achievement and Class Size



Published by New Prairie Press, 2017

Figure 5 depicts the number of teachers required for 60 pupils in relation to the predicted achievement level. As teachers are added, so does the predicted achievement, moving from one teacher to two, or class size from 60 to 30. However, there is a point where the increase in predicted achievement does not warrant the increase in the number of teachers and associated cost. The policy implications derived from the reanalysis portrayed in this graph are different than those from the staff adequacy transformation of Glass and Smith found in Figure 3.

The cost implications from Figure 5 are straightforward. Moving from a class size of 60 to 30 would require an additional teacher, from one to two, essentially doubling the cost. However, there would be a substantial gain in predicted achievement largely justifying the increased cost. But moving from a class size of 30 to 1 would require another 58 teachers with the amount of achievement gain largely uncertain.

Conclusions

The generalizations made in this section were based on a subset data from the Glass and Smith study. The conclusions were reached by comparing the curves generated using the Glass and Smith regression methodology with the data-driven curve methodology used in the reanalysis. No attempt has been made to include data, findings, or conclusions from other class size research.

In the graph below, four relationships between achievement and class-size are depicted, all based on the revised data set. (See Figure 6.) Three are based on Glass and Smith's regression analysis, and the fourth is based on the reanalysis. The first relationship removes the S^2 term from the Glass and Smith regression equation to form a line; the second, the original equation, includes an S^2 term producing a single-bend curve (parabola); the third includes a S^3 term adding another critical point producing a double-bend curve; and the fourth is the data-driven curve. The three regression curves are continuous curves, so the consistency property transformation is not applied for the reason provided earlier. (See Appendix B for regression coefficients and statistics.)

As can be seen in Figure 6, the line is the most straightforward representation of the relationship between achievement and class size. Predicted achievement decreases as class size increases. The line is inconsistent with the data-driven curve, especially for class sizes in the range of 15 to 35. The single-bend curve (the Glass and Smith regression curve) predicts achievement to decrease as class size increases to about 33, at which point the interpretation becomes counterintuitive—achievement increases.[14] This curve does not resemble the data-driven curve or the linear representation. The double-bend curve suggests a complex relationship between achievement and class size. It somewhat resembles the data-driven curve, but in a different phase. In each of the cases, a problem of interpretation arises:

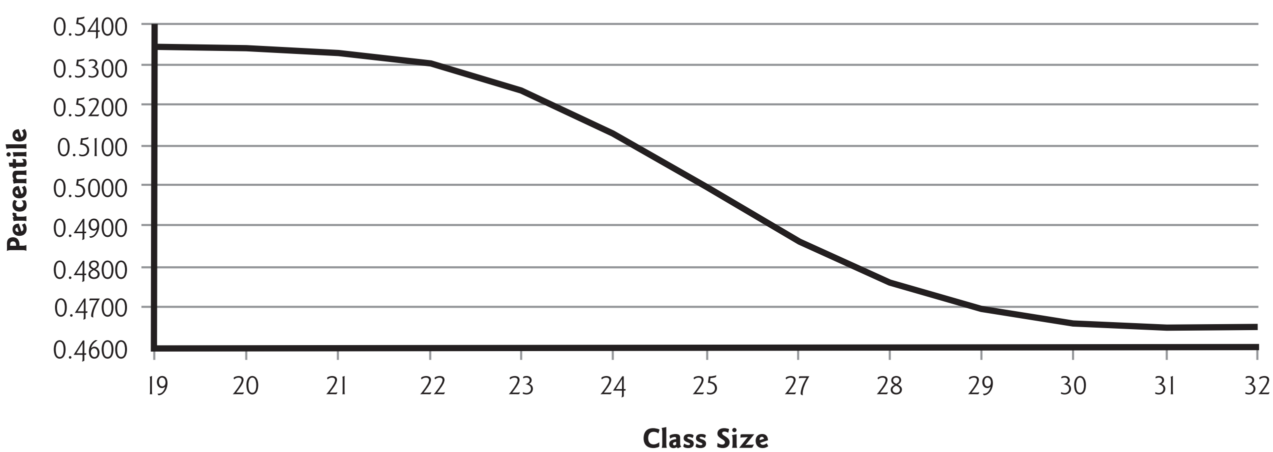
- The line and all curves indicate a gain in estimated achievement as class size moves smaller than about 15.
- There is a predicted gain in estimated achievement as class size moves larger than about 15 for the data-driven curve and about 30 for the two regression curves. The line does not indicate a gain.
- The data-driven curve indicates a drop of estimated achievement as class size moves larger than about 35 while the double-bend curve indicates a drop in achievement as class size moves larger than about 55.

