Appendix: Data Sources, Methodology, and Cost-Effectiveness Calculations

Graff (2026)

Companion appendix to "School Choice Competition vs. New Education Spending: Estimating the Academic Benefits for Public School Students"

---

1. Overview

This appendix documents the data sources, methodology, and calculations underlying the cost-effectiveness comparison presented in the "Scale vs. Spend" policy brief. The analysis addresses a straightforward policy question: given a fixed amount of public resources, which approach produced larger gains for public school student achievement in Florida – creating competition through scaling school choice program or directing those same resources toward increased K–12 education spending? To answer this question, the analysis compares the achievement gains produced by scaling Florida's Tax Credit Scholarship (FTC) program, as measured by the competitive effects on public school students documented in Figlio, Hart, and Karbownik (2023), to the gains that would have resulted from spending the cumulative $2.84 billion (in 2018 dollars) on direct K–12 education funding increases, using the consensus spending effect size from Jackson and Mackevicius (2024).

Three policy counterfactual scenarios are constructed, each distributing the same total funding across progressively narrower student populations. Counterfactual #1 distributes funding across all K–12 public school students in Florida (the broadest scenario and upper bound on the effectiveness multiplier). Counterfactual #1 is not an apples-to-apples comparison because it distributes school spending effects across all K-12 students in the state, not the 55.8% in higher competition schools. However, this is perhaps the most likely policy counterfactual (corresponding to a general increase in the state education budget). Counterfactual #2 distributes funding across only the 55.8% of students in schools facing above-median competition from nearby private schools – the population that directly matches Figlio et al.'s treatment group. Counterfactual #3 further restricts to above-median students in grades 3–8 only, matching the exact test-score sample (the narrowest scenario and maximally conservative). Across all three scenarios, the competitive effects of scaling the tax credit scholarship program substantially outperform what the same resources would have achieved through direct spending increases. The preferred specification (Counterfactual #2) finds that competitive effects were approximately 11 times more cost-effective than equivalent spending increases.

All calculations, data sources, and assumptions are documented below in sufficient detail to permit independent replication. The analysis intentionally relies on conservative assumptions at each decision point, as detailed in Section 5.

2. Data Sources

2.1 Competitive Effects of the Florida Tax Credit Scholarship

The primary source for competitive effect sizes is Figlio, Hart, and Karbownik (2023), "Effects of Maturing Private School Choice Programs on Public School Students," published in the *American Economic Journal: Economic Policy*. This study examined the Florida Tax Credit Scholarship (FTC) program over its first 15 years of operation (2003–2017), making it the longest panel study of competitive effects in the school choice literature.

Figlio et al. (2023) linked Florida state birth records to student-level public school administrative records, creating a near-universal panel of public school students in tested grades. They constructed a competitive pressure index for each public school based on the proximity, density, enrollment capacity, and diversity of nearby private school options that accepted FTC scholarships. Schools were classified into above-median and below-median competitive pressure areas. Approximately 55.8% of Florida public school students attended schools in above-median competitive pressure areas (Figlio et al., 2023, Appendix Table A2).

The key finding used in this analysis is the endpoint effect for reading achievement: public school students in above-median competitive pressure areas experienced a 0.166 standard deviation improvement in reading scores after 15 years of program operation, relative to students in below-median areas. This effect emerged gradually, consistent with a maturation hypothesis in which competitive effects strengthen as the private school market develops and public schools adapt. The study also found that low-income students experienced even larger effects – 0.196 SD after 12 years – and that the program reduced suspensions and absences, indicating benefits beyond test scores.

2.2 Public School Enrollment

Public school enrollment data for 2003–2017 were obtained from the Florida Department of Education (FL DOE) Fall Membership counts, which report PreK–12 enrollment as of October of each school year. Over the study period, enrollment ranged from approximately 2,541,814 (2002–03) to 2,817,076 (2016–17). Data were assembled from the following FL DOE sources:

- **2002–03:** FL DOE Statistical Brief, available at: https://www.fldoe.org/core/fileparse.php/7562/urlt/03-04.pdf
- **2003–04:** FL Economic and Demographic Research, available at: https://www.edr.state.fl.us/Content/special-research-projects/education/Characteristicsofstudents.pdf
- **2005–13:** Annie E. Casey Foundation Kids Count Data Center (sourced from FL DOE), available at: https://datacenter.aecf.org/data/tables/5342-public-school-student-enrollment?loc=11&loct=2
- **2014–17:** FL DOE PK–12 Public School Data Publications, available at: https://www.fldoe.org/accountability/data-sys/edu-info-accountability-services/pk-12-public-school-data-pubs-reports/archive.stml

An important feature of the FL DOE Fall Membership counts is that they do not include most Voluntary Prekindergarten (VPK) students. Florida's VPK program, enacted in 2005, is primarily delivered through private providers and community-based organizations; VPK students are not captured in public school fall membership counts. The PreK students who do appear in the FL DOE data are those enrolled in public preschool programs (e.g., Head Start partnerships, district-operated PreK, and special education PreK). These students constitute only about 2.2% of the total fall membership count. Because the FL DOE data already exclude most VPK students, the K–12 enrollment figures used in this analysis require only a minimal adjustment (see Section 3.1).

