WestEd

FACSA
FLORIDA ASSOCIATION OF CHARTER SCHOOL AUTHORIZERS

Charter School Authorizer Practices and Evaluations in Florida

A Report Funded by the MIT Blueprint Charter Research Collaborative

Olivia Szendey, Kelly Wynveen

August 2026

Suggested citation: Szendey, O., & Wynveen, K. (2026). *Charter school authorizer practices and evaluations in Florida: A report funded by the MIT Blueprint Charter Research Collaborative*. WestEd.

WestEd is a nonpartisan, nonprofit organization that aims to improve the lives of children and adults at all ages of learning and development. We do this by addressing challenges in education and human development, increasing opportunity, and helping build communities where all can thrive. WestEd staff conduct and apply research, provide technical assistance, and support professional learning. We work with early learning educators, classroom teachers, local and state leaders, and policymakers at all levels of government.

For more information, visit WestEd.org. For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the *E-Bulletin*, our semimonthly e-newsletter, at WestEd.org/subscribe.

Executive Summary

Charter school authorizers play a critical role in determining which charter school applicants are approved to open schools and ultimately create educational opportunities for students and families. Despite the importance of these decisions, there is limited empirical evidence regarding whether application-stage evaluations successfully identify schools that will remain viable and effective over time. To address this gap, WestEd and the Florida Association of Charter School Authorizers (FACSA) partnered on a pilot study, funded through the MIT Blueprint Research Collaborative, examining the relationship between charter school application evaluation ratings and long-term school outcomes in the state of Florida.

Florida provides a unique context for this research because all district authorizers use a common state-developed charter school application evaluation instrument and have used this instrument over the past decade. The statewide use of a single rubric created an opportunity to examine whether application ratings predict subsequent school performance and organizational sustainability and to identify which aspects of applications are most closely associated with long-term success.

This pilot study examined 209 approved charter school applications from schools opened between 2013 and 2023 across 22 Florida school districts. Evaluation instrument data were collected from 31 district authorizers and merged with additional school-level information, including operational status, enrollment, demographics, and accountability outcomes. Using descriptive statistics, exploratory factor analysis, and multilevel regression modeling, the study investigated both the overall predictive value of evaluation scores and the contribution of individual application components to school outcomes.

Several notable findings emerged. First, approved charter school applications generally received very high scores, with schools averaging approximately 86 percent of available points. This limited variation suggests that the evaluation instrument may function more effectively as a threshold screening tool than as a mechanism for distinguishing quality among approved applicants. Second, exploratory factor analysis revealed that rubric items align more closely with three empirically derived dimensions of school readiness—Governance and Operations, Curriculum and Design, and Culture and Support—than with the instrument's current organization into Educational Plan, Organizational Plan, and Business Plan sections.

The study found little evidence that application scores predict Florida school accountability grades. However, application ratings demonstrated a meaningful relationship with school sustainability. Schools receiving lower overall rubric scores were significantly more likely to close, even after accounting for enrollment, student demographics, and geographic factors. At the item level, stronger ratings in governance, human resources planning, English Language

Learner planning, and startup readiness were associated with lower probabilities of closure. These findings suggest that operational readiness and organizational capacity may be more important predictors of long-term viability than educational program design alone.

The analysis also revealed substantial variation in scoring practices across districts. Some districts scored applications considerably more strictly than others, resulting in a 32-point difference between the lowest and highest scoring districts. However, districts that scored more strictly did not produce schools with stronger accountability outcomes or lower closure rates. This finding suggests that score severity alone is not a meaningful indicator of authorizing effectiveness, and it highlights the need for greater calibration and consistency in how evaluators apply rubric criteria.

The findings have important implications for policy and practice. As Florida continues to review and revise its charter school application evaluation instrument, the results provide empirical evidence that can inform those revisions. Specifically, the findings suggest opportunities to strengthen the instrument's focus on operational readiness, improve alignment between rubric structure and underlying dimensions of school readiness, and support greater consistency in application reviews through interrater reliability training and calibration activities. More broadly, the study contributes to a growing body of evidence on charter school authorization and offers insights that may be valuable to authorizers in other states.

Ultimately, this pilot study demonstrates that application-stage evaluations contain useful information about the long-term sustainability of charter schools while also highlighting opportunities to strengthen how authorizers assess school readiness. By refining evaluation tools and supporting evidence-based review practices, authorizers can improve decision-making and help ensure that approved schools are positioned for long-term success.

Table of Contents

Project Description

WestEd, the Florida Association of Charter School Authorizers (FACSA), and the MIT Blueprint Research Collaborative partnered on a pilot study focused on understanding the relationship between authorizer evaluations of Florida charter school applications and school quality, closure, and renewal outcomes.

Florida employs a standardized evaluation instrument (Form IEPC-M2) for charter school applications that is published on the Florida Department of Education's (FLDOE's) website. This tool sets benchmarks for district authorizers to use to determine whether a charter school applicant's academic, organizational, and financial plans meet the state standards. This evaluation instrument has been in use since 2009, when State Board of Education Rule 6A-6.0786 was proposed to "reference and implement the model application form, evaluation instrument, charter format, and charter renewal format developed by the Department." This evaluation instrument has been updated since, most notably in 2019 when sections were added. Currently, the instrument contains 22 sections, all of which are rated on a scale of "Meets the Standard," "Partially Meets the Standard," and "Does Not Meet the Standard." The Florida Charter School Application Evaluation Instrument includes definitions for these ratings and encourages authorizers to align their application review process with the Florida Principles and Standards for Quality Authorizing.