What conclusions can be reached given these indications? The single-bend curve is not supported by the evidence of the data-driven curve or the double-bend curve. While the evidence tends to support the notion that achievement would increase for class sizes smaller than 15, the evidence also supports the notion that such class size reductions are cost prohibitive. The evidence supports the notion that class sizes over a certain size are associated with a decrease in achievement; the exact critical point is in doubt based on these data and analyses. In contrast, the evidence does support lowering class sizes from the large extremes, and there are indications that the potential gain would offset the marginal cost. The influence of class size between about 15 and about 45 is unclear, other than the general conclusion that the relationship between achievement and class size is indeed complicated as Porwell (1978) suggested.[15]

Representing the Relationship between Achievement and Class Size Based on a Normal Curve

If one would make some basic assumptions regarding a class size curve, what would those assumptions be? First at the larger

Figure 7
Theoretical Relationship between Achievement and Class Size: Hypothetical Curve



https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

class sizes, it would be fair to assume that by adding one student to a class of 100 students, there would be little if any difference in achievement. With this assumption, a well-matched curve would show gradually decreasing achievement as class size increased approaching a lower bound; i.e., a lower-bound asymptote. Second, at the smaller end of class size, it would be fair to assume that the difference in achievement by removing one student from a class of 5 students would show gradually more achievement as class size decreased approaching an upper bound of 1; i.e., an upper-bound asymptote. Third, it would be fair to assume that the average class size would predict the average achievement level. Finally, it would be fair to assume that all class sizes above the average would predict achievement below the average and vice versa for class sizes below the average predicting achievement above average. These assumptions address the difficulties with the data-driven and two-bend curves presented previously.

There is a curve meeting these conditions. This curve has its roots in normal curve statistics and provides a more reasonable explanation than the other curves. The details are explained fully elsewhere (Phelps, 2008). In summary, the amount of variance explained by a regression equation can be converted to the curve in Figure 7. With the dependent and independent variables measured in standard scores (Z-scores), the amount of variance explained (R^2) can be converted into a normal curve with the same area. When the normal curve is integrated (cumulative area under the curve), the result is an S-shaped curve, asymptotic at the upper and lower bounds, with the average class size predicting the average achievement.

Determining the amount of variance explained by class size is complex because class size is likely to be correlated with other important variables such as socioeconomic status (SES), expenditures, teacher qualifications, support staff, and instructional materials. Studies with these variables could provide estimates of possible ranges of the variance attributable to class size; these estimates can be instructive in policy decision-making (Phelps, July 2008). While the data set from the Glass and Smith study is not suitable for this type of analysis, at least an example can be offered. This example has an average class-size of 25, a standard deviation of 2, and an R^2 of .07 (the average R^2 of the three regression curves is .07).

In reality, class size does not range from 1 to 70, as does the data set, but is more likely to be in the range suggested above. More likely, the curve has a consistently downward slope. It would seem that the likely relationship between achievement and class size is more similar to the curve suggested in Figure 7 than the complex curves depicted in Figure 6.

In summary, there is a likely relationship between class size and achievement, but the relationship is exceedingly complex. At the same time, the financial cost of reducing class size as a primary method of increasing achievement is not warranted. The conclusion to be drawn from these three points is that the substantial influence of Glass and Smith (1978) in changing policy related to class size was/is probably unwarranted. In the final analysis, the class size policy question comes down to what is believed and what is accepted. Does one believe in the analytical results and accept the methodology, or does one believe in the methodology and then accept the results?

References

Glass, Gene V., and Mary Lee Smith. *Meta-Analysis of Research on the Relationship of Class-size and Achievement*. San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978.

Phelps, James L. "Optimizing Educational Resources: A Paradigm for the Pursuit of Educational Productivity." *Educational Considerations* 35 (Spring 2008): 3-18.

____________. "Measuring and Reporting School and District Effectiveness." *Educational Considerations* 36 (Spring 2009): 40-52.

Porwell, P.J. *Class Size: A Summary of Research*. Arlington, VA: Educational Research Service, Inc., 1978.

Endnotes

[1] All subsequent references to Glass and Smith in this article refer to Gene V. Glass and Mary Lee Smith, *Meta-Analysis of Research on the Relationship of Class-Size and Achievement* (San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978).

