A Sixty Million Dollar Statistical Issue Arising in the Interpretation and Calculation of a Measure of Relative Disparity: Zuni Public School District 89 v. U.S. Department of Education

JOSEPH L. GASTWIRTH
Department of Statistics, George Washington University, Washington DC 20052, USA
(This paper has been accepted, subject to minor editing, by *Law, Probability and Risk*)
In 1994, the U.S. Congress rewrote the statute providing educational funds to towns and cities of the nation where the federal government owns substantial amounts of land, which cannot be taxed locally. Usually, this impact aid goes to the local school districts; however, if a state has a program for subsidizing education in all areas so that the per-pupil expenditures are approximately equal throughout the state, then the state receives a large share of the federal funds. The previous statute authorized the Department of Education to establish a formula for determining whether a state's educational funding was "equalized". The new law specifies its own formula. Both formulas place an upper limit on a percentile-based measure of disparity calculated from data on average per-pupil expenditures in the local educational areas of the state. While the formulas are similar, they describe the percentiles to be used in the calculation differently. In 1996, when the Department of Education published its regulations for administering the new law it interpreted the new formula for assessing whether the educational funding in a state was "equal" to mean the identical calculation it had previously issued. In 1999 the state of New Mexico claimed that its school funding system met the criterion for "equalization" so it deserved 75% of the federal impact funds allocated to all impacted school districts in the state. Two local districts filed suit claiming that the state's funding did not meet the new statute's criterion for being "equalized" so the entire federal payment should go to the affected school districts. This article describes the calculation most statisticians would make after reading the new law. It will be seen that none of the calculations presented by the parties is technically correct. Furthermore, one of the assertions made by the Department of Education and accepted by the majority in a split appellate opinion is not mathematically correct. An *en banc* opinion of all twelve Circuit judges split on the interpretation of the statute and the U.S. Supreme Court has accepted the case for review. The last section suggests that alternative disparity measures based on data on expenditures in schools rather than districts would give a more accurate indication of the disparity in school funding in a state. Under either interpretation of the current law, a disparity in educational funding that courts previously have found to support a claim of racial segregation or discrimination, can satisfy the criteria for "equal" educational funding.

Keywords: calculation of percentiles, disparity measures, educational funding, Federal Impact Aid Act, statutory interpretation,

1. Introduction

Education in the United States is considered a local or state responsibility and is supported, in large part, from local property taxes. Since facilities owned and operated by the federal government are exempt from state and local taxes, the federal Impact Act[1] provides for payments to an affected school district or local educational agency (LEA), to offset the district's costs incurred in educating children connected with the federal presence in the area. The federal payments are to be used for the education in the affected LEAs rather than to subsidize the education in the state as a whole. Hence, the Impact Act[2] generally prohibits the state from counting the federal aid as a local resource when determining the amount of state money that an LEA receives.

A complication arises when states supplement local funds for education funds in poorer districts with state-wide funds. Essentially, these states transfer money from wealthier to poorer districts to "equalize expenditures" per-pupil across the state. The Impact Act allows a state having an educational funding "equalization" system to share an appropriate portion of the federal funds allocated to the federally impacted districts since all of the districts in the state bear the increased cost of educating children living in federally impacted districts.

Since it is unreasonable to expect an equalization program to produce per-pupil expenditures that are exactly equal in every LEA, it is necessary to decide when those expenditures are sufficiently uniform for the exception to apply. As originally adopted in 1974, the law left it to the Secretary of Education to define the term "equalize expenditures."[3] In 1976, the Department of Education (DOE) adopted the criterion of a 25% disparity or less between the LEA at the 95th percentile and the LEA at the 5th percentile, where the percentiles are determined according to the number of pupils in each LEA.[4] For example, the DOE calculates the 5th percentile by first ordering the LEAs in increasing order of their expenditures and then accumulates their pupil counts until it reaches 5% of the total number of pupils. The expenditure of that LEA is the 5th percentile.

In 1994, Congress replaced the old statute[5] with the current version. The replacement incorporates the 25% limit, and it looks to the difference between the 95th and 5th percentile. However, it does not state that the percentiles should be determined by the number of students in the districts as did the original regulation. The new statute also provides for the Secretary to consider special expenditures in the state's aid to LEAs that reflect special circumstances, e.g., geographic isolation.

---

[1] 20 U.S.C. Sec. 7709 *et seq.*
[2] 20 U.S.C. Sec. 7709 (a).
[3] *Id.* at Sec. 240(d)(2)(B)(1982).
[4] 41 Fed. Reg. 26320, 26327 (June 25, 1976).
[5] 20 U.S.C. Sec. 240(d).

Following passage of the new statute the DOE issued a revised regulation, which interpreted the disparity formula in the new statute to specify the very same calculation it issued in its 1976 regulation. Using the old formula, in 1999, New Mexico submitted a request to be certified as "equalized" and thus, entitled to 75% of the 62 million dollar federal payment to federally impacted school districts for the 2000 school year alone. Otherwise, the entire payment would be allocated to the federally affected school districts. The DOE approved the state's application. The Zuni School District and the Gallup-McKinley School District filed an objection with DOE, claiming that the new statute states that the percentiles should be calculated by from the number of districts. The 25% rule is not met by their calculated disparity, and hence, the state was not entitled to any portion of the federal assistance for the students in the two districts.

An administrative law judge in the DOE found for the state of New Mexico and the Secretary of Education (who is the head of the DOE). The districts sought review of this decision in the Court of Appeals for the Tenth Circuit. A three judge panel upheld the administrative determination, with one judge dissenting.[6] On rehearing the case *en banc*, the Tenth Circuit split 6-6, letting the original decision by the administrative law judge stand. The school districts petitioned the Supreme Court for a writ of certiorari. The Court granted the writ and placed the case on its docket for the 2006-2007 term.[7]

Courts typically defer to the interpretation of the relevant agency if the law gives the agency the authority to promulgate rules and regulations or when a gap in the law needs to be filled or the statute is unclear or ambiguous.[8] In contrast, when Congress does not give the agency the power to issue the rules and regulations less deference may be given to the interpretation of the agency.[9] Indeed, in *United States v. Mead*[10] the Court cites several cases stating that the weight courts should give an agency's interpretation of a law it administers varies with the Department's degree of care, relative expertise, consistency and its use of formal procedures.

Apparently, the Court accepted the case to clarify the weight given to an agency's interpretation of a statute. This important legal question is not the focus of this paper. Rather, we examine the statute and its formula, which involve statistical concepts,[11] from

[6] Zuni Pub. School Dist. No. 89 v. U.S. Dep't. of Educ., 393 F.3d., 1158 (10th Cir. 2004).
[7] Zuni Pub. Sch. Dist. No. 89 v. Dep't. of Educ. No. 05-1508. Sept. 26, 2006.
[8] Chevron USA., Inc. v. Natural Resources Defense Council, Inc. 467 U.S. 837, 842-43 (1984); FDIC v. Meyer, 501 U.S. 471, 476 (1994).
[9] United States v. Mead Corp., 533 U.S. 218, 220 (2001) (citing EEOC v. Arabian American Oil Co. 499 U.S. 244 at 257 (1991).
[10] *Id.* at 219.
[11] In defining percentile, two respected dictionaries note that it is a statistical term. The RANDOM HOUSE WEBSTER'S UNABRIDGED DICTIONARY 1437 (2001) defines percentile as "*Statistics*. 1. one of the values of a variable that divides the distribution of the variable into 100 groups having equal frequencies. *Ninety percent of the values lie at or below the ninetieth percentile, ten percent above it.*" THE NEW OXFORD AMERICAN DICTIONARY 1262 (2d ed. 2005) defines "percentile" as "Statistics…each of the 100 equal groups into which a population can be divided according to the distribution of values of a particular variable—each of the 99 intermediate values of a random variable that divide a frequency distribution into 100 such groups." The definition provided by THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE1304 (4th Ed. 2006) defines a "percentile" as "one of a

the perspective of a statistician. Hopefully, this viewpoint will be useful to the legal community in assessing the "meaning" of the statute and how well the process used by the DOE in issuing its interpretation of the formula in the current law meets the criteria mentioned in *Mead*.

Both the current statute and the accompanying regulation, which was identical to the one the DOE created under the previous law, are described in Section 2. The two calculations presented by the parties, along with a brief summary of their arguments supporting their interpretations, are also discussed there. The third section explains how a statistician would read the statute and make the calculation. It shows that none of the calculated disparities presented to the Tenth Circuit is completely accurate. Using the correct definition of a percentile calculated from data on every member of the statistical universe, the disparity under the plaintiffs' interpretation is 32.4% instead of 26.9%. The defendant's disparity measure increases only slightly to 15.1% from 14.43%. Furthermore, the statement in the published regulations that exclusion based on a percentage of school districts, rather than percentage of the pupil population, might exclude a substantial percentage of students in states with a small number of large districts is only true when there are at least twenty districts in the state.[12] This section also reports the results of an informal survey of statisticians on how they read and interpret the formula. Virtually all agreed, conceptually, with the plaintiff's primary interpretation and the dissenting opinion by Judge O'Brien. The last part of section 3 questions a formula the DOE established for combining separate calculations of its disparity measure in LEAs with different grade systems into a statewide aggregate. The fourth section discusses issues that are more relevant to improving the measurement of educational disparities in any future law. It stresses the advantages of using school-wide expenditure data rather than the highly aggregate LEA per-pupil data. The section proposes an alternative disparity measure, the coefficient of dispersion, which has been used for many years in assessing the fairness of tax assessments, for consideration by DOE and Congress. The data and more detailed calculations of the various measures are presented in an Appendix.

2. The Statute and Interpretations of the Formula in the Case

a. Summary of Relevant Parts of the Statute and Regulatory Formulas

While states are generally prohibited from considering Federal impact aid to an LEA in determining the amount of the state's contribution to the LEA, an exception exists for states that have a program for equalizing per-pupil expenditures throughout the state. When Congress rewrote the law in 1994, it adopted the following disparity test to assess whether a state's funding program was equalized: Equalization is present--

---

set of points on a scale arrived at by dividing a group into parts in order of magnitude. For example as score equal to or greater than 97 percent of those attained on an examination is said to be in the 97th percentile." While this dictionary did not mention statistics, the definition given is very similar to the previous ones.

[12] *See* 41 Fed. Reg. 26,324 (1976), which was cited in the first appellate opinion, *supra* n.6 at 1164 and the government's brief in opposition to granting cert. at 4. The mathematics supporting our assertion is given in Section 3a

if, in the second fiscal year preceding the fiscal year for which the determination is made, the amount of per-pupil expenditures made by, or per-pupil revenues available to, the local agency in the State with the highest per-pupil expenditures or revenues did not exceed the amount of such per-pupil expenditures made by or per-pupil revenues available to, the local educational agency in the State with the lowest such expenditures or revenues by more than twenty-five percent (25%).

In making a determination under this subsection the Secretary shall—

(i) disregard local educational agencies with per-pupil expenditures or revenues above the 95th percentile or below the 5th percentile of such expenditures or revenues in the state; and

(ii) take into account the extent to which a program of State aid reflects the additional cost of providing free public education in particular types of local educational agencies, such as those that are geographically isolated, or to particular types of students, such as children with disabilities.

Before describing the original law, readers should think about the data they would need to obtain and the calculation they would make to follow the statute. The previous statute[13] gave the Secretary of the Department of Education (DOE) the authority to develop the disparity test to be used in determining whether public school funding in a state was "equalized". After carrying out the normal rulemaking procedure and considering the comments submitted to it, in 1976 the DOE published its disparity measure and described its calculation:[14]

If there is a disparity of no more than 25 percent in revenues per pupil (or other unit of pupil need used in the state program) available to the 95th and 5th percentile school districts (those with the 95th and 5th percentiles of the total number of pupils after being ranked in order of revenue per pupil) the program would be deemed to qualify under section 5(d) (2).

The notice states that this standard was chosen because it is a method of evaluating school finance programs in terms of equalization that has been used by courts

[13] Pub. L. 81-874 Section 5(d).