The universal use of a single evaluation instrument created an unprecedented opportunity to study how different components of the charter application are evaluated and how these evaluations predict the long-term success of charter schools. Charter authorization is a high-stakes decision point, yet there is currently limited evidence on whether application-stage evaluations reliably identify proposals that will lead to high-performing, sustainable schools. The existing research focuses more on identifying distress signals after schools are already operational.[1] Florida also has one of the largest charter school sectors in the United States, with approximately 14 percent of all public school students enrolled in charter schools.

[1] National Charter School Resource Center. (2021). *Identifying indicators of distress in charter schools: Tools to support authorizer data collection*. Manhattan Strategy Group.

Primary Research Questions

This study was designed to examine whether information collected about charter applications during the review process can help charter school authorizers identify schools that are likely to be successful over the long term. Although charter application reviews serve a critical gatekeeping function within the authorization process, relatively little empirical evidence exists regarding whether application ratings are associated with subsequent school outcomes. Drawing on 10 years of charter application evaluation data from Florida, this study examined the relationship between ratings at the school application stage and long-term success.

Research Question 1: To what extent do composite scores on the Florida evaluation instrument predict the long-term success of charter schools?

The first research question focused on the predictive validity of the Florida evaluation instrument as a whole. Specifically, the study examined whether schools that received higher overall ratings from authorizers subsequently demonstrated stronger outcomes, including school accountability performance and continued operation. If overall application scores were associated with long-term outcomes, this would provide evidence that the evaluation instrument is functioning as intended by helping authorizers distinguish between stronger and weaker applications.

Research Question 2: Which elements of the evaluation instrument most significantly predict a charter school's success or nonrenewal or closure?

The second research question focused on the individual components of the evaluation instrument. While the overall score may provide a general indication of application quality, different sections of an application may vary in their relevance to long-term success. For example, operational planning, governance structures, financial management, educational program design, and community engagement may not contribute equally to a school's ability to remain viable and effective over time. Understanding which application components are most strongly associated with positive outcomes can help authorizers focus their review efforts on the areas that provide the greatest insight into future school performance.

Together, these research questions were intended to provide evidence that could inform both policy and practice. By examining the relationship between application evaluations and subsequent outcomes, the study sought to better understand whether current authorizing

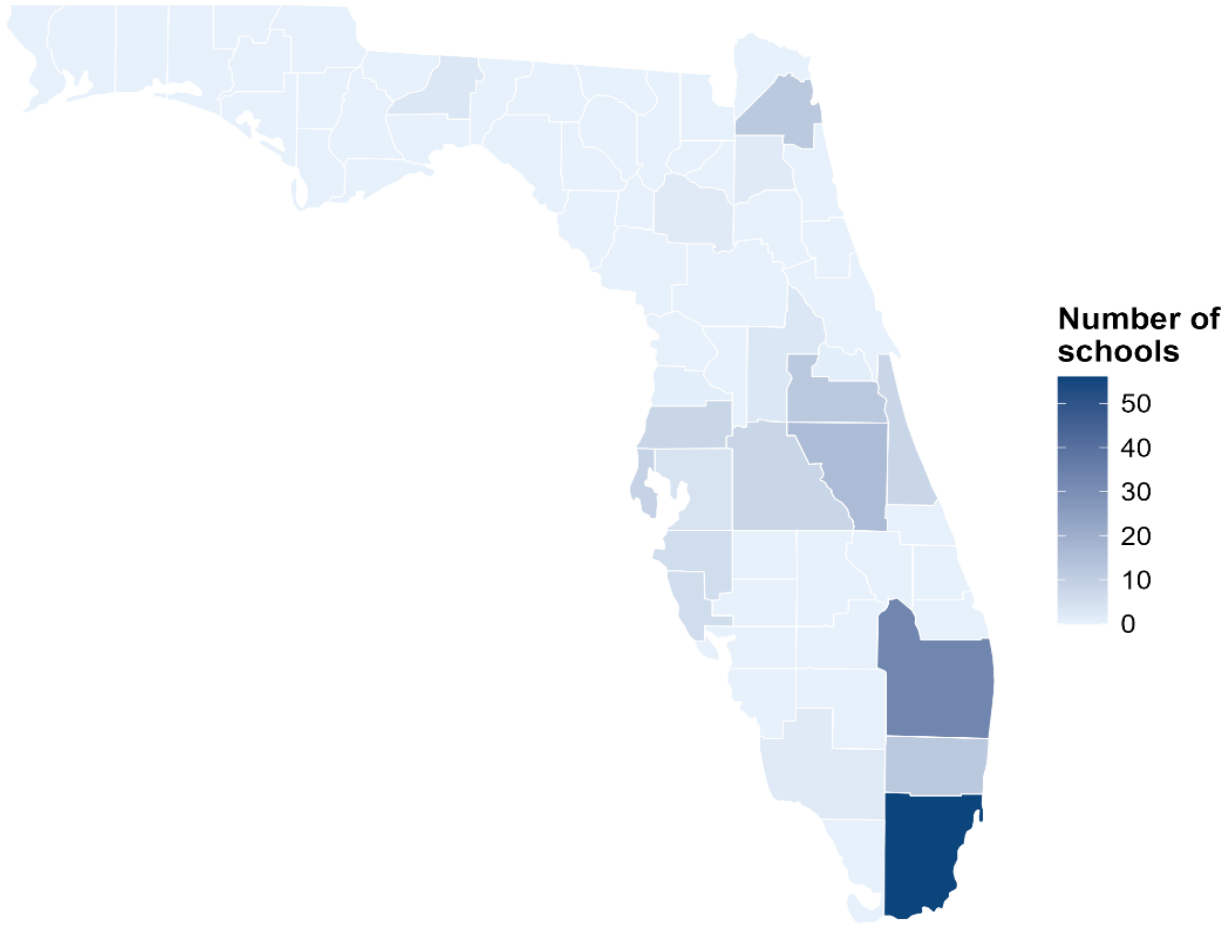
processes identify characteristics associated with successful schools and whether revisions to the evaluation instrument could improve its effectiveness as a decision-making tool. The findings are particularly relevant as Florida considers updates to its charter school evaluation instrument and as authorizers nationwide seek to strengthen evidence-based authorization practices.