As a sensitivity check, enrollment figures were also obtained from the National Center for Education Statistics (NCES) Common Core of Data. The NCES totals are very similar to the FL DOE figures and produce virtually identical cost-effectiveness multipliers (see Section 6).

2.3 School Spending Effect Size

The spending counterfactual uses the effect size from Jackson and Mackevicius (2024), "What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States," published in the *American Economic Journal: Applied Economics*. This meta-analysis synthesized 31 plausibly causal studies of the effect of school spending on student achievement, requiring each included study to use a credible identification strategy (regression discontinuity, instrumental variables, difference-in-differences, or randomized controlled trial).

Jackson and Mackevicius report an average effect of 0.0079 standard deviations per $1,000 per pupil per year, expressed in 2018 dollars. This estimate has become the most frequently cited consensus estimate of school spending effects in the education economics literature. It is broadly consistent with recent evaluations of ESSER (pandemic-era) spending, which have found effects in the range of 0.0049–0.0086 SD per $1,000 per pupil. It is also consistent with the conclusions of Handel and Hanushek's (2023) review, which finds positive but modest average effects of spending on achievement, with substantial heterogeneity across contexts.

The Jackson and Mackevicius estimate is used as the per-dollar achievement return in the spending counterfactual: if the $2.84 billion spent on the FTC program had instead been distributed as per-pupil funding increases, what achievement gain would we expect? This framing isolates the question of comparative cost-effectiveness, holding total resources constant.

2.4 Tax Credit Scholarship Funding

Annual funding data for the Florida Tax Credit Scholarship program were obtained from Figlio et al. (2023) and Florida Legislature appropriations records. Nominal funding grew from approximately $50 million in 2003 to $539 million in 2017, driven in large part by a statutory escalator provision: when claimed tax credits exceeded 90% of the annual cap, the cap automatically increased by 25% the following year. This mechanism produced rapid, market-driven program growth without requiring annual legislative action.

All nominal dollar amounts were converted to constant 2018 dollars using Bureau of Labor Statistics Consumer Price Index for All Urban Consumers (CPI-U) deflators. The cumulative total over the 15-year study period (2003–2017) is $2,837,658,352 in 2018 dollars (approximately $2.5 billion in nominal terms). This figure represents the total cost of the FTC program and serves as the numerator in the cost-effectiveness calculation: both the competitive effects and the spending counterfactual are evaluated against this same resource base.

3. Methodology

3.1 PreK Enrollment Adjustment

Because the FL DOE Fall Membership counts report PreK–12 enrollment, a small adjustment is required to obtain K–12 figures. As noted in Section 2.2, the FL DOE counts do not include most VPK students – only PreK students enrolled in public preschool programs. Based on FL DOE enrollment breakdowns by grade level, PreK students constitute approximately 2.2% of total fall membership, or roughly 0.022 grade-equivalents.

The K–12 enrollment for each year is calculated as:

$$\textit{K–12 Enrollment} = (\textit{PreK–12 Enrollment} / 13.022) \times 13$$

The divisor of 13.022 reflects 13 K–12 grade levels plus 0.022 grade-equivalents of public preschool PreK enrollment. Because FL DOE already excludes most VPK students, this adjustment is very small – typically reducing the enrollment figure by less than 0.2% – and the resulting K–12 figures are highly robust to alternative PreK assumptions. The sensitivity analysis in Section 6 confirms that using NCES enrollment data (which report K–12 directly) produces identical multipliers.