[14] 41 Fed. Reg. 26320, 25327 (June 25, 1976). In addition to the statistical formula the notice indicates the types of expenditures that should be included or excluded. For example, funds used for capital outlays or debt service by the LEAs are excluded as the law concerns current expenditures on education. Although the proper classification of expenditures has an important role in ensuring the accuracy of the data used in the disparity calculation discussed in this paper that topic will not be considered here.

and authorities in school finance.[15] It also states that "The exclusion of the upper and bottom 5 percentile school districts is based on the accepted principle of statistical evaluation that such percentiles usually represent unique or non-characteristic situations."[16]

The 1976 publication gives directions for calculating the disparity measure. First, the districts in a State will be ranked on the basis of current expenditures or revenue per pupil and those districts which fall above the 95th percentile or below the 5th percentile in terms of the number of pupils in attendance in the schools of those agencies will be excluded from the calculation. Indeed, the regulation emphasizes that the percentiles will be determined on the basis of numbers of pupils and not on the number of districts. The stated purpose of the exclusion is to eliminate anomalous characteristics of the distribution of expenditures. The notice observes that if a state had a large number of small districts deleting the highest and lowest 5 percent might exclude an insignificant fraction of the population of pupils. In contrast, in states with a small number of large districts, an exclusion based on a percentage of school districts might exclude a substantial portion of the pupil population.[17]

After the law was changed in 1994, the DOE interpreted part **(i)** of the new disparity standard to require the identical calculation it established in 1976. While implicit in the calculation of the average per-pupil expenditure (or available revenue), the number of pupils in a district is not mentioned in the new statute, much less used to determine the 5th and 95th percentiles. In contrast the regulation DOE established focuses on the distribution of per-pupil expenditures in the state, explicitly considering both the number of pupils and the per-pupil expenditure in each LEA.

. b. A Brief Description of the Case

In February, 1999, New Mexico notified the DOE that intended to take Impact Aid payments into account in allocating school funds for July 1, 1999-June 30, 2000 and submitted revenue and student population data (reproduced in the Appendix) for the 1997-1998 school year. Following the DOE formula, the state calculated that Hobbs was the LEA at the fifth percentile and had a per pupil revenue of $ 2,848. The Penasco LEA was the 95th percentile with revenue of $3,259 per pupil. The difference is $411 so the disparity measure = 411/2848= 14.43%. Hence, the DOE certified that the state's

---

[15] *Id.*at 26320. Although references to the literature are not given, we note that STEPHEN J. CARROLL and ROLLA EDWARD PARK (Ballinger, 1983) THE SEARCH FOR EQUITY IN SCOOL FINANCE focus on the weighted measures but refer to related research where data for each LEA is analyzed, *Id.* at 32-33. These authors acknowledge that their research was supported by the National Institute of Education.

[16] *Id.* at 26320. Again, no reference to a Statistics text or peer-reviewed article is cited in support of this assertion.

[17] Unfortunately, the notice, *supra* n. 14 at 26324, did not discuss these statements in depth nor did it report actual data from a state illustrating the issue. Later in Section 3 it will be seen that even in a state with many small districts and a few large ones, under the DOE formula it is possible that no data will be deleted rather than a large fraction of the student population of the state.

educational funding to be equalized so it could consider the Impact Aid funds in allocating the state contribution, i.e. the state would receive the equivalent of 75% of the federal payment.

In late 1999 two districts, Zuni and Gallup-McKinley filed an objection to the certification. These districts argued that this calculation directly conflicts with the new statute, which they read to require that the five percent of the LEAs at the top and bottom of the distribution of available revenue be deleted. Accordingly they allowed the state to delete the five largest observations (LEAs) and five smallest from the data. Thus, they considered the 95th percentile as $3,591 (Maxwell) and the 5th percentile is $2829 (Gadsden). Therefore, the disparity equals 762/2829=26.94%, exceeding the 25 percent criteria for equalization status. Apparently, the Zuni School district also submitted another method for calculating the disparity, which is not as consistent with the statute or statistical principles.[18]

In the first federal appellate review of the administrative law decisions, which supported the DOE's and New Mexico's interpretation of the new statute as meaning the old formula, a panel of the 10th Circuit also affirmed its use by a vote of 2-1. The two judge majority noted that if a statute speaks clearly to the precise question, then courts must effectuate the unambiguous intent of Congress. If the statute, however, is silent or ambiguous, then the interpretation of a government agency should be sustained if it is based on a permissible construction of the statute.[19] The majority opinion reviewed the history of the original regulation and the justification the Department gave for its choice, which is described above.[20] It noted that while the prior regulations made clear that the percentiles were to be based on the student population, the majority felt that the new statute is not as explicit. The two-judge majority agreed with the DOE that statement **(i)** in the statute directing the Secretary to "disregard local educational agencies with per-pupil expenditures or revenues above the 95th percentile of such expenditures" is ambiguous as the statute does not contain a specific implementation of the disparity test.[21]

---

[18] That method, mentioned in the Zuni panel brief at 13 and the U.S. government's supplemental brief for the *en banc* hearing at 46, would exclude all LEAs whose average expenditures were greater than 95% of the per-pupil expenditure of the LEA with the highest per-pupil expenditure as well as the LEAs whose expenditures were less than 105% of the per-pupil expenditure of the LEA with the lowest expenditure. Since these cut-off points are not determined by the percentiles of the distribution of per-pupil expenditures in the LEAs, they are not consistent with the statute. Moreover, in our informal survey of statisticians described in Section 3b, *infra*, no respondent mentioned this method as a possible interpretation. In a situation where the LEAs with the smallest and largest average expenditures are large so that no data are deleted, a formula of this type would increase the allowable disparity in the ratio of their expenditures from 25% to 25 x (1.05/.95) = 27.63%.

[19] *Supra* note 6 at 1162, citing Yellow Transp. Inc. v. Michigan, 537 U.S. 36, 45 (2002) quoting Chevron, USA v. Natural Res. Def. Council. Inc. 467 U.S. 837, 842-43 (1984).

[20] See notes 14 thru 16 and accompanying text.

[21] *Supra* note 6 at 1166-67. The opinion did not explain why a statute should contain a specific implementation or example of the use of a formula in order to be unambiguous.

In a dissent, including an Appendix with a spreadsheet containing the underlying data on average per-pupil expenditures for each LEA along with the 5th and 95th percentiles obtained from the Excel program, Judge T. O'Brien asserted that the language in **(i)** is unambiguous.[22] He emphasizes that a percentile is a mathematical concept not admitting of multiple interpretations. He also notes that the purpose of the law was to promote control by local educational agencies with little or no Federal or State involvement[23] and that the entire basis of the case is the difference in language between the 1994 statute and the pre-existing regulations. The dissent questions whether the majority's characterization of the difference between them as an "unexplained and slight alteration" is correct, since courts should presume that Congress chooses its words carefully with knowledge of relevant administrative interpretations.[24] Judge O'Brien used Excel to make the calculation advocated by the plaintiffs, obtaining a disparity of 26.9%.[25] As this figure exceeds 25%, the distribution of school revenue in the New Mexico LEAs would not meet the "equalization" requirement and the plaintiffs' districts would keep all of their federal Impact Aid funds.

Since the panel opinion upheld the DOE decision awarding most of the federal funds to the state, the plaintiffs appealed to the 10th Circuit for an *en banc* hearing. In a *Per Curiam* ruling the twelve judges were evenly divided on the interpretation of the statute and the required disparity calculation.[26] This left the original DOE decision intact; however, the Supreme Court will now decide the matter.

Formal statistical descriptions of both interpretations of the statute are given in Appendix A. The DOE reads the statute as using a distribution that weights the average per-pupil expenditure in an LEA by the fraction of all pupils in the state being educated

---

[22] *Supra* note 6 at 1170. *See supra* note 11 for several dictionary definitions of a percentile. *See also* HARDEO SAHAI & ANWER KHURSHID (2002) A POCKET DICTIONARY OF STATISTICS 196 (McGraw Hill) defining percentiles by "The percentiles divide a data set into 100 equal parts, each of which contains 1% of total observations. More precisely, a 100pth percentile is a value such that 100p% of the items in the data set are less than or equal to its value and 100(1-p) % of the items are greater or equal to it".

[23] 393 F.3d 1158, 1172 (10th Cir. 2004).

[24] *Id.* at 1172. The dissent cites Lorillard v. Pons, 434 U.S. 575, 578 (1978) where it states that when Congress re-enacts a statute without change, it is presumed to be aware of the existing administrative interpretation, which implies that the interpretation remains in force.

[25] The program used by Judge O'Brien calculated the percentiles with a formula appropriate for data that are obtained from a random sample. In section 3 we will use the formula that follows the definition of a percentile in a population as we have data for all LEAs, i.e. data from a census of all LEAs in the state. It will be seen that the disparity is 32.4% when the more precise formula is used. This should not be interpreted as a criticism of Judge O'Brien as in Section 3, we will see his interpretation is the one many statisticians would make.

[26] 437 F. 3d 1289 (10th Cir. 2006). The opinion did not present any additional arguments for either interpretation of the statute.

in the LEA. Judge O'Brien and the plaintiffs interpret the wording of the statute to mean that all LEAs are considered equally in the calculation.

3. A Statistician's Reading of the Statute

In its brief to the panel the DOE discusses the statutory provision, 20 U.S.C. Sec. 7709(b)(2)(B)(i) reproduced above in Section 2a. It states that "While the intent of this provision is clear—namely, the Secretary is to exclude from the disparity calculation at the highest and lowest ends of the distribution of LEAs according to their per-pupil expenditures or revenues—the language of the statute does not readily translate into a mathematical formula".[27] This section argues that there is a straightforward translation of the provision; conceptually the one given by Judge O'Brien in his dissenting opinion.

At the beginning of most statistics textbooks, the concept of a population consisting of individual units, e.g. people, school or legislative districts or objects is introduced.[28] Each unit has one or more characteristics of interest and measurements of their value are obtained. If measurements on all units in the population are collected, one has taken a census. More often, in practice one takes a random sample of units of the population, obtains the value of the variable(s) of interest for them and makes inference on the distribution of the variable in the population from the sample.

To a statistician, the statute specifies a population of LEAs, the individual units, and the characteristic of interest is the average per-pupil expenditure (AE) of the LEA. By arranging the LEAs in increasing order of their expenditures, we obtain the distribution of the average per-pupil expenditures in the universe of LEAs in the state (see Appendix A for details). The disparity calculation is made by first taking difference, D, say, between the 95th percentile ($E_{.95}$) and the fifth percentile ($E_{.05}$) of the average per-pupil expenditures of the LEAs. The 95th and 5th percentiles are used in the calculation as the statute clearly states that LEAs with expenditures (or revenues) *above* the 95th percentile or *below* the 5th percentile should be disregarded. Then the disparity measure is the ratio of D to $E_{.05}$. Since one has data for the entire population, one determines the percentiles from the formula appropriate for data from the complete population.[29]

---

[27] Brief of Respondent, United States Department of Education in Zuni Public School District 89 v. U.S. Department of Education No. 01-9541 (March 7,2002) at 15.

[28] *See* BRUCE L. BOWERMAN, RICHARD T. O'CONNELL and MICHAEL L. HAND 3-4 (McGraw-Hill-Irwin(2d Ed 2001), JOSEPH L. GASTWIRTH, 1 STATISTICAL REASONING IN LAW and PUBLIC POLICY 45 (Academic Press, 1988) or JAMES T. McCLAVE and TERRY SINICH 6-9 (10th Ed, Pearson Prentice-Hall 2006)

[29] For distributions that may be either discrete or continuous the p^{th} quantile or $100*p^{th}$ percentile is defined as the smallest value, x, such that the cumulative distribution is at least p. The cumulative distribution function, F(x) or a variable, gives the fraction of the population who have values of the characteristic less than or equal to x. *See* ANTHONY C. DAVIDSON STATISTICAL MODELING at 22 (Cambridge Univ. Press, 2003) for the formal mathematical definitions of the cumulative distribution function and quantiles expressed in terms of the inverse of the cumulative distribution function. The Lorenz curve, which reports

a. Calculation of the Disparity Measures Using the Definition of a Quantile or Percentile of a Variable When Information on the Entire Population is Available

Before calculating the disparity measures for the 89 LEAs in the case, we consider a few smaller and simpler data sets. We first make the calculation described in the previous paragraph, which is a slightly more precise version of the one given by Judge O'Brien in his dissent. Then we calculate the DOE formula.