Sample and Research Methods Overview

Authorizer evaluation instrument data were requested from all district authorizers in Florida via email. Additional evaluation instruments were compiled from searching BoardDocs meeting notes. The sample itself was limited by districts' responsiveness and record keeping. Ultimately, the pilot study compiled a data set of 209 approved evaluation instruments across 22 Florida school districts that opened between 2013 and 2023. Figure 1 displays the number of schools in the sample by district.

The evaluation rubrics were quantified by providing 1 point for "Meets the standard" rating and 0.5 point for "Partially meets the standard" ratings. Then the overall percentage of points was calculated based on the number of items the school was graded on to account for additional items added to the evaluation rubric within the past decade.

Figure 1. Map of Schools in Sample by District



Note. For a data table of the information included in Figure 1, see Appendix A.

WestEd researchers merged evaluation data with additional school-level information; the full list of variables and sources is provided in Appendix B. The team analyzed evaluation instrument data using descriptive statistics of each item by school quality grade and operational status, exploratory factor analysis (EFA) to explore the factor structure for alignment with the sections of the rubric, and multilevel regression and logistic analysis to determine the extent to which evaluation instrument scores predict school quality scores and operational status.

Table 1 presents descriptive characteristics of the full sample and by school operational status. The sample includes 209 Florida charter schools across 22 districts, of which 164 (78.5%) had remained open as of 2024 and 45 (21.5%) had closed. Closed schools tended to have been smaller, to have operated for fewer years, and to have served higher proportions of Black students compared to open schools. Open and closed schools were similar in their overall rubric scores, though open schools scored modestly higher on average (86.6% versus 83.9%).

Table 1. Sample Descriptives

Characteristic	Full Sample (*N* = 209)	Open Schools (*N* = 164)	Closed Schools (*N* = 45)
School characteristics: *N* (%)	209 (100%)	164 (78.5%)	45 (21.5%)
School characteristics: Years in operation, mean (SD)	4.7 (4.4)	5.5 (4.5)	1.6 (1.8)
Locale: % City	37.9%	34.8%	48.9%
Locale: % Suburb	54.9%	56.5%	48.9%
Locale: % Rural	7.3%	8.7%	2.2%
Evaluation instrument: Overall rubric score, mean (SD)	86.0 (12.3)	86.6 (12.0)	83.9 (13.3)
Evaluation instrument: % scoring 80% or above	71.8%	74.4%	62.2%
Evaluation instrument: % scoring 90% or above	46.4%	48.2%	40.0%
Academic outcomes: % earning Grade A (2025)	52.1%	52.1%	—
Academic outcomes: % earning Grade A or B (2025)	67.7%	67.7%	—
Academic outcomes: Mean accountability score, mean (SD)	63.5 (11.9)	63.5 (11.9)	—
Student demographics: Enrollment, mean (SD)	497 (393)	547 (402)	228 (175)
Student demographics: % FRPL eligible, mean (SD)	54.3 (28.9)	55.9 (28.4)	46.4 (30.7)
Student demographics: % Black students, mean (SD)	21.7 (26.2)	19.1 (23.5)	35.6 (35.0)

Characteristic	Full Sample (*N* = 209)	Open Schools (*N* = 164)	Closed Schools (*N* = 45)
Student demographics: % Hispanic students, mean (SD)	47.0 (30.3)	50.0 (30.1)	30.2 (25.6)
Student demographics: % White students, mean (SD)	25.3 (22.6)	25.1 (22.4)	26.2 (23.7)

Note. Academic outcomes reflect 2024–25 Florida school grades and accountability scores. Closed schools did not receive 2025 grades and are excluded from these rows. Demographics drawn from Common Core of Data records for the most recent year each school was observed. SD = standard deviation. FRPL = free or reduced-price lunch program.

Findings

Use of the Evaluation Instrument

Before examining whether rubric scores predicted school outcomes, the project team first sought to understand how the instrument is used in practice across districts—both to contextualize the study's analyses and to ground the interpretation of the findings. The review found that although districts followed policy and scored applications using the rubric, how they applied it varied considerably. Some districts used multiple reviewers; others relied on a single rater. Some applied strict standards of evidence; others appeared more lenient, resulting in an average 32 percentage point gap between the lowest and highest scoring districts. With this variation in mind, scores across districts should not be interpreted as perfectly equivalent measures of application quality.

Finding 1: Most charter applications that were approved received high scores on the evaluation instrument.

The average approved application in Florida received approximately 86 percent of available points on the evaluation instrument, with most schools meeting standards across the majority of application components. Because most schools scored so similarly, the instrument has limited ability to distinguish stronger applications from weaker ones, functioning more as a minimum threshold screener than a tool for evaluating the relative strength of an application.