[2] Meta-analysis is a research method that takes data from many individual studies and combines them into a new analysis.

[3] This is a curious statement. What if the situation were shifted to a supermarket where the price of potatoes was 1 lb. for 70 cents, 3 lbs. for $2.00, 5 lbs. for $3.10, 10 lbs. for $6.00, and 25 lbs. for $13.00? What would the shopper do?

[4] In the previous section, emphasis was added to three points. These points are critical in later portions of this paper: (1) No obvious and simple alternative; (2) including the S^2 term; and (3) including (L-S) term.

[5] See Appendix A of this article for a detailed discussion of the consistency property.

[6] The sum of the intervals should equal 0, as it does considering the rounding error.

[7] The differences in the values in the Table 1 are due to a different value being set for the large class size.

[8] The equation can be graphed in three dimensions with L being the third, starting with 1 and continuing to the largest class-size in the data set. To determine a point on the surface, an arbitrary value for L must be selected in order to evaluate (L-S).

[9] While the relationship between achievement and class-size—the S variable—is parabolic, the relationship between achievement and (L-S) is linear.

[10] When Glass and Smith added a squared term to their equation representing the relationship between class size and achievement, they applied the same mathematical function used to describe a thrown ball—a parabola. So whether intended or not, their class size curve and a thrown ball should follow the same general path. If their parabola assumption were based on fact rather than supposition, and if their consistency property were mathematically correct, then by mathematical symmetry, a thrown ball would follow the upside-down Glass curve (Figure 1) and would never

come down! Conversely, if the thrown ball path is correct, then their squared term assumption, their consistency property, or both, are faulty.

[11] For example, to find the best price per pound, divide the price by the number of pounds. The shopper determined the cost per pound in cents was 70, 67, 62, 60, and 52. These numbers can be placed into a curve depicting the price per pound for various packaging weights.

[12] Author's correspondence with Gene Glass.

[13] See observation #369, study #55, and observation #373, study #4.

[14] Achievement at class-size 1 and 61 (rather than 65.81) is the same because of the change in the data set.

[15] The data-driven curve generated by the reanalysis is complicated to explain; that is, why is a class-size of 33 be the best level for achievement? One must take into consideration that the data in the reanalysis may not be representative, and hence other data sets and other paradigms should be used to test the underlying question.

https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

APPENDIX A

Discussion Regarding the Consistency Property

Glass and Smith (1978, 17) stated [italics added for emphasis]:

Fitting this model by least-squares will result in the curved regression surface:

$\hat{\Delta}_{(S-L)} = \hat{\beta}_0 + \hat{\beta}_1 S + \hat{\beta}_2 S^2 + \hat{\beta}_3 (L-S)$

The problem now is to find the set of $\hat{\Delta}$'s in this surface that can be depicted as a single curved line relationship in a plane. The property that must hold for a set of $\hat{\Delta}$'s before they can be depicted as a connected graph in a plane is what might be called the consistency property:

$\Delta_{n1-n2} + \Delta_{n2-n3} = \Delta_{n1-n3}$

for n1 < n2 < n3. *If this property is not satisfied, then one is in the strange situation of claiming that the differential achievement between class-size 10 and 20 is not the sum of the differential achievement from 10 to 15 and then from 15 to 20.*

When the consistency property is imposed on [regression equation] *(2)*, it follows that:

$\hat{\beta}_0 + \hat{\beta}_1 n_1 + \hat{\beta}_2 n_1^2 + \hat{\beta}_3 (n_2-n_1) + \hat{\beta}_0 + \hat{\beta}_1 n2 + \hat{\beta}_2 n2^2 + \hat{\beta}_3 (n_3-n_2) =$
$\hat{\beta}_0 + \hat{\beta}_1 n_1 + \hat{\beta}_2 n_1^2 + \hat{\beta}_3 (n_3-n_1)$ (3)

Simple algebraic reduction produces the following:

$\hat{\beta}_0 + \hat{\beta}_1 n_2 + \hat{\beta}_2 n_2^2 = 0$

The two solutions to the quadratic equation...are points n_2 such that the $\hat{\Delta}$ is measured with n_2 as either the larger, L, or smaller, S, class size then the resulting set of $\hat{\Delta}$'s will lie on the four dimensional regression curve...but can be depicted as a single line curve in a plane. Since n_2 becomes the point around which values of n_1 and n_3 are selected, it will be called the pivot point [emphasis in original]. That there are two solutions for n_2 is perplexing; fortunately in the analyses to be reported the two corresponding curves were virtually parallel in practice.