Example 1: Consider a state with only 10 LEAs, with average expenditures (AE) of 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000. Suppose each LEA educates the same number of pupils so that 10% of students in the state are taught in each LEA. To calculate the 5th percentile we need to find the smallest value, x, of the AE such that *at least* 5% of districts have an AE less than or equal to x. This is clearly the first district with an AE of $3100 as there are only 10 LEAs. To calculate the 95th percentile, we need to find the smallest value of x, for which at least 95% of the LEAs have a value less than or equal to x. This is $4000 because only 9 or 90% of the LEAs have an AE less than $3900. Notice that there are *no* LEAs with an AE less than the 5th percentile or greater than the 95th so the disparity measure is based on the ratio of the difference between the maximum (4000) and minimum (3100) to the minimum value, i.e. 900/3100=29.03% and the state's expenditures are not "equalized.

To calculate the 5th percentile using the DOE formula we need to find the smallest value, x, of the AE so that the totality of students in LEAs with an AE less than or equal to x, form *at least* 5% of all students in the state. Since 10% of the student population is in the first LEA with an AE of 3100, that value is the 5th percentile. Similarly the 95th percentile is the smallest AE, x, such that the totality of students in LEAs with an AE less than or equal to x, form *at least* 95% of the state's students. As each LEA educates 10% of the students, the 95th percentile is again $4000. The calculated disparity again is 29.03% and the state's expenditures are not "equalized". The reason both disparity calculations agree is because the numbers of students in each LEA are the same, so each LEA counts as one-tenth in both calculations.

Example 2: Now consider a state with 10 LEAs with the same expenditure data, ranging from $3100 to $4000 in $100 steps. The two LEAs in this state with an AE of $3100 and

---

the fraction of the total income individuals up to and including the 100pth percentile receive is a basic tool in the analysis of income inequality, is defined in terms of inverse cumulative distribution. Examples of its use are given in Joseph Gastwirth, *A General Definition of the Lorenz Curve*, 39 ECONOMETRICA 1037-1038 (2002), BARRY J. ARNOLD: MAJORIZATION and the LORENZ ORDER: A BRIEF INTRODUCTION at 31 (Springer, 1987) and John S. Chipman (1985) *The Theory and Measurement of Income Distribution*, 135 at 141-42, in ADVANCES IN ECONOMETRICS: ECONOMIC INEQUALITY: SURVEY, METHODS AND MEASUREMENT (R. L. BASMAN & GEORGE F. RHODES Jr., Eds.) For a discussion of various methods of calculating percentiles *see* Rob J. Hyndman and Yanan Fan, *Sample Quantiles in Statistical Packages*, 50 American Statistician, 361-365(1996) . While this article is concerned with estimating quantiles or percentiles from sample data, the first method it describes uses the inverse of the empiric distribution, which also is the distribution appropriate for the data under the interpretation of Judge O'Brien and the plaintiffs (see Appendix A).

$4000 respectively each educate only 2% of the state's students. The remaining LEAs each educate 12% of the state's pupils. The ordered data and relevant computations are given in Table 1.

Table 1: Data and Calculation of the Two Disparity Measures: Example 2

LEA	Avg. Exp.	Cumulative Fraction of LEAs	Fraction of Pupils	Cumulative Fraction of Pupils
1	**3100**	**.10**	.02	.02
2	*3200*	*.20*	*.12*	*.14 >.05*
3	3300	.30	.12	.26
4	3400	.40	.12	.38
5	3500	.50	.12	.50
6	3600	.60	.12	.62
7	3700	.70	.12	.74
8	3800	.80	.12	.86
9	*3900*	*.90*	*.12*	*.98 >.95*
10	**4000**	**1.00**	.02	1.00

The statistician's calculation described early in this section, finds that the 5th percentile is reached at the first LEA in Table 1 (in boldface) as it alone forms one-tenth of the LEAs. Thus, the 5th percentile is $3100. Similarly, the 95th is $4000 because we only reach a cumulative fraction of LEAs that is greater or equal to .95 at the tenth LEA (see col. 3 of Table 1). Thus, disparity for this state is (4000-3100)/3200 = 29.03% and the state would not be "equalized".

The 5th percentile under the DOE approach is $3200 because only 2% of the students are in the LEA with an AE of $3100 so we need to include the second LEA to reach a cumulative fraction of students (column 5 of Table 1) of at least .05. This occurs at the second LEA (italicized line in Table 1) as 14% of the state's pupils are in the first two LEAs. Due to the discreteness of the data, we cannot obtain a value x of expenditures less than $3200 containing *at least* five percent of the student population. Similarly, the 95th percentile is $3900 as 98% of the state's students belong to LEAs with an AE less than or equal to $3900 but only 86% are in LEAs with an AE less than or equal to $3800. Notice that the disparity using the DOE formula is ($3900-$3200)/$3200=21.88% and the state's expenditures are "equalized".

COMMENTS (1) the second example shows that a portion of the statement in the original notice of DOE, cited in the briefs in opposition to certiori of the Department and New Mexico is questionable. The relevant part is "In States with a small number of large districts, an exclusion based on percentage of school districts might exclude from the

measure of disparity a substantial percentage of the pupil population in those states."[30] In fact when there are a small number of LEAs the first calculation will *never* delete any observation as one must have at least 20 LEAs before the LEA with the maximum AE will not be the 95th percentile and 21 LEAs before the LEA with the minimum AE will not be the 5th percentile.[31] Only in the case when there are many small LEAs as well as a few large (containing more than 5% of pupils in the state) ones and one or two of the large ones are at the extremes of the distribution can deleting LEAs exclude more than ten percent of all students. On the other hand when a few small LEAs are at the lowest and highest ends of the expenditure range, some of their values will be deleted under the DOE formula. In particular, at the low end the population of the LEAs will be accumulated until it reaches 5% of the states total pupil population and the data for those LEAs will be deleted. The same procedure would be carried out at the upper end.
(2) The same section of the notice and briefs also states that the reason for the exclusion is to eliminate anomalous characteristics of the distribution of expenditures and that in states with many districts eliminating the extreme LEAs might exclude an insignificant fraction of the student population.[32] While this is correct, if one has many school districts but the two extreme ones happen to be large enough to contain at least 5% of the student population, then the DOE method will not eliminate any LEAs. Since provision **(ii)** of the new statute allows the Department to adjust the expenditure data to account for special situations such as geographical isolation or the education of pupils with "special needs" many "anomalous" aspects of the data should be removed in this process. Hence, one might question whether any of the adjusted data should be deleted.

(3) If Congress wanted to use the pupil population in the various districts to contribute to the disparity measure, one can ask why it did not use the average per-pupil expenditure in the state as the denominator of the measure. This average does weight the LEA averages by their student populations. This alternative measure would guarantee that the difference in expenditures between the 95th and 5th percentiles was not more than 25% of the average expenditure in the state. For most data sets, the 5th percentile is usually noticeably less than the average so one should also lower the 25% limit, say to about 10-15%, in this approach. Had Congress used this pupil weighted average in the denominator, then the interpretation of the DOE would be more reasonable as one should calculate the components of a disparity measure from the same statistical distribution.

---

[30] *See* 41 Fed. Reg. 26,323-24 (Jun. 25, 1976), brief for New Mexico at 23 and brief for the Department of Education at 13.

[31] Suppose one has n LEAs. Consider the case when n = 20. Then 19/20=.95, implying that after the 20 districts are in rank order of their expenditures, the first 19 districts contain 95% of the LEAs so the AE of the 19th district, having the second highest AE, is the 95th percentile. If one has n districts, the first n-1 of them contain the fraction (1-1/n) of the LEAs. Since 1/n decreases as n increases, for any n less than 20, 1/n is greater than .05 so the fraction of all LEAs formed by the first (n-1) is less than .95 and the LEA with the highest AE is the first one for which we have *at least* 95% of all LEAs. Similar reasoning shows that one needs to have at least 21 LEAs before the district with the second lowest AE, rather than the district with the lowest AE is the 5th percentile. This slight asymmetry results from the fact that with 20 LEAs, 1/20 is exactly .05 so the first district determines the 5th percentile.

[32] *Supra* note 30.

Next we calculate the disparity measures from the data in the case; details are given in Appendix B. After arranging the LEAs in increasing order of their AE, the fifth one is the 5^{th} percentile as 4/89=.0449 while 5/89=.0562, i.e. after including the fifth LEA we have at least 5% of the LEAs. This LEA was Lake Arthur, with an AE of $2787. Similarly, the 95^{th} percentile is the 85^{th} LEA, Hondo Valley with an AE of $3690. The disparity measure is (3690-2787)/2787= 32.4%, which clearly exceeds the 25% guideline. Notice that our calculation yields a greater disparity than either the plaintiffs or Judge O'Brien obtained as they apparently used a computer program designed for calculating percentiles from a random sample.[33]

Following the formula specified by the DOE one finds that we reach at least 5% of the student population at the Hobbs LEA with an AE of $2848. Similarly, we reach 95% of the students at Ruidoso, which is the 74^{th} LEA, i.e. 15 LEAs are excluded. As the AE in Ruidoso is $3278, the disparity is 15.098%. Since one reached 94.95% of the student population of the state at the 73^{rd} LEA, Tatum, one might have used its AE ($3266) if one rounded the original data. Then the disparity would be 14.68%. While both these disparity measures are slightly larger than the value (14.43%) submitted by New Mexico and accepted by the DOE, this does not affect the classification of the state as "equalized" under the DOE formula. In the process of calculating the weights from the data in the Appendix, I obtained a total of 300776.5 students in the state, which differs from the 317,777 given in the submission by the state.[34] This might contribute to the slight discrepancies in the calculations.

b. A Brief Summary of Informal Surveys of Statisticians and Seminar Attendees

Since the interpretation of a statute depends on either the way the word or expression in question is understood in its common usage or as a term of art, how professional statisticians would make the specified calculation should be a consideration in determining whether the statute is reasonably clear.[35] As I did not have the resources to conduct a true random sample, I decided to follow the approach Judge Weinstein took in *U.S. v. Fatico* [36] who asked his colleagues on the bench to provide him estimates of the

[33] *See* Hyndman and Fan, *supra n.* 29, for an extensive discussion of various methods. Fortunately, the calculation of a percentile when data are available for all members of the statistical universe, i.e. a census has been taken is more straightforward.

[34] *See* Appendix A No.01-9541, filed Feb. 23, 2006 at 213a. Both Excel and R gave the figure I report. Although one would expect such an error to be caught during the administrative and legal reviews, the change has a very slight effect on the calculated disparity.

[35] I use the term reasonably clear rather than ambiguous since the classification of a statement as ambiguous in this context is a legal rather than statistical issue. My colleagues in the Humanities specializing in hermeneutics have told me that virtually all sentences inherently have some degree uncertainty or ambiguity.

[36] U.S. v. Fatico 458 F. Supp. 388, 410 (E.D. N.Y. 1978). The data and other statistical aspects of the case are described in JOSEPH L. GASTWIRTH, 2 STATISTICAL REASONING IN LAW AND PUBLIC POLICY, 700-704.(Academic Press, 1988).

probabilities they associate with various standards of proof. Because the sample is relatively small and respondents were not randomly selected from a list of statisticians and quantitative researchers, the results should not be interpreted as an estimate of the percentage of statisticians who would read the statute according to the plaintiffs' or the defendant's interpretation. Rather it provides us some insight into the way statisticians react to the statute and indicates that the description given at the beginning of the section is quite standard.

My first step was to send a copy of the statute to about 10 statisticians as a "pilot study". Almost all interpreted the statute as Judge O'Brien did; however, several noted that the statute was somewhat unclear. In particular, some concern was expressed about the use of actual expenditure data *or* revenue available data as a state could use both sets of data and then present the most favorable result to the Department of Education. Also, one person questioned the use of data from two years before the one in question was used. Since those issues were tangential to the definition of the percentiles that are essential to the calculation of the disparity, I wrote two simpler versions of the statute, which were sent to colleagues with whom I had professional correspondence during the period.

The first version (A) asked respondents to describe the data they would ask for and how they would make the calculation. It also asked whether any other way of calculating the disparity came to mind. The second version (B) indicated that the available data for each LEA would report *both* the number of pupils and the average expenditure per-pupil. The second version was designed to see whether having the population data would stimulate the respondents to give the interpretation made by the Department of Education.[37] The two versions are given in Appendix C.

