432 Ed. Law Rep. 506

West's Education Law Reporter
April 10, 2025

Education Law into Practice

Zorka Karanxha, Perry A. Zirkel, Kristine Y. Hogarty[aa1]

A COMPARATIVE ANALYSIS OF THE LOWER AND APPELLATE DECISIONS UNDER THE IDEA FOR EACH FEDERAL CIRCUIT[a1]

Within the relatively extensive published research concerning the frequency and outcomes of court decisions under the Individuals with Disabilities Education Act (IDEA),[1] only two analyses included a comparison among the regions demarcated by the federal circuit courts of appeals.[2] The first analysis, which Zirkel and D'Angelo coauthored, was limited to the early period ending in 2000 and included Section 504 along with IDEA cases.[3] Within these delimitations, their circuit-by-circuit analysis found a rather wide range in the frequency of IDEA decisions, with one circuit (2d) as the clear leader and another circuit (10th) distinguishably at the low end.[4] For outcomes, they found a district-favorable skew in eight circuits, led by two of them (4th and 5th); a relatively even balance between districts and parents in three circuits (2d, 3d, and 9th); and a parent-favorable skew in one circuit (D.C.).[5]

The second study, which served as the springboard and predecessor for this article, covered a much longer and more up-to-date period--the twenty-five years from 1998 to 2022.[6] More specifically, in their circuit-by-circuit analysis, Zirkel and Karanxha found that four circuits (2d, 9th, 3d, and D.C. in descending order) led in the overall frequency of published decisions, with a few circuits (8th, 11th, and 10th in descending order) being at the low end of the range.[7] For outcomes, they found that the distribution was skewed in favor of ***507** school districts in all circuits, with two circuits (8th and 5th) at the high end and two others (D.C. and 6th) on the low end.[8]

However, these regional analyses mixed together, without differentiation, the decisions of both the federal district courts and the circuit courts of appeals.[9] This follow-up analysis of the 1998-2022 court decisions provides an exploratory differentiation of the frequency and, more particularly, the outcomes between these two judicial levels.

Method

The purpose of this follow-up study for published court decisions under the IDEA during the 25-year period from January 1, 1998 to December 31, 2022 was to differentiate the three-category outcome distributions[10] between the decisions of the U.S. Circuit Courts of Appeals (hereinafter referred to as ""appellate" courts) and those of U.S. District Courts (hereinafter referred to as "lower" courts) and examine the extent of their relationship.

More specifically, the research questions were as follows:

> 1. On an overall basis, i.e., for all the circuits taken together, is the outcomes distribution of the appellate decisions significantly different from that of the lower court decisions, and, if so, is the difference particularly strong?

2. On a circuit-by-circuit basis, what is the outcomes distribution for the appellate as compared to the lower court decisions?

3. For each of the sufficient-frequency circuits (1st, 2d, 3d, 4th, 6th, and 9th) and for the combination of the remaining circuits, is the outcomes distribution of the appellate decisions significantly different from that of the lower court decisions, and, if so, is the difference particularly strong?[11]

Although based on the same cases as the predecessor study,[12] this analysis made the following customized refinements: (a) excluding the dwindling proportion of state court ***508** decisions[13] and the handful of U.S. Supreme Court decisions;[14] (b) conflating decisions in the same case,[15] except attorneys' fees decisions[16] and those on the merits that were independent of each other;[17] and (c) combining the intermediate outcomes into a "mixed" category, thus yielding a three-category outcomes scale.[18] As a result, the analysis was based on a total of 1,193 court decisions.[19]

For research questions 1 and 3, we applied a two-step analysis. As the first step, we applied the chi-square test of independence to measure the difference between the observed and expected frequencies of the outcomes of cases at the two levels. If the chi-square determined that the difference was statistically significant, the second step was to apply Cramer's V for strength, or power, with the generally accepted categorization of effect size for the practical meaning of statistical significance.[20]

*509 Results

For research question 1, Table 1 presents the three-category outcomes distribution of the appellate and lower courts overall (i.e., for all the circuits taken together), along with the findings for statistical significance and effect size.

TABLE 1

Chi-Square Results for the Overall Sample

	Conclusively for District	Mixed	Conclusively for Parent
Appellate	66%	16%	18%
Courts	(n=413)	(n=98)	(n=115)
Lower Courts	50%	23%	27%
	(n=285)	(n=129)	(n=153)
$X^2=30.25^{a2}$			
Effect size=0.16			

Footnotes

a2 $p < .001$.