A single line curve in a plane can be constructed by solving for one or the other values of n_2 in (4) and constructing a set of $\hat{\Delta}$'s values. *These values will give the standardized mean differences in achievement between n_2 and any other class size.* The curve that connects these $\hat{\Delta}$'s has no non-arbitrary starting point. *One can assume for convenience sake that the achievement curve (z), instead of the differential achievement curve ($\hat{\Delta}$)* is centered around an arbitrary class size, e.g., something like the national average in the low 20s (pp. 17-19).

The purpose of this discussion is to test the assumptions underlying the consistency property as described above. (Note the italicized passages.)

1. Under what circumstances is the differential achievement between class size 10 and 20 the sum of the differential achievement from 10 to 15 and then from 15 to 20?
2. Can the consistency property be logically imposed on the regression equation?
3. If the consistency property cannot be logically imposed on the regression equation, is there an alternative formulation?
4. What is the nature of the achievement variable? The achievement variable in the data set is $\Delta_{(S-L)}$, but why has the interpretation changed to a Z-score after the regression coefficients have been applied to the equation?
5. What are the consequences of the alternative formulation?

In order to critique the "imposition" of the consistency property (equation (3)) on the regression equation (equation (2)), three achievement values must be obtained—one each for three sequential and equidistant class sizes (e.g., class-sizes of 10, 15, and 20 as suggested). For the critique, the selected coefficients values are: $\beta_0 = 2$, $\beta_1 = -.1$, $\beta_2 = 0$, and $\beta_3 = .01$. These values have been set to make the calculations simpler and clearer (eliminating the squared term making the relationship linear). The selection of the values does not affect the underlying principles or conclusions. Substituting these values, regression equation (2) becomes: $\Delta = 2 - .1S + .01(L-S)$. The consistency property in equation (3) can be expressed as three equations where the sum of the first two equals the third ($\Delta 1 + \Delta 2 = \Delta 3$):

$\Delta 1 = \beta_0 + \beta_1 n_1 + \beta_2 n_1^2 + \beta_3 (n_2-n_1)$ or $2 - .1 * 10 + .01(15-10) = 2 - 1 + .05 = 1.05$
$\Delta 2 = \beta_0 + \beta_1 n_1 + \beta_2 n_1^2 + \beta_3 (n_2-n_1)$ or $2 - .1 * 15 + .01(20-15) = 2 - 1.5 + .05 = 0.55$
$\Delta 3 = \beta_0 + \beta_1 n_1 + \beta_2 n_1^2 + \beta_3 (n_2-n_1)$ or $2 - .1 * 10 + .01(20-10) = 2 - 1 + .1 = 1.10$

The algebraic reduction of the equations (3) becomes:

$\beta_0 + \beta_1 n_2 = 0$ or $\beta_0 = - \beta_1 n_2$ (4)

Equation (3) is false ($\Delta 1 + \Delta 2 \neq \Delta 3$). Also, equation (4) is false ($2 + (-.1 * 15) \neq 0$). Equation (3) will be true only when $n2 = - \beta 0 / \beta_1$, or -2/- .1, or a class size of 20 which contradicts the initial condition of n2 = 15. The equations proposed by Glass and Smith for meeting the consistency property conditions are unsatisfactory. The task is to identify a workable alternative formulation.

The solution to consistency property equations will be clearer if the regression equations are graphed. Graphing the expression $\Delta = \beta_0 + \beta_1 S$ is straightforward: the expression is represented by a line with a slope of -.1 and the Δ intercept of 2 (at S = 0, Δ = 2). Graphing the expression $\Delta = \beta_3$ (L-S) is problematic; while the slope is .01, there is not a consistent intercept. For Δ, the intercept is 15 (when S = 15,

APPENDIX A *continued*

(L-S) = 0 and Δ = 0). For the other two equations the intercept is 20. In other words, L is the intercept, and it is not the same in each equation. As a result, the (L-S) term produces a family of lines and not a single line, as with the other expression. This difference between the two expressions is critical.

Looking for an alternative, there are two primary criteria: (1) Δ1 + Δ2 must = Δ3; and (2) because the equations are linear (by setting the squared term to 0) and the class-sizes are sequential and equidistant, the values of Δ1, Δ2, and Δ3 must also be sequential and equidistant.