Shortly after the Court granted certiori, I stopped asking collecting answers as respondents might hear about the case. As of Oct. 4, 2006, the results show that virtually all professional statisticians agreed with the plaintiffs and Judge T. O'Brien and his colleagues. Of sixteen individuals answering version B, which mentioned the population data, fourteen gave with the plaintiffs' interpretation, i.e. the statute says consider the data for each LEA without weighting by population. The other two indicated that while a literal reading of the statute specifies the unweighted calculation, it would be fairer if the LEAs were weighted by their population, i.e. they would prefer the measure created by the Department of Education. One of them felt that the law as written did not reflect the intent of Congress as it mixed per-capita and total dollar expenditures of the LEAs so he would present several alternative calculations, including both proposals. Three of these fifteen respondents questioned the clarity of the calculation *after* the percentiles were

[37] The purpose of using question B in the survey was to determine whether statistician reading the statute would conclude it meant the DOE formula even when a "hint" that population data were available was given. Thus, it differed from the ordinary use of survey evidence in Lanham Act cases as questions having a tendency to "lead" are improper as noted in Coors Brewing Co. v. Anheuser Bush Co's. , 802 F. Supp. 965, 972 (S.D.N.Y.1992). *See* Shari S. Diamond (2000*) Reference Guide on Survey Research* in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, 229-272 (Federal Judicial Center, Washington D.C., 2000 2nd Ed.) and 2 GASTWIRTH, *supra* note 36 at 467-541, for a more detailed discussion of the issues in developing a proper sample and survey instrument for use in legal cases.

determined. Two made a calculation similar to that of Judge O'Brien and the plaintiffs while one thought that the difference between the 95th and 5th percentiles should be divided by the 95th instead of the 5th. One person questioned the propriety of removing the "outliers" since one is concerned with equal education for all students but realized that there might be special circumstances.

All eleven respondents to version A interpreted the law to mean the plaintiff's calculation of the percentiles from the per-pupil expenditures of the LEAs and made a disparity calculation similar to that of Judge O'Brien. Three respondents also mentioned the same question about the choice of denominator at the second stage noted by three respondents to version B. One was unsure about the use of both expenditure and available revenue data and thought that an LEA above (below) the 95th (5th) percentile of the LEAs ordered by either measure would be excluded. Five of the twelve respondents to Version A questioned whether the data should be trimmed. One of them argued that the law should require a "floor" or minimum for all pupils. Another noted said that the use of symmetric percentiles for the trimming needed validation. Several wondered why there was a problem in the first place.

Among all 27 respondents only two suggested that the Department's approach came to mind as a possible alternative calculation in question 2 of Version B or question 3 of Version A. One other respondent to version B mentioned the Department's approach occurred to him after he gave the plaintiffs' interpretation but only because the population data was given in the question. Thus, a substantial majority of the small set of statisticians in this informal survey understood the statute to mean the unweighted calculation, which is consistent with the interpretation of the plaintiffs and Judge O'Brien. Our survey suggests that if there is an ambiguity in the law, it concerned how to proceed after calculating the difference, D, between the 95th and 5th percentiles. Six were not sure whether the ratio of D to the 5th or 95th percentile should be taken in order to check whether it was less than .25 or 25%. Five decided on the 5th percentile but one used the 95th.

On Sept. 18, 2006 I gave a seminar at Columbia University and asked the audience of about 35 graduate students and faculty to interpret Version B. One individual volunteered the interpretation advocated by the plaintiffs and Judge O'Brien. I asked the audience if they agreed and about two-thirds of the people raised hands.[38] One person asked whether the population counts should be considered but did not suggest the calculation specified by the Department's regulation. Apparently he felt that the measure in the law should have incorporated the varying pupil populations of the LEAs but agreed with the others that that the law did not say this.

In sum, while some details of the calculation described in the statute were not absolutely clear, most statistical readers were not confused about which type of percentiles should be calculated when the statute is read literally. Even though sub-grouping the LEAs by grade groups is *not* mentioned in the new statute, the DOE also

---

[38] Since many attendees were students, it is not surprising that some would be hesitant to provide an answer in public. None of them, however, suggested the DOE interpretation.

retained an alternative formula that weights the disparity of LEAs of these different types to obtain a weighted average disparity for the state. Although this alternative method was not applied in the *Zuni School District 89* case, the next sub-section describes several statistical deficiencies with that calculation.

3c. Statistical Problems with the Department of Education's Alternative Computation

Both the original and current DOE regulations also allow a state to request separate disparity computations for different groups of LEAs in the state that have similar grade levels of instruction and obtains an overall disparity index by weighting the disparity for each group in proportion to their share of the population of students.[39] While the statistical soundness of this formula was not an issue in the case, several statistical issues concerning it deserve discussion.

First, it allows a state's educational funding program to be classified as "equal" even though if there is very wide disparity in *per pupil* expenditures in the LEAs of different types. To see this consider a state with 300,000 pupils of which 100,000 are in systems with grades 1-6. The LEAs in this group spend between $1,000 and $1,200 per pupil. Another 100,000 students are in systems with grades 7-12 that spend between $ 3,000 and $3600 per pupil. The remaining 100,000 students are in systems with grades 1-12 that spend between $2,000 and $2,400. For each type of school the disparity measure cannot exceed .20, so that the weighted disparity cannot be greater than .20. If one considers all the LEAs as one group, however, the fifth percentile of the population of schools ranked by their LEA revenue or expenditure must lie in the first group and is therefore *less* than $1200. Similarly, the 95th percentile must lie in the second group and is therefore *at least* $3,000. Thus, the disparity index of all districts is *at least* 150% (1800/1200). Such a result is not consistent with the 25% guideline in the current statute, which is expressed in per-pupil expenditures, without any reference to the grade levels of the students or LEAs.

Secondly, if the disparity measure was based on the ratio of the difference between the maximum and minimum to the minimum, i.e. the upper and lower five percent of the data was not omitted from the calculation, then mathematically the weighted disparity *will always* be less or equal to the disparity calculated from the totality of the data, i.e. the per-pupil expenditures of all LEAs.[40] Thus, states with LEAs

---

[39] The regulation, 34 CFR Sec. 222 Appendix to Subpart K, illustrates this computation. Assume 80,000 pupils are in LEAs operating schools with grades 1 thru 6 and their disparity measure is 18.0%, 100,000 pupils are in LEAs with grades 7-12 with a disparity measure of 22.0% while 20,000 students are in systems with grades 1-12 with a disparity of 35.0%. As there are 200,000 pupils in the state, the weighted disparity measure equals
.4 x 18.0% +.5 x 22.0% + .1 x 35% = 7.2%+11.0% + 3.5%=21.70%.

[40] This results from the fact that the minimum revenue in all LEAs is less than or equal to the minimum of revenue in each of the subgroups (three in the example in the regulation) while the maximum of the revenues in all the LEAs is greater than or equal to the maximum revenue in each of the subgroups. Hence the disparity (max-min)/min calculated from the data on all LEAs will be greater than or equal to the disparity in each subgroup and consequently *less* than the weighted average of the subgroup disparities.

organized by grade level are more likely to be classified as having an "equalization" program than other states even though their overall disparity in per-pupil funding is the same or even greater.[41]

Finally, the weighting formula in the DOE regulations allows a very large disparity in one category of schools to be masked by small disparities in the other categories. In the hypothetical from the regulations, reproduced in footnote 39, notice that the disparity in the LEAs with grades 1-12 could be as large as 68% before the overall disparity would equal the 25% threshold.[42] Given the purpose of the federal government's funding of education in impacted areas, did Congress intend that a statute with the objective of ensuring that a state has an equalization program and does not even mention grade levels allow one large category of schools (or districts) to have a disparity greater than 50%?

4. Other Statistical Aspects Concerning the Choice of Data and Measure of Equal Per-pupil Funding

This section discusses statistical issues that Congress and the DOE might consider when they develop similar formulas to ensure equal education opportunity in the future. The first section explains why one should obtain data on expenditures per school rather than for each district in assessing educational inequality. It will be seen that the use of data on the average per-pupil expenditure in an LEA does not reflect any of the variation between schools in that LEA. This means that substantial inequality in the educational funds available to the children in the state may not be detected when disparity measures are calculated from data on LEA averages. Section 4b criticizes the approach to trimming the data used to calculate both measures at issue in the case. In particular, deleting the bottom five percent of the data, in either approach at issue in the case, is questionable from both a statistical view[43] and from a policy view if one desires to ensure an adequate

---

[41] Because the calculation trims the data *below* and *above* the fifth and ninety-fifth percentiles in each of the grade-groups of LEAs, it is no longer a mathematical certainty that the weighted summary measure of disparity will be less than the disparity calculated on the revenue data for all of the LEAs. For example, if there are 100 LEAs in the state. One group consists of 4 LEAs, two with average expenditures of \$2000 and two of \$3000, a second group consists of 4 LEAs, two with average expenditures of \$6000 and two of \$9000 and the third group of 92 LEAs have average expenditures ranging from \$3600 to \$4499. Of these 92 LEAs, assume the lowest five have an AE of \$3600, the sixth an AE of \$3601 and the largest five have an AE of \$4499 an the next an AE of \$4498. The 5th percentile of all 100 LEAS is \$3600 and the 95th is \$4499, so the disparity measure is 24.91% and just satisfies the 25% criteria for equalization. As the two smaller groups do not satisfy the 25% criteria, the weighted disparity measure will also fail to satisfy the criteria (.92x24.91 +.08x 33.33 = 22.917 + 2.666= 25.58%). Nonetheless, the weighted disparity will tend to be less than the overall one in most realistic data sets.

[42] This figure is obtained by determining the maximum value the disparity, d, in the last group can have so that the final total is 25%. Thus .1xd=25%-16.2% or d=68%.

[43] *See* ALAN STUART 7 J. KEITH ORD, KENDALL'S ADVANCED THEORY OF STATISTICS 59, (5Th Ed. Oxford Univ. Press 1987) at 59 (noting that robust measures of spread should be used with care as when the extreme values are genuine, using one of the robust measures may systematically understate the variability in the population). Several important robust measures of location or the center typically delete

education for all pupils.[44] In the third part of the section we suggest that the coefficient of dispersion, a standard measure of relative inequality used to evaluate the uniformity or fairness of tax assessments, or a weighted version of it be considered.[45] During the process of calculating these measures it will be seen that the LEAs with the largest contributions to the disparity measure can be identified. Then their individual situations can be examined to see if their data should be adjusted for a “special situation” as specified in 20 U.S.C. Sec. 7709(b)(2)(B)**(ii)** of the 1994 statute.

4a. Limitations of the District Level Data Used in the Assessment of the Equality of Expenditures per Pupil.

Both the measure originally proposed by the DOE and the one in the new statute focus on the distribution of expenditures per-pupil; however, they both refer to the average per-pupil expenditure of an LEA. Assume for a moment that one could determine the cost of educating each pupil, individually. Then, one could then calculate the average cost per pupil in each classroom. Then weighting each class in proportion to the number of students in it, one obtains the average cost per school. Finally, one obtains the data used in the calculations in the case by weighting the expenditures in each school in proportion to the size of the student body. From a statistical viewpoint each averaging process *reduces* the variation in the data, i.e. the variance of the averages of several subsets of a data set is *less than* the variance in the total data set.[46]

These considerations imply that a modest degree of variability in the expenditures per LEA, weighted or not, can mask a much larger degree of variability in per-school inequality, much less per-pupil inequality. For example, a district with 20 schools of similar size could allocate \$2,200 to ten schools and \$3,800 to the other ten. Both the mean and median equal \$3,000, so this LEA would lie in the middle of the data reported in the Appendix (median=\$3059, weighted median=\$2992). Under either calculation at

---

some data above or below an extreme percentile, *Id.* at 49; however, accurate data in the extremes reflects the variability of the characteristic in the population being studied.

[44] The formal adoption of the 1976 notice reproduced in the Appendix to the plaintiffs’ petition for certiori, A No.01-9541, filed Feb. 23, 2006 at 149a states that the following should be an objective of a state equalization program: “Such program is designed to ensure the provision of financially adequate educational programs and supportive services for every pupil in the State who is enrolled in public schools”.