Finding 2: The evaluation instrument items align with different constructs than how it is currently organized. The constructs that the instrument aligns well with are Governance and Operations, Curriculum and Design, Culture and Support.

The evaluation instrument is formally organized into three sections: an Educational Plan, an Organizational Plan, and a Business Plan. To understand whether this formal structure reflects how the items function, the research team examined which items tended to move together—that is, whether schools that scored well on one item also tended to score well on a related set of items.

What emerged was a somewhat different grouping than the instrument's current structure. Rather than mapping neatly onto Educational, Organizational, and Business planning, the items are clustered into three dimensions: Governance and Operations, Curriculum and Design, and Culture and Support. Only items available across all years of data were included; items added in more recent revisions were not part of this analysis. The full results are presented in Appendix C.

In practice, this means the rubric is capturing three relatively distinct dimensions of school readiness even if it is not currently organized that way. A reviewer evaluating an application section by section may not be assessing these dimensions holistically; an applicant's governance capacity, for example, is reflected across multiple sections rather than concentrated in one place.

Predictive Ability of the Evaluation Instrument

The research team next examined whether scores on the evaluation instrument were associated with two key outcomes: school accountability scores and school closures. To do this, the analysis used regression modeling that accounted for differences in school size, student demographics, and location, ensuring that any relationships that were found were not simply reflections of which types of schools tended to score higher on the rubric. The analysis examined outcomes at three levels: the overall rubric score, the three constructs identified through the factor analysis (Governance and Operations, Curriculum and Design, and Culture and Support), and individual rubric items.

Finding 3: There is a limited relationship between application scores and accountability grades.

Florida assigns each operating charter school an annual accountability score—both a numerical score and a letter grade—based on student performance. The study examined whether scores on the evaluation instrument at the time of application were associated with either measure in subsequent years. Overall, the analysis found little evidence that rubric scores predict how well a school will perform on Florida's accountability system. This was true both for overall points earned and for likelihood of earning a Grade A.

Florida's accountability grades are heavily influenced by student demographics, particularly rates of economic disadvantage, making them a difficult outcome for any application-based measure to predict. Additionally, because this study's sample consists only of schools that were approved, the rubric may have already screened out the weakest applicants, leaving limited variation to be detected.

Finding 4: Overall rubric scores significantly predict closure—lower scores were associated with a higher likelihood of closure.

The examination of whether rubric scores were associated with school closure found a meaningful relationship: After differences in school size, student demographics, and location had been accounted for, schools that scored lower on the evaluation instrument at the time of application were substantially more likely to close in subsequent years.

Schools that scored higher on operationally focused items—governance, human resources, English Language Learner planning, and startup planning—were significantly less likely to close.

When the study looked beyond the overall rubric score to examine individual items, a consistent pattern emerged: It was not the educational program items that predicted whether a school would close but rather the operational ones. Schools that demonstrated stronger planning around how they would govern themselves, hire and manage staff, serve English Language Learners, and execute their startup plan were significantly more likely to remain open.

This suggests that what distinguishes schools that survive from those that close may have less to do with the strength of their educational vision and more to do with whether they had a concrete, realistic plan for how they would actually operate.

These item-level associations should be interpreted with caution: When a Benjamini-Hochberg correction for multiple comparisons is applied across the 22 items tested, none of the four items reported above as significant (governance, human resources, English Language Learner planning, and startup planning) remain significant at a false discovery rate of .05. The pattern is suggestive and worth pursuing further, but on this sample it should be treated as exploratory rather than confirmed.

Enrollment is the strongest predictor of closure.

Of all the factors examined, school size was the most powerful predictor of whether a school closed. Larger schools were substantially less likely to close, and this was consistent across all the analyses. What drives this relationship is not fully clear from the study's data. It may reflect the financial vulnerability of smaller schools, the challenge of attracting and retaining enough students to remain viable, or both.

Finding 5: Districts apply the evaluation instrument in different ways, but the differences in rater severity were not predictive of accountability grades and school closure.

Districts varied considerably in how they scored applications, as described earlier in this report. The study examined whether this variation mattered for outcomes—specifically, whether schools in districts that scored more strictly tended to perform better academically or were less likely to close. Districts that applied higher standards in their scoring did not produce schools with better accountability grades, and stricter scoring districts did not have lower rates of school closure.

Policy Relevance

This study provides timely practical evidence that can inform charter school authorizing practices in Florida specific to charter school application reviews, and it has potential impacts for charter authorizers across the country. As one of the first studies to examine how ratings on a charter school application evaluation instrument relate to long-term school outcomes, the findings contribute to a limited but growing body of research on the role of authorizer practices in shaping charter school quality and sustainability.

Carried out by a research–practice partnership between WestEd and the FACSA, this study was intentionally designed to generate findings that could be translated directly into policy and practice. The research is particularly relevant given Florida's ongoing efforts to review and revise its charter school application evaluation instrument. The FACSA recommends that the FLDOE use the findings from this study to inform those revisions, providing an opportunity to incorporate empirical evidence into future versions of the state's review process.

Several other findings have direct implications for the design and use of the evaluation instrument. Although overall application scores demonstrated limited relationships with subsequent school accountability grades, they were significantly associated with school closure outcomes. More importantly, a small number of operationally focused application components, including governance, human resources planning, English Language Learner planning, and startup readiness, showed the strongest relationships with long-term school survival. These findings suggest that authorizers may benefit from placing greater emphasis on the operational and organizational capacity of founding teams during the application review process. The results also raise important questions about whether the current equal-weighting approach used across application components reflects the relative importance of different areas of school readiness. These findings could impact changes to the evaluation instrument itself or to how authorizers use the instrument and/or to how they conduct follow-up capacity interviews of applicants, knowing that these areas may merit greater attention or weight.