Examination of Table 1 reveals that the outcomes were skewed in favor of school districts at both court levels, with the skew significantly more pronounced at the appellate level. Although statistically significant, this difference was quite limited in its effect size.[21]

For research question 2, Figure 1 provides the outcomes distribution of the lower and appellate courts within each of the twelve federal circuits.

See last page of this article for Figure 1

***510** This figure shows that (1) the appellate courts for all circuits but D.C. had a pronounced district-favorable skew;[22] (2) the 5th and 8th circuits appeared to be particularly pro-district at both levels; and, with the exception of the two circuits (8th and 10th) that had so few decisions at the lower level to preclude this comparative analysis, (3) the 5th and D.C. circuits appeared to have the closest outcomes distributions between the lower and appellate levels.

For research question 3, the differences between the appellate and lower court levels for each of the sufficient-frequency circuits (1st, 2d, 3d, 4th, 6th, and 9th) and for the combination of the remaining circuits was not statistically significant.[23]

*511 Discussion

The IDEA litigation in this analysis was largely limited to officially published court decisions concerning IDEA issue categories of direct and primary interest to educators, such as identification, FAPE, LRE, and remedies.[24] Thus, they represent only the visibly significant tip of the "litigation iceberg" for school practitioners.[25] Thus, the generalizations in the descriptive and inferential statistical analysis herein are limited to selected case law that has weighty precedential effect but is not necessarily representative of the other levels of activity, including largely hidden settlements, in the adjudicative forum under the IDEA. Nevertheless, this segment of case law is practically significant due to not only the vertical and horizontal effect of precedent,[26] but also the prominence of the federal courts for IDEA litigation.[27]

Other methodological delimitations included that only the final decision on the merits was included in the source compilation of cases.[28] More specifically, any of the federal appellate decisions in the compilation eliminated any previously included federal district court decision for the same case. This procedure avoided double-counting but decreased the relationship between the appellate and lower court outcomes. It also reflected the fluid state of the case law for the last two-three years of the compilation, because some of those lower court decisions were subject to being superseded after the time lag for the appeal process.

For the research question 1, the finding that the outcomes skewed in favor of school districts at both the federal district and appellate courts showed that the composite results of the predecessor analysis applied to each of these judicial forums alone.[29] The related finding that this school-friendly skew was significantly more pronounced at the federal appellate level is more novel. This tendency may be attributable to the selection process, which depends on the extent that each side wins at the lower level, which side appeals, and whether the case ends before issuance of an appellate decision with a settlement. Nevertheless, while bearing in mind the weak effect size, this differential finding suggests caution for the parent side in appealing adverse federal district court decisions, because the odds for a reversal on appeal and, indirectly, for settlement become lower. Moreover, this

more pronounced skew in favor of schools at the appellate level would seem, within the boundaries of precedent, to signal continuing momentum that makes a balancing shift all the more difficult.

For research question 2, the finding that the courts, led by the 5th and 8th circuits, the decisions at the appellate level in each federal judicial region except the D.C. circuit evidenced a pronounced pro-district skew generally aligned with the composite results of the predecessor analysis.[30] However, the relative positions of each circuit were not the same for ***512** the appellate levels as they were for the composite analysis.[31] The comparison finding that the 5th and D.C. circuit court appeared to have the most similar outcomes distributions between the lower and appellate levels was of no importance upon the inferential statistical analysis for research question 3.

For research question 3, the lack of a statistically significant difference between the outcome distributions in each of the sufficient-frequency circuits (1st, 2d, 3d, 4th, 6th, and 9th) and the combination of the remaining circuits was surprising. However, this finding is cautionary due to the relatively limited cell sizes[32] and the conservative value of $p < .004$.[33] Alternatively, the lack of a significant difference may be attributable to the deferential standard for appellate review of federal district court decisions.[34]

Recommendations for future research include extending the comparative analysis between lower and appellate federal courts in the IDEA context to include decisions that are unpublished and, at the same time, to the rulings within major issue categories, such as identification and FAPE. Qualitative research concerning judicial ideology, the selection and settlement processes for appeals, and the availability and efficacy of parent attorneys will also deepen our understanding of the role of courts in the implementation of the IDEA. A dual approach for quantitative and qualitative follow up research would be particularly appropriate for differentiating and deciphering the impact of the "mixed" outcomes category.