[45] This measure is also used by CARROLL & PARK, *supra* note 15 where it is called the relative deviation. We follow the terminology of the U.S. Bureau of the Census where the coefficient of dispersion is used to assess the fairness of property tax assessments.

[46] This type of decomposition of the overall variance in a data set is the basis of the Analysis of Variance method for comparing the average values of several data sets to assess whether or not they are the same. The total variance is the sum of the variance between the groups and the variance within each group. Ignoring the per-pupil expenditures within each of the 89 LEAs implies that the true variation in per-pupil expenditures is substantially under-estimated. Thus, either of disparity measures calculated from LEA-average per-pupil expenditures usually under-estimates the true statewide-inequality in per-pupil expenditures.

issue in *Zuni School District 89*, however, the disparity index within this LEA equals 1600/ 2200 =72.3%, greatly exceeding either criteria. Thus, it is clearly preferable to use expenditure or revenue per-pupil data for each school in the state as the calculated disparity would now include the variation in per-pupil expenditures within each of the LEAs as well as the variation in spending per-pupil between the LEAs. Although the 1976 notice stated that "school-by-school identification of per pupil expenditures would provide a more precise picture within a state" it was felt that this would be a "tremendous" administrative burden for the state.[47] The importance and relevance of data on expenditures by schools is demonstrated by its use in cases concerning school segregation or discrimination.[48] The fact that the Department did not require this more accurate data also undercuts the argument that their calculation of the disparity measure was an approximation to the true disparity in per-pupil expenditures in the state.

4b. Statistical issues concerning both measures of disparity

Before suggesting alternative measures of disparity, it is useful to examine the data in the Appendix. Notice that a large proportion of the school population of the state is in few districts, e.g. a third of the student population resides in Albuquerque or Las Cruces. Suppose one of them had the highest revenue and the other the lowest, then calculating the 5th and 95th percentiles using their data in the government's disparity index might have raised a question. Formally, they would have been the 5th and 95th percentiles and but removing them, as the original notice seems to suggest,[49] would effectively delete over 33% of the data rather than 10%. As discussed in Section 3a, those percentiles should not be deleted and the DOE's disparity measure would be calculated from their average expenditures. Under the calculation of Judge O'Brien, however, the data for these two LEAs would be deleted as commented on in Sec. 3a. Neither measure seems appropriate when the LEAs with extreme per-pupil expenditures are large.

---

[47] The notice, fn. 2 at Sec. 115.62 as reprinted in Appendix B of plaintiff's petition for certiorari, at p. 131a. Given modern computerized financial systems, the administrative burden should be much smaller today.

[48] In Gebhart v. Belton*,* 91 A. 2d 137, 150 (S. Ct. Del. 1952), a companion case to Brown v. Board of Education, the court among the inequalities noted between two schools, one black and one white, was the fact that the black school received $ 137.22 per pupil while the white school received $ 178.13. For these two schools the disparity measure is the ratio of the difference, $40.91 to the minimum ($137.22) or 29.8%, which exceeds the 25% criteria. More recently, the Court's opinion in the school segregation case Freeman et al. v. Pitts et al., 503 U.S. 467, 483-484 (1992) noted that the fact that per-pupil expenditures in primarily white schools exceeded those in black schools is an indicator that the schools still were racially identifiable. This point is also made by Justice Souter, *Id*. at 508. Moreover, in a concurring opinion at 511, Justice Blackmun, joined by Stevens and O'Connor noted that expenditures per-pupil were under the control of the school administration. The analysis considered in Freeman v. Pitts, reported in an appellate opinion, 755 F.2d 1423, 1442, compared the average expenditures of majority white schools of $ 2833 to the corresponding figure of $ 2492 in majority black schools. Schools where the minority fraction increased in recent years had an average expenditure of $ 2540. The disparity between the white and black schools of 341/2492= 13.68% sufficed to support a claim of discrimination. It is smaller than either calculation at issue in the Zuni School District case, which suggests that the "equalization" criterion is not especially stringent.

[49] *Supra* n. 30.

The choice of the 5th and 95th percentiles as the values to trim the data when estimating a measure of spread or disparity, especially for skewed data, is not supported by the statistical literature. In the 1960's and 1970's substantial statistical research focused on the deletion of outliers or unusual observations when one estimated the *center* or *typical* value of the data. The average of the observations between the 5th and 95th percent, known as the 5%-symmetrically-trimmed mean, has good statistical properties when the underlying data is roughly bell-shaped or normal with a modest fraction of data from a heavier-tailed distribution.[50] The revenue data in the Appendix is far from normal and is not symmetric about a central value.[51] Hence, the choice of symmetric percentiles even for eliminating "unusual" observations for estimating the center of the revenue data is not appropriate. For such skewed data, most statisticians would use the median to estimate the *central* or *typical* value rather than the mean.

To estimate the scale parameter, which is related to the variability or spread of the data, of a skewed distribution, statisticians often only trim a small fraction of data in the upper-tail.[52] Keeping the LEAs with low amounts of available revenue or expenditure per-pupil in the calculation of disparity would also be consistent with a state having an equalization program. Since part **(ii)** of the new statute *allows* the state to adjust the expenditure or revenue data to account for appropriate special situations, the further use of trimming to reduce the potential impact of an unusually large or small observation

---

[50] *See* HERBERT A. DAVID & H.N. NAGARAJA, ORDER STATISTICS 211-12 (3d Ed, John Wiley, 2003) for a brief summary of and references to the literature. For trimmed means, in particular, *Id.* at 217 they cite John W. Tukey, A *Survey of Sampling from contaminated distributions*. In: INGRAM OLKIN et al. (Eds.) CONTRIBUTIONS TO STATISTICS,448-485 (Stanford Univ. Press , 1960), Edwin L. Crow & M.M. Siddiqui, *Robust Estimation of Location*, 62 J. AMER. STATIST. ASSOC. 353-389 (1967) and Joseph L. Gastwirth & Martin L. Cohen, *Small Sample Behavior of Robust Linear Estimators of Location* 65 J. AMER. STATIST. ASSOC. 946-973 (1970). *See also,* PETER SPRENT, DATA DRIVEN STATISTICAL METHODS 68-75 (Chapman-Hall, 1998) for a nice introduction to the basic concepts used to develop robust estimators.

[51] Although tests of statistical significance are primarily designed to check the assumptions underlying an analysis of data obtained from a random sample of the relevant universe, it is often helpful to use them to examine the distribution of the characteristic of interest on data from a census. The p-value becomes more of an indicator of the fit of the data to an assumed model instead of a pure test of significance. The commonly used Shapiro-Wilk test of normality is .514, much below its expected value near 1.0. The p-value is < .0001; strongly indicating that if the data came from a random sample it would be far from normally distributed. Three tests of symmetry also reject that hypothesis with very low p-values. References for the three tests are: Cabilio, P. and Masaro, J., *A simple test for symmetry about an unknown median*, 24 CANADIAN J. OF STATIST., 349-361 (1996), A. Mira *Distribution-free tests of symmetry based on Bonferonni's measure*. 26 J. OF APPLIED STATIST. 959-972 (1996) and Miao, W., Gel, Y. and Gastwirth, J.L. (2006). *A new non-parametric test of* symmetry (to appear in RANDOM, WALK, SEQUENTIAL ANALYSIS and RELATED TOPICS – A FESTSCHRIF IN HONOR OF YUAN-SHIH-CHOW (ed. A. Hsiung, C.H. Zhang and Z. Ying).

[52] See ROBERT G. STAUDTE & SIMON J. SHEATHER, ROBUST ESTIMATION AND TESTING 68-70 (John Wiley, 1990). For *symmetric* distributions such as the normal curve, the authors show, at p. 123-126, that a symmetrically trimmed estimate of the standard deviation is an appropriate robust estimator when one is concerned with possible contamination by a heavier tailed *symmetric* distribution. For recent results concerning robust estimation from skewed data, *see* Brenton R. Clarke, David K. Gamble and Tadeusz Bednarski, *A Note on Robustness of the β-Trimmed Mean*, 42 AUSTRALIAN & NEW ZEALAND J. STATIST. 113-117 (2000).

seems unnecessary. After those adjustments are made, the per-pupil expenditures should be reasonably equal if the state had an "equalization" program in effect. Since a true "equalization" program should be concerned with the funds available to educate children in the poorest districts, the deletion of the lowest five percent of the adjusted revenue or expenditure data appears inconsistent with the stated purpose of the law.[53]

The numerators of both disparity measures are based on the difference between suitably defined 95th and 5th percentiles which is a measure of spread. Thus, it is sensible to analyze the data by well established statistical measures of relative inequality. Such measures, e.g. the Gini index or coefficient of dispersion are ratios of a measure of spread to a central value. Thus they gauge how large a measure of the spread of the data (revenue per-pupil) is relative to a typical value (the mean for the Gini index or median for the coefficient of dispersion). Since it is less sensitive to "outliers" we discuss the coefficient of dispersion in the next sub-section.

4c. The Coefficient of Dispersion and its Weighted Version

Using the median, M, to estimate the center of the data, a natural measure of the spread or variation of the data is the average distance of all the observations from M; called the mean deviation from the median (MDM). The ratio of this measure to the median is a measure of relative disparity as it compares the spread of the observations to their typical value. This measure, the coefficient of dispersion (CD), has been used to evaluate the fairness of tax assessments and its statistical properties are well known. One can also define a weighted version (WCD) corresponding to the approach of the DOE. It is helpful to define these measures precisely and apply them to the LEA revenue data. Denoting a set of n observations by $x_1,\ldots,x_n$ and their median by M, the CD is

$$CD = \frac{(1/n)\sum_{i=1}^{n} | x_i - M |}{M}.$$

The CD is just the average absolute difference between the observations and their median. Technically, it can also be defined as the expected absolute deviation of a random variable with distribution $F_n(t)$, defined in Appendix A, from its median.

The weighted CD is the CD of the distribution, $F_D(t)$, defined in Appendix A, which weights each LEA in proportion to their fraction (w_i) of students in the state. Denoting the per-pupil revenue or expenditure of the i^{th} LEA by s_i, the weighted CD is

[53] The original regulation, fn. 23 at Sec. 115.62 or Appendix E of plaintiff's petition for certiori No.01-9541, filed Feb. 23, 2006 at 149a states that a program of state aid shall not be determined to be a program designed to equalize expenditures among LEAs unless "Such a program is designed to ensure the provision of financially adequate educational programs and supportive services for every pupil in the State who is enrolled in public schools".

$$WCD = \frac{\sum_{i=1}^{n} w_i \mid s_i - M_w \mid}{M_w}$$, where M_w is the median of the weighted distribution.

In the case, this median, M_w is the 50th percentile of revenue according to the calculation described by the DOE, i.e. the median of the distribution $F_D(t)$. To calculate it, one orders the data on the LEAs by revenue and adds their weights w_i in succession until one reaches one-half. If that LEA is the k^{th} one, s_k is the median of the weighted distribution. The WCD is the expected or average absolute deviation of a random variable with distribution $F_D(t)$, defined in Appendix A, from its median, M_w.

The ordinary CD which gives equal weight to each LEA is .0861 (263.48/3059) implying that the relative spread of the revenues amongst the LEAs of the state is slightly below 10% of the median value \$3059 (San Jon LEA).[54] In contrast, the weighted median is 2992 (Roswell LEA) and the WCD = .0422 (126.35 /2992). These two measures of relative disparity differ by a factor of two, which is near the ratio of the two disparity measures in the case. The smaller value, indicating more equality, is obtained when the population-weighted distribution is used. In the process of calculating the weights from the data in Appendix B, I obtained a total of 300776.5 students in the state, which differs from the 317,777 given in the submission by the state.[55]

One can also assess the relative contribution of each LEA to the WCD by looking at the product of its weight (share of students) and its distance from the center or weighted median. There are two districts with quite large contributions to the numerator of the WCD. The first is Los Alamos having weight .0117 but a large revenue (\$5611) that exceeds the weighted median by \$2619. The second is Albuquerque, the largest city in the state, which has a large weight (.2786) and per-pupil revenue (\$3071) near both the weighted median (\$2992) and ordinary median (\$3059).