The study further found that the evaluation instrument appears to measure three underlying dimensions of readiness—Governance and Operations, Curriculum and Design, and Culture and Support—rather than the Educational, Organizational, and Business Plan structure that currently organizes the rubric. As Florida considers future revisions to the instrument, these findings offer an opportunity to examine whether the rubric's organization, scoring structure, and reviewer guidance could better align with how application components function in practice.

The results also suggest opportunities to improve consistency in application reviews across districts. While Florida benefits from having a common statewide evaluation instrument,

authorizers varied considerably in how they applied the rubric, resulting in substantial differences in average scores across districts. Although districts with stricter scoring practices did not produce schools with better outcomes, these findings highlight the potential value of additional calibration activities, reviewer training, and shared guidance to promote more consistent interpretation of rubric standards across authorizers. This training could be offered by state organizations such as the FACSA or the FLDOE or by national organizations such as WestEd or the National Association for Charter School Authorizers.

Beyond Florida, the findings have potential relevance for charter authorizers nationwide. The Florida evaluation instrument is one of the few publicly available and consistently used charter application review tools in the country. Through the FACSA's participation in the National Network for District Authorizing (NN4DA), lessons learned from this study can be shared with authorizers in states where districts play a primary authorizing role, including California, Colorado, Georgia, Maryland, Oklahoma, Pennsylvania, Tennessee, and Wisconsin. Because many authorizers across the country rely on application review processes that address similar domains of academic, organizational, and financial planning, the findings may inform broader discussions about evidence-based charter authorization practices.

This study also complements other Blueprint-funded research focused on authorizer accountability and oversight after schools begin operation. Together, these efforts contribute to a more comprehensive understanding of the charter school accountability continuum, from application review and approval decisions to ongoing monitoring, performance management, and renewal. By generating empirical evidence regarding both the front-end and back-end components of charter school accountability systems, this body of work supports the development of more effective, data-informed authorizing practices.

Finally, the findings suggest that application review processes may be streamlined and refined over time. If future research continues to identify application components that are weakly related to school outcomes, authorizers and policymakers may be able to reduce unnecessary burden on applicants and reviewers by simplifying portions of the application process while focusing greater attention on areas that are most predictive of long-term success. Ultimately, the findings from this study provide a foundation for improving charter authorization practices and supporting the development of high-quality, sustainable charter schools from approval through renewal.

Next Steps

This pilot study was designed to identify where relationships between application evaluations and school outcomes appear to exist, not to confirm which application components are causally predictive of long-term success. The patterns identified here are informative, but at this stage they should be considered preliminary rather than definitive. Triangulating these quantitative findings with qualitative research, including authorizer focus groups and interviews, and testing whether the patterns replicate in additional cohorts or states are necessary before drawing firm conclusions about what drives school outcomes. At this time, the project team is not recommending specific revisions to the evaluation instrument; the next steps outlined below suggest the additional research and validation needed to move from these exploratory findings toward evidence-based recommendations.

This pilot study identified several application components, including governance, human resources planning, English Language Learner planning, and startup readiness, that showed a stronger relationship with long-term school sustainability than other components did. As Florida continues to review and refine its authorization processes, these patterns suggest that the current organization and weighting of the rubric may be worth revisiting to better reflect the dimensions of school readiness that appear most relevant to future success. Given the exploratory nature of this pilot, this suggestion should be considered a direction for further investigation rather than a confirmed set of changes to implement. The next steps below outline the additional validation needed before specific revisions can be recommended with confidence.

Strengthen Interrater Reliability

The study identified substantial differences in scoring practices across districts despite the use of a common statewide evaluation instrument. These findings highlight an opportunity to improve consistency in how applications are reviewed and scored. The FACSA plans to use the results of this study to support professional learning opportunities focused on reviewer calibration and interrater reliability. These efforts may include training sessions, scoring exercises using sample applications, and structured discussions among reviewers to establish a more consistent interpretation of rubric standards. Increasing consistency in scoring practices should strengthen confidence that evaluation scores reflect differences in application quality rather than differences in local review practices.

Engage Practitioners in Interpreting and Applying Findings

A key next step is sharing the results with charter school authorizers and other interest holders throughout Florida. Presentations and facilitated discussions will allow practitioners to reflect on the findings and consider how they align with their own experiences reviewing applications and overseeing charter schools. Authorizer feedback will help identify practical implications of the research and inform future revisions to both the evaluation instrument and reviewer guidance. These conversations will help ensure that improvements to charter authorization practices are grounded in both empirical evidence and practitioner expertise.

Continue Research by Exploring Alternative Approaches for Scoring and Weighting and by Conducting Authorizer Focus Groups

While the quantitative analyses identified several relationships between application ratings and school outcomes, they do not fully explain why those relationships exist. A logical next phase of this work is to conduct focus groups and interviews with charter school authorizers across Florida. These discussions would provide valuable insight into how reviewers interpret rubric criteria, what characteristics they associate with successful schools, and how application review processes vary across districts.