For education practitioners as well as parents of students with disabilities, awareness of IDEA adjudication, including the frequency and outcomes trends at the judicial level, can contribute to making informed choices about collaboration as well as litigation. In general, focusing on the shared interest in improved student outcomes supersedes the differentiated statistical analysis of judicial outcomes.

Footnotes

a1 *Education Law Into Practice* is a special section of the Education Law Reporter sponsored by the Education Law Association. The views expressed are those of the author and do not necessarily reflect the views of the publisher or the Education Law Association. Cite as 432 Educ. L. Rep. 506 (April 10, 2025).

aa1 Dr. Karanxha is Professor of Educational Leadership and Policy Studies at the University of South Florida. Dr. Zirkel is a retired professor of education law and a past president of the Education Law Association who shares his work at *perryzirkel.com*. Dr. Hogarty is Director of Assessment in the College of Education at the University of South Florida.

1 *See, e.g.,* Perry A. Zirkel & Benjamin Frisch, *Longitudinal Trends of Judicial Rulings in K-12 Education: The Latest Look*, 407 Educ. L. Rep. 409 (2023) (tracking frequency trends of judicial decisions in public school context, including the segment in special education); Perry A. Zirkel & Cathy A. Skidmore, *National Trends in the Frequency and Outcomes of Hearing and Review Officer Decisions under the IDEA: An Empirical Analysis*, 29 Ohio State J. Disp. Resol. 525, 529, 531-32 (2014) (summarizing frequency and outcome analyses at the judicial level).

2 Perry A. Zirkel & Zorka Karanxha, *Longitudinal Trends in Special Education Case Law: An Updated Analysis*, 37 J. Special Educ. Leadership 42 (2024); Perry A. Zirkel & Anastasia D'Angelo, *Special Education Case Law: An Empirical Trends Analysis*, 161 Educ. L. Rep. 731 (2002).

3 Zirkel & D'Angelo, *supra* note 2, at 737-38. The cases were limited to IDELR-published decisions, and the outcome scale was limited to the three categories designated in their case abstracts. *Id.* at 738.

4 *Id.* at 741.

5 *Id.* at 747.

6 Zirkel & Karanxha, *supra* note 2, at 44. The court decisions were limited to those that were officially published or in the *Federal Appendix*. *Id.* at 42.

7 *Id.* at 47.

8 *Id.*

9 Both the Zirkel & D'Angelo and Zirkel & Karanxha analyses also included, without differentiation, the small and dwindling minority of state court decisions.

10 The three categories were: conclusively in favor of parent, mixed, and conclusively in favor of district. *Infra* note 18.

11 "Sufficient-frequency" in this context refers to those circuits that have, for the purpose of our statistical analysis (*infra* note 20), all the cells in the three outcome categories at the lower and appellate levels having a count of at least five decisions, which is the generally accepted standard for use of the chi-square test. *See, e.g.*, Dan Yates, David S. Moore, & George McCabe, The Practice of Statistics 734 (1999) (recommending that for chi-square analysis the expected cell count for at least 80% of the cells be a minimum of five). Based on this rigorous application of the accepted standard, the 5th, 7th, 8th, 10th, 11th, and D.C. circuits did not qualify and, thus, comprised the combined "other" category.

12 Zirkel & Karanxha, *supra* note 2. The database was the annually updated compilation of IDEA court decisions that is available at the second author's website, *perryzirkel.com*. We used the April 2023 version of the compilation to make sure that the data collection of the published decisions for the last few months of 2022 was complete in light of the limited time lag for official publication that is attributable to the technological speed of the underlying Westlaw electronic database. The compilation is systematically exhaustive of the officially published and Federal Appendix court decisions concerning the IDEA issue categories of direct and primary interest to educators, such as identification, free appropriate public education (FAPE), least restrictive environment (LRE), discipline, and the remedies of tuition (or IEE) reimbursement and compensatory education. It does not extend to those decisions limited to technical adjudicative issues, which are more directly and primarily of interest to litigators, such as jurisdiction, statute of limitations, exhaustion, and additional evidence. Finally, the compilation's coverage of the hybrid category of attorneys' fees is limited to a sampling of the representative issues, such as eligibility for and the scope of these fee awards.