3.2 Treatment Population Size (55.8%)

Figlio et al. (2023) report in the Appendix (Table A2, p.5) that out of their public school student sample of 6,971,914 student-years, there were 3,890,161 student-years from schools in above-median competition schools (i.e., subject to higher levels of competition). This ratio comes out to 55.8% of Florida public school student-years in the sample, which is drawn from roughly 1.2 million unique students from 2002–2003 to 2016–2017. This proportion is then applied to the K–12 enrollment total to estimate the treatment population size, which then determines the public per-pupil spending amount in each counterfactual. For my preferred specification Counterfactual #2, I estimate the treatment population size in the following way:

$$\textit{Above-Median K–12 Enrollment} = \textit{K–12 Enrollment} \times 0.558$$

This proportion is held constant across all years, which is standard practice in policy analysis and is consistent with the Figlio et al. (2023) design, which defines treatment status based on time-invariant school characteristics. In practice, the share of students in above-median areas may vary slightly year to year due to enrollment shifts, but the 55.8% figure represents the best available estimate from the source study.

3.3 Grades 3–8 Estimation

The achievement effects in Figlio et al. (2023) are measured using Florida's standardized test scores, which are available only for students in grades 3–8. To estimate the number of above-median students in these tested grades, the analysis applies a uniform cohort assumption:

$$\text{Above-Median Grades 3–8} = (\text{Above-Median K–12} / 13) \times 6$$

This assumes that each of the 13 grade levels (K–12) contains an equal share of the total student population. In practice, lower grades tend to be slightly larger than upper grades due to grade retention in early years and dropout in later years. This means the 6/13 fraction likely underestimates the true grades 3–8 population. The direction of this bias is conservative: underestimating the 3–8 population means overstating the per-pupil spending in the counterfactual, which yields a larger counterfactual effect and a smaller effectiveness multiplier. The 5.2x lower-bound estimate is therefore a conservative one.

3.4 Counterfactual Construction

The core of the analysis is a year-by-year counterfactual calculation. For each year t in the study period (2003–2017), the analysis asks: if the FTC program funding for that year had instead been distributed as a per-pupil spending increase to the relevant student population, what achievement gain would Jackson and Mackevicius's (2024) spending effect imply?

The formula for each year t is:

$$\text{Counterfactual Effect} = (0.0079 \times \text{FTC Funding}) / (\text{Enrollment} \times 1{,}000)$$

Where:

- **0.0079** = the Jackson and Mackevicius (2024) estimate of the achievement effect of $1,000 per pupil per year of additional school spending, in standard deviations (2018 dollars)
- **FTC Funding** = total tax credit scholarship funding in the given year, expressed in 2018 dollars
- **Enrollment** = the relevant student population for the counterfactual scenario in the given year
- **1,000** = denominator converting the per-$1,000 effect to a per-dollar effect

All funding amounts and the Jackson and Mackevicius effect size are denominated in 2018 dollars, so no additional inflation adjustment is needed within the formula. The cumulative counterfactual effect is the sum of the annual effects across all 15 years. The effectiveness multiplier is then computed as the ratio of the observed competitive effect (0.166 SD from Figlio et al.) to the cumulative counterfactual spending effect:

$$\text{Effectiveness Multiplier} = 0.166 / \text{Cumulative Counterfactual Effect}$$

3.5 The Three Counterfactual Scenarios

Counterfactual #1: All K–12 Students (Upper Bound). This scenario distributes the $2.84 billion in FTC funding across all K–12 public school students in Florida, totaling 40,057,093 student-years over the 2003–2017 period. This yields an average spend of $70.84 per pupil per year and a cumulative spending counterfactual of 0.0092 SD, producing an effectiveness multiplier of 20.1x. This is the upper-bound estimate because it distributes funding across the broadest possible population, including students in low-competition areas who are not in Figlio et al.'s treatment group. Distributing the same dollars over a larger population dilutes the per-pupil spending increase and therefore reduces the counterfactual achievement gain, making the competitive effect appear relatively more impressive.

Counterfactual #2: Above-Median K–12 Students (Preferred Specification). This scenario restricts the distribution to the 55.8% of K–12 students in above-median competitive pressure areas, totaling 22,351,858 student-years. This yields an average cost of $126.95 per pupil per year and a cumulative counterfactual of 0.0165 SD, producing an effectiveness multiplier of 11.2x. This is the preferred specification because it matches the counterfactual population to the treatment population in Figlio et al. (2023). The competitive effect of 0.166 SD was estimated for students in above-median areas, so the relevant policy counterfactual asks: what if those same resources had been spent on those same students? This scenario provides the most apples-to-apples comparison and is the most policy-relevant of the three.

Counterfactual #3: Above-Median Grades 3–8 Students (Lower Bound). This scenario further restricts to above-median students in grades 3–8 only, totaling 10,316,242 student-years. This yields an average cost of $275.07 per pupil per year and a cumulative counterfactual of 0.0357 SD, producing an effectiveness multiplier of 4.6x. This is the lower-bound estimate because it matches the exact population for which test-score data are available. However, it is overly restrictive as a policy counterfactual because competitive effects certainly extend to grades beyond the 3–8 testing window. Students throughout the whole K-12 grade span were eligible for the FTC program and students in grades K–2 and 9–12 also attended (sometimes the same) schools that are very likely responding to competitive pressure; restricting to tested grades artificially restricts the hypothesized population to which the Figlio et al. (2023) study generalizes.