This study suggests that some application components may provide stronger signals about future school sustainability than others. Future analyses should examine whether alternative scoring or weighting approaches improve the evaluation instrument's ability to identify schools that are prepared for long-term success. Such work could explore whether greater emphasis on operational readiness and organizational capacity increases predictive validity while preserving fairness, transparency, and ease of implementation for both applicants and authorizers.

Future research should also explore additional outcomes, including enrollment stability, renewal decisions, financial performance, student growth measures, and other indicators of organizational health. Additional analyses may also examine whether the findings identified in Florida are replicated in future cohorts of charter schools or in other states that use similar authorizing practices. Expanding the evidence base will help determine which aspects of charter application reviews are most strongly associated with long-term school effectiveness and sustainability.

Disseminate Findings Nationally

The project team will continue sharing findings with interest holders in Florida and across the nation. Results will be presented to the FLDOE, FACSA members, and other state education leaders as part of ongoing efforts to strengthen charter school authorization practices. Through

the FACSA's participation in the NN4DA, lessons learned from this study can also inform conversations in other states where district authorizers play a prominent role. Broad dissemination of the findings can contribute to the growing body of evidence on effective charter school authorization and support efforts to develop more data-informed approaches to application review, oversight, and accountability.

Appendix A. Data Table for Figure 1

District	Number of schools
Miami-Dade	56
Palm Beach	33
Osceola	16
Broward	12
Duval	12
Orange	12
Pinellas	9
Brevard	8
Pasco	8
Polk	8
Manatee	6
Sarasota	6
Hillsborough	4
St. Lucie	4
Lake	3
Leon	3
Alachua	2
Clay	2
Collier	2
Hernando	1
Seminole	1
Tallahassee Collegiate Academy	1

Appendix B. Variables and Data Sources

Table B1. Evaluation Instrument Variables and Data Sources

Variable name	Description	Source
Overall rubric score	Percentage of available points earned on the evaluation instrument, calculated as total points earned divided by total items scored, multiplied by 100; accounts for variation in the number of items scored across rubric versions	Florida Charter School Application Evaluation Instrument
Individual item scores	Score on each of the 22 evaluation instrument items, scored as: Meets the Standard (1), Partially Meets the Standard (0.5), or Does Not Meet the Standard (0); items added in later rubric revisions have higher rates of missing data	Florida Charter School Application Evaluation Instrument
Factor 1: Governance and Operations	Mean score across five items identified through EFA as loading on the Governance and Operations factor: Budget (Item 20), Management and Staffing (Item 11), Exceptional Students (Item 6), Financial Management and Oversight (Item 21), and Governance (Item 10)	Derived from evaluation data
Factor 2: Curriculum and Design	Mean score across three items identified through EFA as loading on the Curriculum and Design factor: Curriculum Plan (Item 4), Educational Program Design (Item 3), and Mission and Guiding Principles (Item 1)	Derived from evaluation data
Factor 3: Culture and Support	Mean score across three items identified through EFA as loading on the Culture and Support factor: School Culture and Discipline (Item 8), Student Recruitment and Enrollment (Item 14), and Facilities (Item 16)	Derived from evaluation data
Relative rubric score	Each school's overall rubric score minus the mean rubric score of all schools in the same district, capturing how a school's score compares to others reviewed by the same authorizer, net of district-level scoring tendencies	Derived from evaluation data

Table B2. School Outcome Variables and Data Sources

Variable name	Description	Source
School closure	Binary indicator of whether a school closed prior to 2024, defined as a school's last appearance in the CCD occurring before the 2023–24 school year, with open schools defined as those still appearing in the CCD in 2023–24	CCD, National Center for Education Statistics
Florida accountability score	Percentage of total possible accountability points earned under Florida's school grading system, reflecting student achievement, learning gains, graduation rates, and other performance indicators	Florida Department of Education (FLDOE) School Grades
Florida school grade	Annual letter grade (A–F) assigned to each school by the FLDOE based on the accountability score, with schools earning insufficient data assigned an "I" designation and excluded from analyses	FLDOE School Grades

Note. CCD = Common Core of Data.

Table B3. School Characteristics Variables and Data Sources

Variable name	Description	Source
School district	Name of the Florida school district in which the school is located, used as the grouping variable for district-level random effects in all mixed effects models	District authorizers
First year in CCD	First year the school appeared in the CCD, used as the proxy for school opening year	CCD, National Center for Education Statistics
Last year in CCD	Last year the school appeared in the CCD, representing the final year of operation for closed schools and 2023–24 for open schools	CCD, National Center for Education Statistics
Years in operation	Number of years between a school's first and last appearance in the CCD, reflecting observed operational lifespan	Derived from CCD

Variable name	Description	Source
Locale	School urbanicity category—City, Suburb, Rural, or Town—derived from the NCES Urban-Centric Locale classification system, with Suburb as the reference category in all regression models; open school locale drawn from FLDOE administrative records; closed school locale drawn from CCD records	CCD (closed schools); FLDOE (open schools)

Note. CCD = Common Core of Data. For each closed school, demographic characteristics were drawn from the most recent year in which the school appeared in the CCD with valid data, which may differ from the year of closure. Open school locale data were drawn from FLDOE administrative records.