13 *E.g.,* Petition of State, 294 A.3d 243, 416 Educ. L. Rep. 1102 (2022); Daniel G. v. Delaware Valley Sch. Dist., 813 A.2d 36, 172 Educ. L. Rep. 883 (Pa. Commw. Ct. 2002). The state cases amounted to approximately 5% of the compilation. Moreover, as Zirkel & Frisch, *supra* note 1, at 6 found, federal courts have increasingly been the forum for most special education decisions in recent years.

14 *E.g.*, Endrew F. v. Douglas Cnty. Sch. Dist. RE-1, 580 U.S. 386 (2017); Forest Grove Sch. Dist. v. T.A., 557 U.S. 230, 245 Educ. L. Rep. 551 (2009).

15 This conflation reduced the frequency count, but the recorded outcomes only changed if the outcome entries for the conflated decisions were different from each other. The conflation of different outcomes often resulted in categorization as mixed. However, for the limited number of cases that appeared under an initial category of identification, FAPE, or LRE and also under the remedies category of tuition reimbursement, the reimbursement outcome superseded the initial outcome due to the applicable multi-part, flowchart-like test. For example, if the FAPE decision was conclusively in favor of the parent (P), but the reimbursement rulings was conclusively in favor of the district (D), the outcome entry was D. This interaction, particularly for the FAPE and reimbursement rulings, requires an adjustment that is even more complicated for analyses on an issue-by-issue basis. *See, e.g.*, Zirkel & Skidmore, *supra* note 1, at 546-47, 553-54 (providing separate revised analysis for the issues categories for remedies).

16 Only a small proportion of the published attorneys' fees decisions also appeared elsewhere in the case compilation for a decision on the merits.

17 *E.g., compare,* Endrew F. v. Douglas Cnty. Sch. Dist. RE-1, 798 F.3d 1329, 321 Educ. L. Rep. 639 (10th Cir. 2015) (ruling on un-appealed procedural FAPE claims), *with* Endrew F. v. Douglas Cnty. Sch. Dist. RE-1, 290 F. Supp. 3d 1175, 353 Educ. L. Rep. 298 (D. Colo. 2018) (ultimate ruling for the substantive FAPE claim).

18 The original outcomes scale consisted of five categories. Zirkel & Karanxha, *supra* note 1, at 4. For the sake of a feasible statistical analysis on the basis of final decisions, not issue categories, we retained the polar categories (conclusively in favor of parent and conclusively in favor of district) but conflated the three intermediate categories into one broadened ""mixed" category. The resultant outcome categories for this analysis were as follows: P=conclusively for parent (i.e., all of the issue category rulings in favor of the parent) M=mixed (i.e., inconclusive rulings, conclusive rulings partially in favor of each side, and combinations) D=conclusively for district (i.e., all of the issue category rulings in favor of the school district)

19 The respective subtotals were 567 lower court and 626 appellate court decisions. Because the unit of analysis, with the limited aforementioned exceptions (*supra* notes 16-17 and accompanying text), was the final decision in the case, the lower court subtotal did not include published lower court decisions that were within the appellate subtotal.

20 The chi-square test is the generally accepted inferential statistic for frequency distributions of nominal data, indicating whether the difference between the two distributions is likely to be generalizable to the accessible population. We tested the null hypothesis of a chi-square test of independence for the appellate and lower courts, i.e., that the court outcomes at the two levels (appellate and lower courts) were not significantly different. *See, e.g.*, Meredith Gall et al., Educational Research 325-27 (2007). For inferential analysis the conventional minimum level of probability (p) is $p < .05$, representing a 95% probability that the results are not due to random chance or measurement error. We used the Bonferroni adjustment to control the probability of committing a type I error (i.e., false positive) when performing multiple statistical tests. *See, e.g.,* Olive J. Dunn, *On Multiple Comparisons among Means*, 56 J. Am. Stat. Ass'n 52 (1961); David C. Howell, Fundamental Statistics for the Behavioral Sciences 148 (2004). Thus, we used a minimum probability level of $p<.004$ to reduce the possibility of false positives. For results meeting this requisite level for statistical significance, effect size measures the magnitude of the difference in terms of its practical significance. *See, e.g.,* Jacob Cohen, *Things I Have Learned (So Far)*, 45 Am. Psych. 1306, 1310 (1990). Although various fields of study differ in their interpretation of effect sizes, we used the framework for social sciences, which identifies a minimum effect size of .2 as small, .5 as medium, and .8 as large. Christopher K. Ferguson, *An Effect Size Primer: A Guide for Researchers and Clinicians*, 40 Pro. Psych. Res. & Prac. 522, 533 (2009).