In the measure adopted by DOE, the Los Alamos LEA is dropped from the calculation as the 95th percentile of the population weighted revenue data (see Appendix B) is \$3259. Rather than deleting this and other observations that have a very large impact on the calculated WCD, it might be fairer to examine them to see whether they have a special situation as described in part 20 U.S.C. Sec. 7709(b)(2)(B)**(ii)** of the new

[54] The state might argue that this value of the CD for the average per-pupil expenditure data is small in comparison to typical values seen in other applications to justify that it has an "equalization program". However, this results from the fact that the average expenditures in each LEA are being used rather than data on each school. For example, in assessing the uniformity of tax assessments CDs of .20 or even .30 are common. In that application, however, the measure is calculated on the assessment to sales ratio of *individual* homes not averages of groups of houses.

[55] *See* petitioners Appendix No.01-9541, filed Feb. 23, 2006 at 213a. Both Excel and R gave the figure I report. Although one would expect such an error to be caught during the administrative and legal reviews, the change has a slight effect on the calculated disparity. The LEA at the 5th percentile of student population again is Hobbs but the LEA at the 95th percentile is Ruidoso, with a revenue per-pupil of \$3278. This only increases the disparity to 430/2848=15.1% so the state remains "equalized" under the DOE formula.

statute. A famous government research laboratory, the successor to the research center that developed the atomic bomb in World War II is located in Los Alamos and is its largest employer. Since it employees many scientists and professionals, the citizens of Los Alamos are likely to be more interested in education and perhaps more willing to pay higher taxes to support their local schools. Given its history, Los Alamos might still be regarded as isolated and needs to pay more to attract teachers. If this is the case, then according to section 7709(b) (2) (B) **(ii)** the value for the Los Alamos LEA might be adjusted appropriately before a disparity measure is calculated. If the Los Alamos LEA should not be considered as a special situation, adjusting or deleting its expenditure makes it more difficult to find that a state's educational revenues are *not* distributed equally. In other contexts where one assesses inequality or discrimination, the outliers or unusual observations can be strong indicators of unequal treatment.[56] Thus, one should be careful about deleting observations unless one truly believes they are unusual or "outliers".

Comment: Given the importance of the Los Alamos Laboratory to defense-related research, why would Congress first allocate funds for the education of the children of its employees and then create a formula that allows most of those funds to be used by the state? The employees of the Laboratory probably desire very good schools for their children, so government support of the school system should assist in the recruitment of qualified researchers.

Technical Comment: When there are an odd number of districts, the one with the median revenue and corresponding weight w will not contribute to the calculation of the CD or WCD as its distance from the median is zero, one might prefer to use (n-1) in the denominator of the CD. Similarly, one could renormalize the WCD by multiplying it by 1/ (1-w) where w is now the weight of the LEA which has the weighted median revenue.

5. Discussion

There has been considerable discussion in the legal literature about the weight judges should give to the interpretation of the relevant governmental agency when they interpret the statute in question.[57] A recent case focusing on whether the operation of a

---

[56] In Craik v. Minnesota State Univ. Board, 731 F.2d 465 (8th Cir. 1984) the defendant criticized a regression analysis which included data on male faculty members who had served as administrators and retained their 12 month salary when they returned to the faculty (paid on a 9 month salary). Naturally, they were identified as outliers or unusual observations, however, no female faculty member had been appointed to an administrative position so these outliers actually reflected the discrimination at issue in the case.

[57] *See* ANTONIN SCALIA, A MATTER OF INERPRETATION: FEDERAL COURTS and the LAW (Princeton Univ. Press, 1997), RICHARD A. POSNER, PRAGMATISM and DEMOCRACY 67-72 (Harvard Univ. Press, 2003), STEPHEN J. BREYER, ACTIVE LIBERTY: INTERPRETING OUR DEMOCRATIC CONSTITUTION. 103-108 (Knopf, 2005) at 103-108 and TONY HONORE, ABOUT LAW (Oxford Univ. Press, 1995) at 87-95. Honore, *Id.* at 91 discusses how a fairly simple law allowing a policeman to stop and administer a blood test to a driver requires interpretation since after the car has stopped; the individual is no longer "driving" the car. Nonetheless, if the test indicates a high level of blood alcohol the individual can be charged with driving while drunk. Thus, the reviewing court needs to consider the meaning of a particular section of a law or regulation in context.

dam may result in a discharge into navigable waters illustrates the process used to assess whether a statute is ambiguous.[58] If the dam may make a discharge, in order to obtain a federal license, the owner of the dam must obtain state certification that its water protection laws will not be violated.[59] Consequently, the Court had to decide what the term discharge meant in the context of the law. The opinion noted that the word "discharge" was not defined in the statute and is not a term of art having a different meaning in the special field relevant to the case; hence it should be construed in accordance with its ordinary or natural meaning.[60] The Court concluded that a dam has the potential for a discharge so state approval is required.

Since the new law specifies the disparity formula, rather than preserving the wording of the Department's previous formula, the weight that should be given to the Department's unchanged interpretation of the disparity calculation is a legal issue.[61] If the Court decides that the definition of a percentile is considered a term of art, then the interpretation given by professional statisticians and statistical textbooks and treatises should be seriously considered.[62] Several other statistical issues discussed here may be relevant to evaluating the expertise and professional care the agency gave in developing its interpretation of the new statute. Our informal survey and discussion at a seminar at Columbia University, statisticians do not have much difficulty interpreting the pertinent part of the statutory formula.

In our review of the case and regulatory background it became evident that statistical issues may not have played a major role in the development of either formula. In the case neither party apparently called a statistician as an expert witness.[63] In addition

---

[58] S.D. Warren Company v. Maine Board of Environmental Protection et al., 547 U.S.----(2006).

[59] *Id.* at 1 (citing Sec. 401 of the Clean Water Act).

[60] *Id.* at 4 citing FDIC v. Meyer, 501 U.S. 471, 476 (1994).

[61] In United States v. Price 361 U.S. 304, 313 (1960), the Court noted that the views of a subsequent Congress form a hazardous basis for inferring the intent of an earlier one. Since the Congress that passed the new statute knew the previous formula created by the Department when it decided to write a somewhat different one, it is reasonable to infer that Congress desired to modify the old formula. Thus, the fact that the Department did not change its interpretation in response to the new law might be relevant.

[62] In Evans v. Utah, 536 U.S.452 (2002) at 467, the Court interpreted the statistical term "sampling" as a term of art since the relevant statute, 13 U.S.C. Sec 195 (1976) placed quotation marks around it. *See* Jason Wejnert *Utah v. Evans and Census Apportionment* 43 JURIMETRICS J. 441-451 (2003) for a summary of the majority and dissenting opinions in the case. Since the statute in question in the *Zuni School District* case did not place quotes around the term percentile, it will be interesting to see if the Court considers it a technical term or a term of art as well.

[63] The Department of Education did submit a declaration from Richard G. Salmon, a Professor of Education specializing in School Finance, which is referred to in the Department's Brief to the Court of Appeals of the 10th Circuit at 17. It stated that basing the percentile exclusion on "students as the unit of analysis [as is done under the regulatory Appendix]…is a reasonable and efficient adjustment" to the disparity test. While this statement supports the Department's claim that its interpretation is a permissible one, it does not seem relevant to the question of whether or not the statute is ambiguous.

to the apparent lack of ambiguity of the part of the statute in dispute, we raised serious questions about the propriety of:

a) The assertion that using data on the LEAs, without considering their student populations, disadvantages states with a small number of large school districts without the caveat that this can occur only when there are many other small districts. Furthermore, we noted that if one or both of the LEAs at the low or high end of the distribution contain at least 5% of the population, under the DOE formula they would be the 5th or 95th percentile and no data below (or above) them would be deleted. Thus, the DOE formula can also fail to remove data from LEAs with anomalous characteristics.

b) The alternative calculation proposed by the DOE that combines the disparities for LEAs with different grade level groupings tends to decrease the calculated disparity. Thus, it is easier for a state in which school districts are organized by grade levels to pass the criterion for having an "equalization" program than other states with the same variation in average per-pupil expenditures.

c) The use of highly aggregated data to assess per-pupil expenditure inequality, which leads to an *underestimate* of the inequality in per-pupil expenditures. This allows states with substantial within district school-wide differentials in per-pupil expenditures as well as noticeable differences between the LEAs to "pass" the "equalization" criteria.

d) If one is concerned with assuring equal educational opportunity as reflected in expenditure or available revenue data on LEAs, it is not reasonable to delete the lowest five percent of the data if one desires to ensure a basic minimum education for all pupils. The pupils in those districts are the ones who are probably not receiving an adequate education relative to the typical child in the state.

e) Even if one desires to delete unusual observations or outliers, the choice of percentiles the DOE made to trim the data is not appropriate for assessing spread or relative disparity from skewed data. If part **(ii)** of the statute allowing for adjusting the basic expenditure data to account for special circumstances is implemented carefully, few "outliers" would remain. Of course, no adjustment system will be perfect and relevant statistical literature suggests that deleting a small fraction of data for the LEAs with very high per-pupil expenditures might well be appropriate.

While it is unreasonable to expect the formulas stated in either the 1976 or current DOE regulations to be perfect, the fact that there are several statistically questionable aspects suggests that in using its expertise the Department did not make a thorough examination of the relevant statistical literature. If courts are to defer to the "expertise" of a governmental agency when interpreting a statute, at a minimum they should check that a reasonable degree of "expertise" is reflected in the regulations.[64] The deficiencies in the first regulation established by the Department and the retention of the alternative formula in its new regulations raise questions about the expertise of the agency. The Department's

[64] Recall that the regulation, fn. 16 and accompanying text, states that the upper and bottom percentile school districts should be deleted based on accepted statistical principles but no references are provided. Under the Dauber*t*, 509 U.S. 579 (1993) criteria, the testimony of a scientific expert making such a statement without supporting citations to the literature might well be deemed inadmissible. In GE v. Joiner, 522 U.S. 136, 146 (1997) the Court went further and said courts need not admit evidence that is connected to existing data only by the *ipse dixit* of the expert.

failure to realize that they had data for the entire population rather than a sample as well as the arithmetic error noted previously[65] indicate that the agency was not very careful.[66]

In the future, both Congress and the Department might explore the large literature on measuring relative inequality before establishing a new formula. While this paper proposes that a measure based on the coefficient of dispersion will be superior to the existing measure, other measures related to the Lorenz curve, such as the Gini index or Pietra index might also be considered. Weighted versions of these measures could be constructed analogous to the formula for the weighted CD (WCD). These measures can be calculated on a per-school basis or a per-LEA basis and a suitable maximum allowable disparity developed. Alternatively, Congress or the DOE could lower the maximum allowable value of the current disparity measure calculated from LEA-wide data and also place a limit on a suitable disparity measure on school-wide expenditures within each LEA.

Without examining actual data from several states it is difficult to make a definitive suggestion as to the most appropriate measure to use. Indeed, how the costs incurred by the central administration of a school system are allocated may affect the cut-off criteria. If they are allocated on equal per-pupil basis, this would decrease the measured disparity so the maximum allowable disparity should be smaller if those expenditures are included. Our brief comparison of the current measures with data from school discrimination cases indicated that the current criteria do not ensure that the educational funding in a state is approximately uniform.[67]

Although the statistical soundness of either interpretation of the measure in the statute or the limitations of the data to which the measure is applied are unlikely to have a role in the ultimate resolution of *Zuni School District 89*, hopefully our discussion of the case and possible alternative measures will stimulate further research. Then a sound statistical literature will be available to Congress to rely on when creating similar measures that affect the allocation of millions of dollars.

Finally, a non-statistical issue that occurs in this law but may also affect other ones is the binary nature of the classification. Either a state's program is "equal" or not. Perhaps law makers could consider a sliding scale. Even with the current measure, one could say that 25% of the Impact Aid funds can be claimed by the state if the measure is less than 25%, 50% if the measure is less than 15% and 75% if the measure is less than 10%. Further examination by the DOE of per-pupil expenditure data for each school, rather than the average expenditure in an LEA, from a number of states with various

---

65 *See* fn. 34 and accompanying text.