Table B4. Student Demographics Variables and Data Sources

Variable name	Description	Source
Total enrollment	Total number of students enrolled, log-transformed in regression models due to right skew in the distribution, drawn from the most recent year with valid data for each school	CCD
% FRPL eligible	Percentage of students eligible for free or reduced-price lunch, used as a proxy for school-level economic disadvantage, drawn from the most recent year with valid data for each school	CCD
% White students	Percentage of enrolled students identified as White (non-Hispanic), included as a demographic control to account for racial composition differences between open and closed schools	CCD
% Black students	Percentage of enrolled students identified as Black or African American, reported descriptively to characterize demographic differences between open and closed schools	CCD
% Hispanic students	Percentage of enrolled students identified as Hispanic or Latino, reported descriptively to characterize demographic differences between open and closed schools	CCD
% Asian students	Percentage of enrolled students identified as Asian, reported descriptively in sample characteristics	CCD

Note. CCD = Common Core of Data. For each closed school, demographic characteristics were drawn from the most recent year in which the school appeared in the CCD with valid data, which may differ from the year of closure. All CCD demographic variables are expressed as proportions in the analysis and converted to percentages for reporting. FRPL = free or reduced-price lunch program.

Appendix C. Technical Methodology and Statistical Results

Exploratory Factor Analysis

To examine the underlying structure of the evaluation instrument, the project team conducted an exploratory factor analysis (EFA) on the subset of items available across all years of data (*n* = 17 items; *n* = 131 schools with complete item-level data). Three items were excluded due to high rates of missing data reflecting their addition in later rubric revisions: Supplemental Programming (72% missing), Professional Development (68% missing), and Parent and Community Involvement (66% missing).

The EFA was conducted using principal factor extraction with promax rotation, which allows for correlated factors—an appropriate choice given that dimensions of school readiness are likely related to one another. Three factors were retained based on the factor loadings, scree plot, and interpretability of the factor solution.

Table C1. Factor 2: Curriculum and Design: Exploratory Factor Analysis—Item Factor Loadings and Assignments

Item	Factor Loading	Highest Loading	Item-Rest Correlation
Mission and Guiding Principles	0.396	F2: Curriculum and Design	0.253
Target Population and Student Body	0.319	F2: Curriculum and Design	0.357
Educational Program Design	0.709	F2: Curriculum and Design	0.53
Curriculum Plan	0.768	F2: Curriculum and Design	0.59

Item	Factor Loading	Highest Loading	Item-Rest Correlation
Student Performance Assessment	0.366	F2: Curriculum and Design	0.467

Table C2. Factor 1: Governance and Operations: Exploratory Factor Analysis—Item Factor Loadings and Assignments

Item	Factor Loading	Highest Loading	Item-Rest Correlation
Exceptional Students	0.539	F1: Governance and Operations	0.623
English Language Learners	0.323	F1: Governance and Operations	0.494
Governance	0.465	F1: Governance and Operations	0.582
Management and Staffing	0.589	F1: Governance and Operations	0.673
Budget	0.711	F1: Governance and Operations	0.643
Transportation	0.351	F1: Governance and Operations	0.429
Financial Management and Oversight	0.472	F1: Governance and Operations	0.322
Startup Plan	0.215	F1: Governance and Operations	0.319

Table C3. Factor 3: Culture and Support: Exploratory Factor Analysis: Item Factor Loadings and Assignments

Item	Factor Loading	Highest Loading	Item-Rest Correlation
School Culture and Discipline	0.608	F3: Culture and Support	0.5
Student Recruitment and Enrollment	0.587	F3: Culture and Support	0.349
Facilities	0.495	F3: Culture and Support	0.39
Food Service	0.278	F3: Culture and Support	0.308

Table C4. Descriptive Statistics for Evaluation Instrument Items 1–9: Educational Plan

Item	*N*	Mean	Standard Deviation	Does Not Meet Standard (%)	Partially Meets Standard (%)	Meets Standard (%)
Mission and Guiding Principles	209	0.952	0.163	1.0%	7.7%	91.4%
Target Population and Student Body	207	0.865	0.233	1.0%	25.1%	73.9%
Educational Program Design	207	0.862	0.250	2.4%	22.7%	74.9%
Curriculum Plan	208	0.817	0.261	1.9%	32.7%	65.4%
Student Performance Assessment	209	0.801	0.273	2.9%	34.0%	63.2%
Exceptional Students	205	0.790	0.284	3.9%	34.1%	62.0%
English Language Learners	208	0.779	0.321	8.2%	27.9%	63.9%
School Culture and Discipline	206	0.898	0.214	1.0%	18.4%	80.6%

Item	*N*	Mean	Standard Deviation	Does Not Meet Standard (%)	Partially Meets Standard (%)	Meets Standard (%)
Supplemental Programming	59	0.932	0.173	0.0%	13.6%	86.4%

Note. Full sample *N* = 209. Items with *n* < 100 were added in later rubric revisions and have substantially higher rates of missing data. These items were excluded from the EFA. Mean scores range from 0 to 1 (0 = Does not meet, 0.5 = Partially meets, 1 = Meets standard). SD = standard deviation.

Table C5. Descriptive Statistics for Evaluation Instrument Items 10–15: Organizational Plan

Item	*N*	Mean	Standard Deviation	Does Not Meet Standard (%)	Partially Meets Standard (%)	Meets Standard (%)
Governance	207	0.874	0.263	4.3%	16.4%	79.2%
Management and Staffing	207	0.845	0.288	5.8%	19.3%	74.9%
Human Resources and Employment	162	0.858	0.246	1.9%	24.7%	73.5%
Professional Development	68	0.897	0.238	2.9%	14.7%	82.4%
Student Recruitment and Enrollment	205	0.898	0.220	1.5%	17.6%	81.0%
Parent and Community Involvement	72	0.924	0.200	1.4%	12.5%	86.1%

Note. Full sample *N* = 209. Items with *n* < 100 were added in later rubric revisions and have substantially higher rates of missing data. These items were excluded from the EFA. Mean scores range from 0 to 1 (0 = Does not meet, 0.5 = Partially meets, 1 = Meets standard). SD = standard deviation.