21 Based on the aforementioned framework (*supra* note 20), this effect size was within the weak category for strength.

22 "Pronounced" in this context means more than 50% conclusively in favor of districts, although the varying results in mixed category moderates such characterizations.

23 Based on the aforementioned adjustment (*supra* note 20) to the conservative level p<.004, none of the twelve tests was statistically significant. As a result, per general convention, we did not display these insignificant statistical data in tables.

24 *Supra* note 12. Moreover, for the particular purpose of this analysis, as expressed in the three research questions, the relatively small number of Supreme Court and published state court decisions were excluded. *Supra* notes 13-14 and accompanying text.

25 *See, e.g.,* Perry A. Zirkel, *The Role of Law in Special Education*, 31 Exceptionality 308, 311-12 (2023) (delineating the litigation iceberg in the IDEA context).

26 *See, e.g.,* Jeffrey C. Dobbins, *Structure and Precedent*, 108 Mich. L. Rev. 1453, 1460-63 (2010).

27 Zirkel & Frisch, *supra* note 1, at 13 (finding that federal courts accounted for more than 90% of the decisions in special education in recent decades).

28 For the source compilation, *see supra* note 12. Another delimiting methodological feature was the broad mixed outcomes category, which included the conflation procedure for separate issue category rulings that were different from each other. *Supra* notes 15-18 and accompanying text.

29 Zirkel & Karanxha, *supra* note 8, at 46.

30 *Id.* at 47.

31 For the percentage conclusively in favor of school districts, the 1st circuit was fourth at the appellate level but tenth on a composite basis, whereas the 4th circuit was tied for eighth at the appellate level but tied for third at the composite basis.

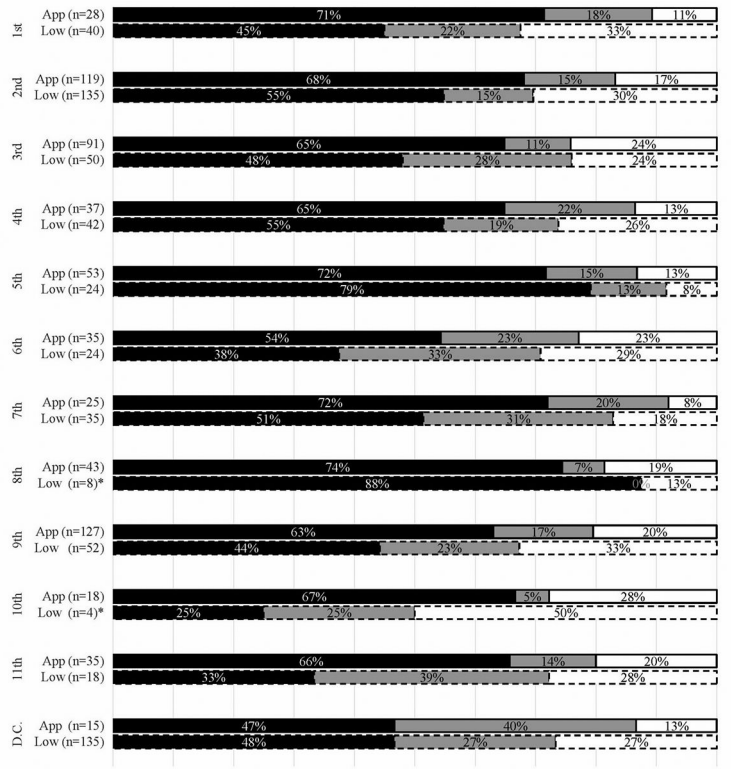
32 *See, e.g.*, William O. Bearden, *Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models*, 19 J. Marketing Res. 425 (1982); Shelby J. Haberman, *A Warning on the Use of Chi-Squared Statistics with Frequency Tables with Small Expected Cell Counts*, 83 J. Am. Stat. Ass'n 402, 555 (1988).

33 *Supra* note 20.

34 The standard is to uphold the lower court decision unless it is clearly erroneous, which is similar to the ample latitude afforded by an abuse of discretion standard of judicial review.

432 WELR 506

Figure 1. Outcome Distributions of Lower (Low) and Appellate (App) Court Decisions per Circuit



* These numbers are so small as to warrant particular caution in interpeting the corresponding outcome percentages.