Of the three scenarios, Counterfactual #2 is the most defensible and policy-relevant comparison. It avoids the dilution problem of Counterfactual #1 (which includes students outside the treatment population) and the artificial restriction of Counterfactual #3 (which excludes students who plausibly benefit from the same competitive dynamics). The headline finding of the policy brief – that competitive effects are at least 11 times more cost-effective – is based on this primary comparison. The range of 5.2x to 20.1x across all three scenarios provides a transparent sensitivity band.

4. Results

Table 1 presents the complete cost-effectiveness comparison across all three counterfactual scenarios. All dollar figures are expressed in 2018 dollars. Effect sizes are reported in standard deviations and translated into percentile-point equivalents (evaluated at the mean of the normal distribution) for interpretability.

Table 1. *Cost-Effectiveness Comparison: FTC Competitive Effects vs. Equivalent Spending Increases*

Metric	CF #1: All K–12	CF #2: Above-Med. K–12	CF #3: Above-Med. 3–8
Total Cost (2018$)	$2,837,658,352	$2,837,658,352	$2,837,658,352
Student-Years	40,057,093	22,351,858	10,316,242
Cost per Student/Year	$70.84	$126.95	$275.07
FTC Competitive Effect (SD)	0.166	0.166	0.166
Spending Counterfactual (SD)	0.0082	0.0148	0.0320
FTC Effect (Percentile Points)	6.1	6.1	6.1
Spending Effect (Percentile Pts)	0.3	0.6	1.2
Effectiveness Multiplier	**20.1x**	**11.2x**	**5.2x**

Note: *All dollar amounts in 2018 dollars. FTC competitive effect from Figlio, Hart, & Karbownik (2023), 15-year endpoint, reading, above-median students. Spending effect from Jackson & Mackevicius (2024), 0.0079 SD per $1,000/pupil/year. Percentile-point conversions evaluated at the mean of the standard normal distribution.*

The results are unambiguous across all three scenarios: the competitive effects of scaling the FTC program produced substantially larger achievement gains than the same resources would have produced through direct spending increases. Under the primary comparison (Counterfactual #2), the FTC program's competitive effect of 0.166 SD (6.1 percentile points) is approximately 11 times larger than the 0.016 SD (0.6 percentile points) that would have resulted from distributing the same $2.84 billion as per-pupil spending increases to above-median-competition students. Even under the most conservative scenario (Counterfactual #3), which restricts the comparison to the exact tested-grade population, the competitive effect is still 5.2 times more cost-effective.

To put the magnitude in perspective: the spending counterfactual under Counterfactual #2 translates to approximately 0.6 percentile points of improvement – a gain that would be difficult to detect in any individual student's performance and would likely be invisible to parents, teachers, and principals. The competitive effect, by contrast, translates to 6.1 percentile points – a substantively meaningful improvement that represents roughly two-thirds of a year of additional learning.

5. Conservative Assumptions and Sources of Understatement

The analysis incorporates several conservative assumptions, each of which biases the effectiveness multiplier downward. That is, each assumption makes the competitive effects of the FTC program look even lower than the true effect in comparison to the spending increase approach.

1. **Equilibrium effects are not captured by the study design.** The Figlio et al. (2023) identification strategy compares students in above-median versus below-median competitive pressure areas. Any competitive effect that is shared by all schools statewide – regardless of their level of private school competition – is absorbed into the common trend and is not captured in the estimated treatment effect. However, supplementary analyses from Figlio et al. provide evidence that competitive spillovers extend beyond the immediate treatment group. For students in above-median competition *districts* whose individual schools are in the below-median comparison group, the authors found positive and statistically significant effects of competition (Appendix, Table A6, p.9), and the main paper concludes that "the effect of district-level competition dominates the school-level measures" for nearly every outcome (p. 287). The true competitive effect for reading scores is then certainly larger than 0.166 SD, which is not accounted for in this analysis. This is a standard limitation of difference-in-differences designs in settings with potential general-equilibrium spillovers.
2. **No fiscal savings are included.** Public schools spend substantially more per pupil than the average FTC scholarship amount. When students leave the public system for private schools, the public school retains the difference between its per-pupil funding and the scholarship cost. Research suggests that school choice programs become fiscally cost-neutral when approximately 57% of participants are students who would otherwise have attended public school (Lueken, 2024). These fiscal savings – which would reduce the net cost of the FTC program – are not included in the analysis. Including them would lower the denominator of the cost-effectiveness ratio and increase the multiplier.
3. **Assumes linear returns to increasing school spending over time.** Jackson and Mackevicius' (2024) meta-analysis relied upon quasi-experimental studies of school spending most often undertaken at the school district level over a comparatively (~3-5 year) intervention period. Their average effect of .0316 SD is only a four year estimate (.0079 per $1,000 per pupil per year). There is comparatively less evidence that permanent spending increases continue to produce linear returns on student achievement after 10+ years. In my analysis, however, I assume linear returns of the effect of additional school spending. The true effect of additional spending is likely lower over 10-15 years, and so this assumption downwardly biases the cost-effectiveness multiplier.
4. **Assumes no decay in the magnitude of the spending effect size at a statewide scale.** The studies included in the meta-analysis largely studied school spending interventions at the