66 This lack of care may have less legal relevance as it concerns the calculations the Department accepted in this case rather than a lack of care during its process of establishing the regulation. The method used to calculate the percentiles, however, is an important aspect of implementing the statute and it is surprising that both the 1976 and current regulations failed to realize that data for the total population are available.

67 *See supra* n. 48. It does not seem appropriate for a calculated disparity of 13% to be evidence of racial discrimination in school funding while a disparity of 24% satisfies the criteria for equality or uniformity of funding. Even a 10% difference in school funding may affect the quality of education students receive.

measures should help it decide on the most suitable one along with appropriate cut-offs to recommend to Congress.

Acknowledgments: It is a pleasure to thank all the statisticians who took the time to answer my question as to how the calculation described in the statute should be carried out. In addition I received very helpful comments on the form of the question used in the pilot study or on a draft of the manuscript from Profs. Yulia Gel, David H. Kaye, Weiwen Miao, Justine Shults, Dylan Small and Tian Zheng and most capable assistance checking the data analysis from Mr. Hiro Hikawa. I am very grateful for the cooperation of two lawyers: Mr. Ronald Van Amberg who represented the Zuni School District and Mr. Mark W. Smith of the Department of Education. They sent me a substantial amount of background material, including briefs and the relevant notices from the Federal Register, even though my research was independent of the case. Some of the statistical methods used in this research were supported in part by grant SES-0317956 from the National Science Foundation. Their application to the data in the *Zuni School District 89* case and discussion of the results are solely the views of the author.

APPENDIX A: Formal Statistical Descriptions of the Two Calculations in the Dispute

The available data consists of the average expenditure (or revenue) per pupil and the number of students enrolled in each of the n LEAs in the state. Formally, we index the n (89 in the case) LEAs by $i=1, \ldots, n$ and let s_i denote the number of pupils and r_i denote the average per-pupil expenditure (or available revenue) in the i^{th} LEA. Thus, we have a universe or statistical population of every LEA in a state and information on two characteristics or variables (population and average per-pupil expenditure) for each of the n LEAs. The plaintiff and Judge O'Brien give each LEA the same weight, 1/n, which is the way most statistical textbooks would do in creating the empiric distribution function[68] from a sample; which here would be the distribution of the variable (revenue) in the population of LEAs. Thus, the fraction of LEAS with a revenue less than or equal to t, say, is

$F_n(t)$= (the number of LEAs with r_i less than or equal to t)/n.

Notice that if one orders the LEAs in increasing order of their revenues (r_i), it is to determine the value of $F_n(t)$. The p^{th} percentile of $F_n(t)$ is the *smallest* value of t for which $F_n(t)$ is greater than or equal to p.

The original regulation of DOE and its interpretation of the formula in the new law weight each LEA by the fraction of the pupils in the state that it educates. Letting $S=\sum s_i$ be the total number of pupils in the state, then the weight given to the i^{th} LEA is $w_i = s_i/S$. Arranging the data for the LEAs in increasing order of their revenue (r_i), the distribution considered by the Department of Education is:

[68] In elementary statistics textbooks this curve is called the ogive or cumulative distribution. For example, see MARIO F. TRIOLA, ELEMENTARY STATISTICS 57-58 (10^{th} ed., Boston: Pearson, 2006) or JOHN VERZANI, USING R FOR INTRODUCTORY STATISTICS 266-67 (Boca Raton: CRC Press, 2005).

$F_D(t)$ = the sum of the w_i for all LEAs with r_i less than or equal to t.

While $F_n(t)$ gives equal weight or probability to each LEA, the distribution $F_D(t)$ assigns each LEA, a weight or probability equal to its fraction of all students in the state. When the data are obtained from a simple random sample of a population, $F_n(t)$ is the appropriate graphical summary of the data. In the context of complex sample surveys where the individual units (LEAs) are sampled with unequal probabilities, the weighted distribution $F_D(t)$ is the appropriate summary.

While the dissent's description of the Department's formula as "directing a complex and mystifying formula for determining which LEA's fall into the 5th and 95th percentiles of per-pupil expenditures"[69] may seem too strong, the term "weighted" is not mentioned in either the new statute or the regulations published by the DOE.

APPENDIX B: The Data and Supplementary Statistical Analysis

Since both the Department of Education and Congress apparently did not consider the usual or weighted version of the coefficient of dispersion as a measure of disparity to assess the uniformity of per-pupil expenditures, it is useful to display the details of their calculation. This will also enable us to determine the observations having a very large impact on the final result. These influential districts can be studied further to ascertain whether their values should be adjusted to account for one or more special factors noted in part (**ii**) of the statute.

In Table 1, the rank and name of the LEAs are listed in increasing order of their average revenue per-pupil. The third column reports the average per-pupil revenue and the number (Mem) of pupils in the LEA is given in the fourth column. The fifth column (ewt) is weight 1/89 given to them in the calculation corresponding to the statistician's disparity calculation and the sixth column simply cumulates these fractions so one can determine the percentiles. Similarly wtd and cumwtd, given in columns 7 and 8 report the weight each LEA receives in the DOE formula and their cumulative values from which the percentiles are determined. The absolute values (abdif) of the difference between the district's revenue and the median revenue ($3059) of all 89 districts is given in column 9 and the corresponding absolute differences (abdifd) between the district's revenue and the weighted median ($2992)for the DOE interpretation are given in column 10.[70] The product of the weight and column (10) gives the contribution (relwt) of the district to the weighted coefficient of dispersion (WCD). This column identifies the districts making the largest contribution to the WCD.

[69] *Supra* note 6 at 1171.
[70] The median of the weighted distribution is obtained by calculating the median according to the formula the Department of Education developed. After ordering the districts by per-pupil expenditure or revenue, total the number of pupils in the consecutive districts until you reach half of their grand total (300776.5). The per-pupil revenue of that district (Roswell) is the weighted median.

The ordinary coefficient of dispersion is the ratio of the average of the differences between the revenue per pupil of each district and their median value to their median value. Thus, the numerator is the average of the values in column 8 (abdif) and the denominator is the median. The numerator is also known as the mean deviation from the median. The differences between the per-pupil revenue of the districts and their median in this column also show us that the data are quite skewed (to the right). Notice that the first two differences from the median, i.e. $ 3461 and $ 2732, are much larger than the differences from the median of the last two districts, i.e. $ 387 and $ 334. In fact nine LEAs have average per-pupil expenditures that exceed the median by more than $ 387, the largest difference of all districts with expenditures less than the median. The data in this column identifies the districts that are major contributors to the coefficient of dispersion (CD) measure of disparity, e.g. the four districts with the largest expenditures.

The skewness and non-normality of the data can also be seen from a plot of the standardized ordered data on the y-axis against the expected values of an ordered sample of 89 observations from a normal curve.[71] If the data followed a normal distribution, the standardized ordered data should be near the 45 degree line in Figure 1. Some points will be above the line while some will be below it but the deviations should not be systematic. Data that have “heavier tails”, i.e. with more larger and smaller observations than a normal distribution with the same mean and standard deviation, will lie above the 45 degree line on the far right and below it on the far left portion of the graph. Right skewed data will lie above the 45 degree line on the right portion of the graph. Notice that the largest nine observations in Fig. 1 are noticeably above the line indicating that the revenue data are skewed to the right and noticeably larger than one expects to see in data from a normal curve.

To compare the plot of the standardized revenue data in the *Zuni School District* case to that of a normal distribution with the same mean and standard deviation. Figure 2 displays a plot of a random sample of 89 observations from such a normal. This plot of the standardized data from a normal distribution is much closer to the 45 degree line than the corresponding plot (Fig. 1) of the expenditure data in the case. Clearly, the two graphs are very different and demonstrate that the data in the case are not normal and not symmetric.

[71] Here we use a robust standardization, i.e. the difference between the observation and the median of the data set divided by the median absolute deviation, a robust estimate of the standard deviation. *See* Yulia Gel, Weiwen Miao and Joseph Gastwirth, *The importance of checking the assumptions underlying statistical analyses: graphical methods for assessing normality*, 46, JURIMETRICS J. 3-29 (2005).

Table 1: Number of Pupils and Average Revenue per Pupil for the 89 Districts and Supplementary Statistical Calculations

	District	Revenue	Mem	ewt	cuewt	wt	cumwtdoe	abdif	abdifd	relwt
1	Des Moines	2672	196.5	0.011	0.011	0.0007	0.001	387	320	0.209
2	Floyd	2725	261.5	0.011	0.022	0.0009	0.002	334	267	0.232
3	Hagerman	2777	475.5	0.011	0.034	0.0016	0.003	282	215	0.340
4	Dulce	2783	718	0.011	0.045	0.0024	0.005	276	209	0.499
5	Lake Arthur	2787	245.5	0.011	0.056	0.0008	0.006	272	205	0.167
6	Gadsen	2829	12000.5	0.011	0.067	0.0399	0.046	230	163	6.503
7	Hobbs	2848	8114.5	0.011	0.079	0.0270	0.073	211	144	3.885
8	Eunice	2849	780	0.011	0.090	0.0026	0.076	210	143	0.371
9	Quemado	2858	214.5	0.011	0.101	0.0007	0.076	201	134	0.096
10	Gallup	2861	13815	0.011	0.112	0.0459	0.122	198	131	6.017
11	Moriarty	2870	4643.5	0.011	0.124	0.0154	0.138	189	122	1.883
12	Carrizozo	2880	222.5	0.011	0.135	0.0007	0.139	179	112	0.083
13	Dexter	2883	1161.5	0.011	0.146	0.0039	0.142	176	109	0.421
14	Cloudcroft	2884	545	0.011	0.157	0.0018	0.144	175	108	0.196
15	Los Lunas	2887	7946.5	0.011	0.169	0.0264	0.171	172	105	2.774
16	Deming	2912	5326	0.011	0.180	0.0177	0.188	147	80	1.417
17	Grady	2915	159.5	0.011	0.191	0.0005	0.189	144	77	0.041
18	House	2936	119	0.011	0.202	0.0004	0.189	123	56	0.022
19	Aztec	2942	3283.5	0.011	0.213	0.0109	0.200	117	50	0.546
20	Truth-Cons.	2945	1771	0.011	0.225	0.0059	0.206	114	47	0.277
21	Belen	2948	4732.5	0.011	0.236	0.0157	0.222	111	44	0.692
22	Farmington	2948	10153	0.011	0.247	0.0338	0.256	111	44	1.485
23	Rio Rancho	2959	8590.5	0.011	0.258	0.0286	0.284	100	33	0.943
24	Capitan	2962	622.5	0.011	0.270	0.0021	0.286	97	30	0.062
25	Lovington	2963	2909	0.011	0.281	0.0097	0.296	96	29	0.280
26	Artesia	2964	3861	0.011	0.292	0.0128	0.309	95	28	0.359
27	Socorro	2968	2204.5	0.011	0.303	0.0073	0.316	91	24	0.176
28	Bloomfield	2968	3358.5	0.011	0.315	0.0112	0.327	91	24	0.268
29	Las Cruces	2974	21365.5	0.011	0.326	0.0710	0.398	85	18	1.279
30	Animas	2975	580.5	0.011	0.337	0.0019	0.400	84	17	0.033
31	Tucumcari	2975	1526.5	0.011	0.348	0.0051	0.405	84	17	0.086
32	Portales	2975	2806.5	0.011	0.360	0.0093	0.415	84	17	0.159
33	Alamogordo	2982	7824.5	0.011	0.371	0.0260	0.441	77	10	0.260
34	Clovis	2983	8639.5	0.011	0.382	0.0287	0.469	76	9	0.259
35	**Roswell**	**2992**	**10528.5**	**0.011**	**0.393**	**0.0350**	**0.504**	**67**	**0**	**0.000**
36	Dora	2996	254	0.011	0.404	0.0008	0.505	63	4	0.003
37	Estancia	3002	950	0.011	0.416	0.0032	0.508	57	10	0.032
38	Elida	3006	127	0.011	0.427	0.0004	0.509	53	14	0.006
39	Santa Rosa	3011	860	0.011	0.438	0.0029	0.512	48	19	0.054
40	Central	3027	314	0.011	0.449	0.0010	0.513	32	35	0.037
41	Pecos	3033	878.5	0.011	0.461	0.0029	0.516	26	41	0.120
42	Grants	3035	3699	0.011	0.472	0.0123	0.528	24	43	0.529
43	Santa Fe	3050	3044	0.011	0.483	0.0101	0.538	9	58	0.587
44	Questa	3054	638.5	0.011	0.494	0.0021	0.540	5	62	0.132
45	San Jon	3059	217.5	0.011	0.506	0.0007	0.541	0	67	0.048
46	Chama	3065	620	0.011	0.517	0.0021	0.543	6	73	0.150
47	Albuquerque	3071	83709.5	0.011	0.528	0.2783	0.821	12	79	21.987