In the first test for an alternative, the large class size is set to a fixed value (L=20), and Δ3 is calculated with the value of the third class size:

$\Delta 1 = 2 - .1*n_1 + .01*(L-n_1)$ or $\Delta 1 = 2 - .1*10 + .01*(20-10) = 1.10$

$\Delta 2 = 2 - .1*n_2 + .01*(L-n_2)$ or $\Delta 2 = 2 - .1*15 + .01*(20-15) = 0.55$

$\Delta 3 = 2 - .1*n_3 + .01*(L-n_3)$ or $\Delta 3 = 2 - .1*20 + .01*(20-20) = 0.00$

Again, Δ1 + Δ2 ≠ Δ3! However, Δ1, Δ2, and Δ3 are sequential and equidistant. The situation does not change if L is set to another value, although Δ1 + Δ2 does = Δ3 at L = 291. But if any of the class-sizes change, so does the value of L; so there are an infinite number of solutions to the equations! Interestingly, the average class size must be 20, for when S =20, achievement is predicted to be 0, and the average class size equals the average achievement (a Z-score of 0). In order to evaluate the regression equation, L must be set to a constant to preserve a consistent relationship among the class-sizes. The achievement variable is not measured in terms of Δ and/or the formulation is incorrect in that the whole is not the sum of the parts but is correct in that the values are sequential and equidistant. Equation (2) is true. Even with the change, equation (3) is not true.

For the second test for an alternative, the achievement variable is assumed to be Z-scores, and the Δ is assumed to be the difference between two Z-scores, or: Δ1 = (Z2 – Z1), Δ2 = (Z3 –Z2), and Δ3 = (Z3-Z1), or $(f(s2) - f(s1)) + (f(s3) - f(s2)) = (f(s3) - f(s1))$.

$Z1 = 2 - .1*n_1 + .01*(L-n_1)$ or $Z1 = 2 - .1*10 + .01*(20-10) = 1.10$

$Z2 = 2 - .1*n_2 + .01*(L-n_2)$ or $Z2 = 2 - .1*15 + .01*(20-15) = 0.55$

$Z3 = 2 - .1*n_3 + .01*(L-n_3)$ or $Z3 = 2 - .1*20 + .01*(20-20) = 0.00$

Substituting, (.55- 1.10) + (.00 - .55) = (.00 – 1.10) or (-.55 - -.55) = -1.1. Both criteria are met. Therefore, when L is set to a fixed value, the achievement variable is measured in Z-scores, and Δ is the difference between two Z-scores, "…the differential achievement between class-size 10 and 20 is…the sum of the differential achievement from 10 to 15 and then from 15 to 20" (p. 18). With this interpretation, the logical condition is met, and the regression equation is graphically portrayed not as a surface but as two lines which, when added together form a "single curved-line in a plane." This interpretation is consistent with the results presented in Table 1 using the actual regression equation.

Under Glass and Smith's overly-complicated consistency property formulation, the logical condition is not met. In practice, they do set L to a fixed value, but make other changes, which are discussed in this article. Based on this analysis, it is inappropriate to apply the consistency property transformation.

APPENDIX B

Table B-1
Regression Coefficients from Glass and Smith Meta-Analysis

Studies	*Intercept*	*S*	*S^2*	*R^2*
Elementary students	0.38503	-0.02995	0.00052	0.255025
Secondary students	0.75539	-0.05024	0.00071	0.192721
Poorly controlled	0.07399	-0.00587	0.00009	0.034969
Well controlled	0.69488	-0.06334	0.00128	0.385641
All	0.57072	-0.03860	0.00059	0.181476

Source: Glass and Smith (1978, 33, 39). R^2 calculated by author from multiple R.

https://newprairiepress.org/edconsiderations/vol39/iss1/3
DOI: 10.4148/0146-9282.1100

APPENDIX C

Table C-1
Regression Coefficients, R^2, and Z-axis Intercepts from Reanalysis

	Intercept	*L-S*	*S*	*S^2*	*S^3*	*R^2*	*Z = 0*	*Z = 0*	*Z = 0*
Line	0.141156	0.002786	-0.004679			0.034	30.76		
S^2	0.356798	0.002891	-0.025273	0.000407		0.084	22.09	40.00	
S^3	0.461896	0.003502	-0.045211	0.001270	-0.000010	0.098	18.32	35.56	69.21