district level, not spending interventions that affected a majority of students across the state. Maintaining that effect size at scale is unlikely. For example, both of the ESSER studies, which increased education spending at scale, found smaller effects on reading test scores than .0079 SD at scale. One of the ESSER evaluations found no effect of additional funding on reading scores. Therefore, assuming a .0079 SD effect is a conservative assumption of additional spending on reading scores at scale, further biasing the multiplier downwards.

5. **The program continued growing after 2017.** The study period ends in 2017, but the Florida legislature subsequently expanded the FTC program and created additional choice programs, including education savings accounts (ESAs) and universal eligibility. If competitive effects continued to mature beyond the 15-year study window – as the maturation trajectory suggests they would – the endpoint effect used in this analysis understates the full long-run competitive impact.

6. **Only reading scores are used.** Figlio et al. (2023) found positive competitive effects on both reading and math achievement, as well as reductions in student suspensions and absences. The analysis uses only the reading effect (0.166 SD) because it is the most precisely estimated and the largest. Including math effects and behavioral outcomes would provide a more complete picture of the benefits of competition, further strengthening the cost-effectiveness comparison.

6. Sensitivity Analyses

To assess the robustness of the findings, the cost-effectiveness calculations were replicated using alternative enrollment data from the National Center for Education Statistics (NCES) Common Core of Data. Table 2 compares the effectiveness multipliers under the primary specification (FL DOE enrollment with 2.2% PreK adjustment) to the NCES-based specification.

Table 2. *Sensitivity of Effectiveness Multipliers to Enrollment Data Source*

Enrollment Source	CF #1: All K–12	CF #2: Above-Med. K–12	CF #3: Above-Med. 3–8
Primary (FL DOE, 2.2% PreK adj.)	20.1x	11.2x	5.2x
NCES Common Core of Data	20.1x	11.2x	5.2x

The results are identical across data sources. The FL DOE and NCES enrollment totals for Florida are very similar over this period, and the 2.2% PreK adjustment applied to the FL DOE data aligns the two series closely. This robustness check confirms that the findings are not sensitive to the choice of enrollment data source or to the method used to handle PreK enrollment.

More broadly, the effectiveness multipliers are most sensitive to two parameters: the Figlio et al. endpoint effect size (0.166 SD) and the Jackson and Mackevicius spending effect size (0.0079 SD per $1,000). Both of these are sourced from peer-reviewed publications in top economics journals and represent the best available estimates in the literature. The analysis does not attempt to construct confidence intervals around the multipliers because the two underlying effect sizes come from different research designs and populations, making formal statistical combination inappropriate. Instead, the three counterfactual scenarios provide a transparent range that captures uncertainty about the relevant comparison population.

References

Figlio, D. N., Hart, C. M. D., & Karbownik, K. (2023). Effects of maturing private school choice programs on public school students. *American Economic Journal: Economic Policy, 15*(4), 255–294.

Florida Department of Education. Fall Membership Reports, various years. Retrieved from https://www.fldoe.org/accountability/data-sys/edu-info-accountability-services/pk-12-public-school-data-pubs-reports/archive.stml

Handel, D. V., & Hanushek, E. A. (2023). U.S. school finance: Resources and outcomes. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 7). North-Holland.

Jackson, C. K., & Mackevicius, C. L. (2024). What impacts can we expect from school spending policy? Evidence from evaluations in the United States. *American Economic Journal: Applied Economics, 16*(1), 412–446.

Lueken, M. (2024). *Fiscal effects of school choice*. EdChoice.

National Center for Education Statistics. Common Core of Data (CCD). Retrieved from https://nces.ed.gov/ccd/