48	Lordsburg	3074	860.5	0.011	0.539	0.0029	0.824	15	82	0.235
49	Reserve	3087	270	0.011	0.551	0.0009	0.825	28	95	0.085
50	Cimarron	3088	666.5	0.011	0.562	0.0022	0.827	29	96	0.213
51	Jal	3091	534	0.011	0.573	0.0018	0.829	32	99	0.176
52	Magdalena	3092	377.5	0.011	0.584	0.0013	0.830	33	100	0.126
53	Ft. Sumner	3100	428	0.011	0.596	0.0014	0.832	41	108	0.154
54	L.V. City	3122	2623.5	0.011	0.607	0.0087	0.840	63	130	1.134
55	Jemez Mt.	3123	415.5	0.011	0.618	0.0014	0.842	64	131	0.181
56	Espanola	3147	4920.5	0.011	0.629	0.0164	0.858	88	155	2.536
57	Clayton	3151	732.5	0.011	0.640	0.0024	0.861	92	159	0.387
58	Carlsbad	3152	6690	0.011	0.652	0.0222	0.883	93	160	3.559
59	Wagon Md.	3154	185	0.011	0.663	0.0006	0.883	95	162	0.100
60	Pojoaque	3155	1937.5	0.011	0.674	0.0064	0.890	96	163	1.050
61	Taos	3164	3326.5	0.011	0.685	0.0111	0.901	105	172	1.902
62	Melrose	3187	287.5	0.011	0.697	0.0010	0.902	128	195	0.186
63	Cobre	3194	2067	0.011	0.708	0.0069	0.909	135	202	1.388
64	Mountainair	3195	365	0.011	0.719	0.0012	0.910	136	203	0.246
65	Loving	3204	556.5	0.011	0.730	0.0019	0.912	145	212	0.392
66	Hatch	3206	1410.5	0.011	0.742	0.0047	0.917	147	214	1.004
67	Mesa Vista	3233	572	0.011	0.753	0.0019	0.918	174	241	0.458
68	L.V. West.	3241	2130.5	0.011	0.764	0.0071	0.926	182	249	1.764
69	Bernalillo	3244	3528.5	0.011	0.775	0.0117	0.937	185	252	2.956
70	Tularosa	3246	1140.5	0.011	0.787	0.0038	0.941	187	254	0.963
71	Raton	3249	1479.5	0.011	0.798	0.0049	0.946	190	257	1.264
72	Penasco	3259	709.5	0.011	0.809	0.0024	0.948	200	267	0.630
73	Tatum	3266	369	0.011	0.820	0.0012	0.950	207	274	0.336
74	Ruidoso	3278	2408	0.011	0.831	0.0080	0.958	219	286	2.290
75	Jemez Val.	3286	513.5	0.011	0.843	0.0017	0.959	227	294	0.502
76	Springer	3295	285.5	0.011	0.854	0.0009	0.960	236	303	0.288
77	Zuni	3320	1696	0.011	0.865	0.0056	0.966	261	328	1.850
78	Texico	3335	498	0.011	0.876	0.0017	0.967	276	343	0.568
79	Silver City	3391	3837.5	0.011	0.888	0.0128	0.980	332	399	5.091
80	Cuba	3404	773	0.011	0.899	0.0026	0.983	345	412	1.059
81	Logan	3484	278	0.011	0.910	0.0009	0.984	425	492	0.455
82	Roy	3516	113	0.011	0.921	0.0004	0.984	457	524	0.197
83	Mora	3530	707.5	0.011	0.933	0.0024	0.986	471	538	1.266
84	Maxwell	3591	152	0.011	0.944	0.0005	0.987	532	599	0.303
85	Hondo Val.	3690	157.5	0.011	0.955	0.0005	0.988	631	698	0.366
86	Vaughn	4641	111.5	0.011	0.966	0.0004	0.988	1582	1649	0.611
87	Los Alamos	5611	3509.5	0.011	0.978	0.0117	1.000	2552	2619	30.559
88	Corona	5791	81	0.011	0.989	0.0003	1.000	2732	2799	0.754
89	Mosquero	6520	57	0.011	1.000	0.0002	1.000	3461	3528	0.669

Source: Impact Aid Disparity Analysis submitted by New Mexico, Petitioner's Appendix at 210a -213a. Las Vegas is denoted by L.V., Valley and Mound are abbreviated as Val. and Md. Because special categories of pupils, e.g. those with special needs or disabilities, are multiplied by weighting factors the pupil population (Mem) need not be an integer. The data have been adjusted for the special factors mentioned in part (ii) of the statute.

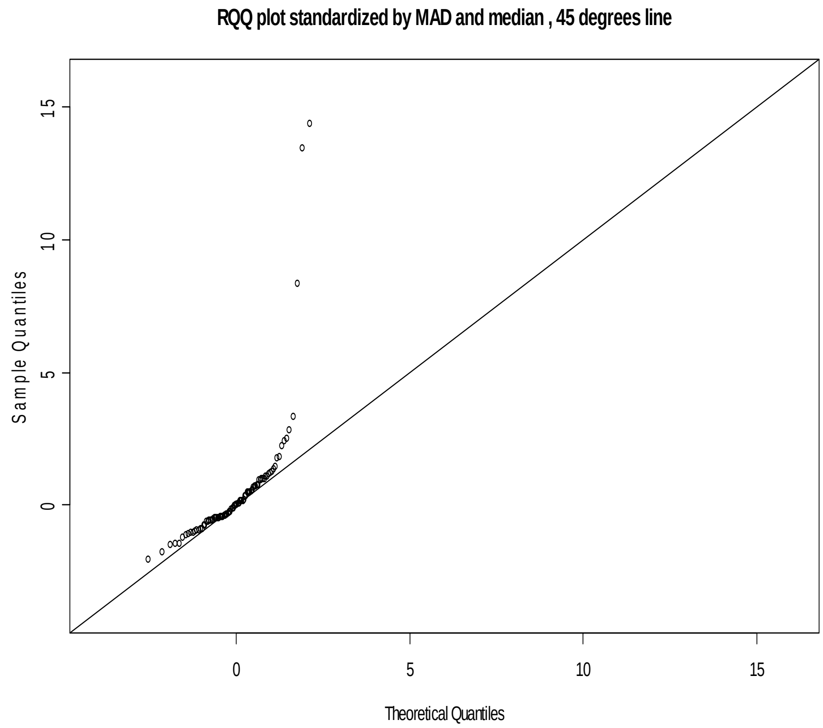

Figure 1: The Robust Quantile-Quantile Plot of the Average Revenue Per-Pupil for each LEA in New Mexico.

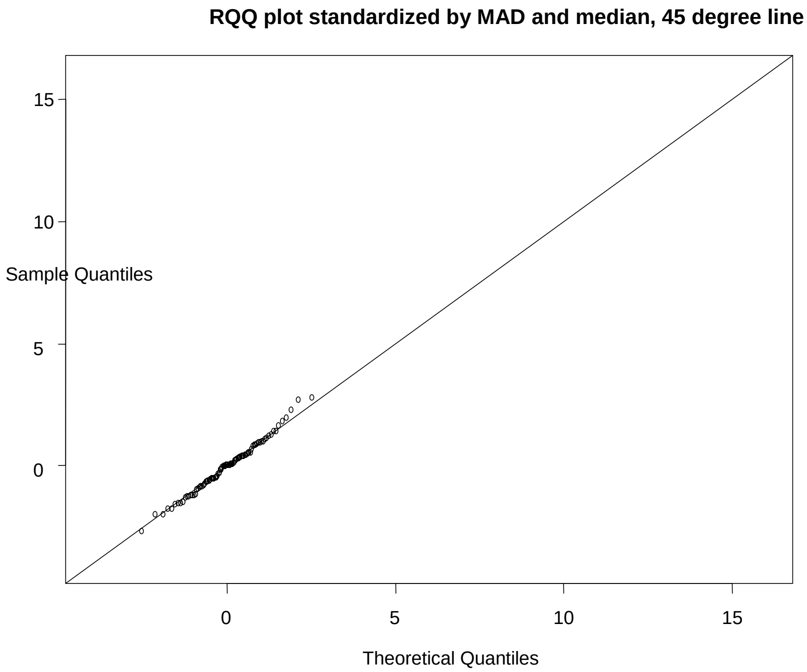

Figure 2: The Robust Q-Q plot for a random sample of 89 observations from a normal distribution with the same mean and standard deviation as the average revenue data.

APPENDIX C: The Two Forms of the Questionnaire

Version A

I am working on a paper on a statistical issue from a law case concerning the equality of the expenditures per-pupil in the school districts in a state. Of course, one cannot expect these per-pupil expenditures to be exactly equal so the law gives a formula to measure whether the per-pupil expenditures are equal in all districts (called local educational agencies or LEAs in the law). If this disparity measure calculated on the data meets the criterion specified in the law the state's funding is deemed **"equalized".** After reading the law, specify the data you would obtain from the state and its LEAs and describe how you would calculate the disparity measure described in the law.

The Law

The statute essentially says that the funding in a state is "equalized"

if the amount of per-pupil expenditures available to, the local educational agency in the State with the *highest* per-pupil expenditures or revenues did *not exceed* the amount of such per-pupil expenditures made by the local educational agency in the State with the *lowest* such expenditures by more than twenty-five percent (25%).

In making a determination under this subsection the Secretary shall—

(i) disregard local educational agencies with per-pupil expenditures or revenues above the 95th percentile or below the 5th percentile of such expenditures or revenues in the state

Now that you have read the **description of the measure of disparity and the criteria for determining** whether the funds available to the local educational agencies or LEAs, are "**equalized**", please describe:

1) The data you would obtain in order to make the stated calculation.

2) Describe how you would make the stated calculation from the data to decide on whether the state's school funding program met the criteria for being "equalized"?

3) Did any other formula or way of making the calculation described in the statute occur to you? If so, please describe it and any other data you might need to implement it.

In case respondents wanted to know more about why the issue was important at the end of both questionnaires I added some historical background at the end of the questionnaire. It did not mention the previous formula or the fact that Congress had taken away the Department of Education's authority to establish the new formula. Here is the information provided to the respondents:

Version B

I am working on a paper on a statistical issue that arose in a law case concerning whether the funding of the school districts in a particular state satisfies the criterion for being the same or equal. The term used in the law is **"equalized".** I would appreciate your help in deciding the calculation that a statistician should make to determine whether the criterion given in the law is satisfied.

The Data and the Law

Assume that for each school district (called **local educational agency or LEA** in the statute) there is reliable data on the number of students in the district (LEA) and the average expenditure per student. Thus, the data look like:

District	Number of Students	Average Expenditure Per-Pupil
1	500	$ 2500
2	750	$ 3200
3	250	$ 4200
4	1100	$ 3500
5	400	$ 2700
.		
.		
.		
N	800	$ 4050

The statute essentially says that the funding in a state is "equalized"

if the amount of per-pupil expenditures in the local educational agency in the State with the *highest* per-pupil expenditures did *not exceed* the amount of such per-pupil expenditures made by the local educational agency in the State with the *lowest* such expenditures by more than twenty-five percent (25%).

In making a determination under this subsection the Secretary shall—

(i) disregard local educational agencies with per-pupil expenditures above the 95th percentile or below the 5th percentile of such expenditures or in the state

Now that you have read the **description of the measure of disparity to be calculated** please describe:

1) How you would make the stated calculation from the given data.

2) Did any other way of making the calculation described in the statute occur to you? If so, please describe it.

For our purposes assume the available data on expenditures per-pupil are correct. **Feel free to choose a particular number, N, of districts that is convenient to illustrate the calculation you describe. In the actual law case N was between 50 and 100.** The **only question** is what calculation you would make if you were the statistical expert for the government agency administering the law.