Table C6. Descriptive Statistics for Evaluation Instrument Items 16–22: Business Plan

Item	*N*	Mean	Standard Deviation	Does Not Meet Standard (%)	Partially Meets Standard (%)	Meets Standard (%)
Facilities	197	0.878	0.238	2.0%	20.3%	77.7%
Transportation	195	0.856	0.268	4.1%	20.5%	75.4%
Food Service	194	0.938	0.187	1.5%	9.3%	89.2%
School Safety and Security	124	0.891	0.226	1.6%	18.5%	79.8%
Budget	197	0.726	0.344	11.2%	32.5%	56.3%
Financial Management and Oversight	198	0.891	0.224	1.5%	18.7%	79.8%
Startup Plan	195	0.892	0.251	4.1%	13.3%	82.6%

Note. Full sample $N = 209$. Items with $n < 100$ were added in later rubric revisions and have substantially higher rates of missing data. These items were excluded from the EFA. Mean scores range from 0 to 1 (0 = Does not meet, 0.5 = Partially meets, 1 = Meets standard). SD = standard deviation.

Mixed Effects Regression Models

All inferential analyses used mixed effects regression models with a random intercept for school district, accounting for the nested structure of schools within districts. The intraclass correlation (ICC) from the null model indicated that approximately 7 percent of variance in closure was attributable to district-level differences, justifying the use of multilevel modeling. Each outcome was examined at three levels: the overall rubric score, the three EFA-derived factor scores (Governance and Operations, Curriculum and Design, and Culture and Support), and individual rubric items. Models were estimated with demographic controls, which included the percentage of students eligible for free or reduced-price lunch (FRPL), the percentage of White students, log-transformed enrollment, and locale (City, Rural, or Town, with Suburb as the reference category). Enrollment was log-transformed due to right skew in the distribution. All continuous predictors were standardized prior to modeling (mean = standard deviation [SD] = 1) to allow comparability of effect sizes across items and constructs. The analytic sample for models including demographic controls was 174 schools, reflecting listwise deletion of cases missing data on any control variable.

School Accountability Outcomes

To examine whether rubric scores were associated with Florida school accountability outcomes, the research team estimated mixed effects linear regression models for overall accountability score (percentage of total points possible) and mixed effects logistic regression models for Grade A status. Each outcome was modeled with the overall rubric score and, separately, with the three EFA-derived factor scores, along with the same demographic controls used in the closure models (percentage FRPL eligible, percentage White students, log enrollment, and locale). The analytic sample for these models was 174 schools, matching the closure analysis. Results from logistic models are reported as odds ratios; odds ratios above 1.0 indicate higher odds of earning a Grade A associated with higher scores on that predictor.

Table C7. Mixed Effects Models: Overall Rubric Score and Exploratory Factor Analysis Factors Predicting School Accountability Outcomes: % of Total Points Possible (standardized regression coefficient) for Model 1

Predictor	Estimate (OR for Grade A)	Standard Error	95% CI Lower	95% CI Upper	*p*
Overall rubric score (standardized)	0.61	0.947	-1.263	2.484	.520
% FRPL eligible (standardized)	-5.765	1.062	-7.867	-3.662	< .001*
% White students (standardized)	-0.121	1.221	-2.566	2.323	.921
Log enrollment (standardized)	0.635	0.945	-1.233	2.503	.503
Locale: City (ref = Suburb)	-2.916	1.88	-6.635	0.803	.123
Locale: Rural (ref = Suburb)	-1.528	3.192	-7.839	4.784	.633

Note. OR = odds ratio for the Grade A models; unstandardized regression coefficients for the % Total Points Possible models. CI = Confidence Interval. All continuous predictors standardized (mean = 0, SD = 1). Reference category for locale = Suburb. *N* = 174 schools. Random intercept by school district included in all models. FRPL = free or reduced-price lunch program.

* = $p < .05$.

Table C8. Mixed Effects Models: Overall Rubric Score and Exploratory Factor Analysis Factors Predicting School Accountability Outcomes: Grade A Versus Not Grade A (odds ratio) for Model 1

Predictor	Estimate (OR for Grade A)	Standard Error	95% CI Lower	95% CI Upper	*p*
Overall rubric score (OR)	1.234	0.259	0.817	1.862	.317
% FRPL eligible (OR)	0.313	0.085	0.184	0.531	< .001*
% White students (OR)	0.756	0.233	0.413	1.385	.365
Log enrollment (OR)	1.034	0.226	0.674	1.586	.878
Locale: City (ref = Suburb)	0.677	0.284	0.298	1.539	.352
Locale: Rural (ref = Suburb)	0.324	0.228	0.081	1.286	.109

Note. OR = odds ratio for the Grade A models; unstandardized regression coefficients for the % Total Points Possible models. CI = Confidence Interval. All continuous predictors standardized (mean = 0, SD = 1). Reference category for locale = Suburb. *N* = 174 schools. Random intercept by school district included in all models. FRPL = free or reduced-price lunch program.

* = $p < .05$.

School Closure

School closure was modeled as a binary outcome (1 = closed, 0 = open), defined as a school's last appearance in Common Core of Data (CCD) records prior to 2024, using mixed effects logistic regression. Results are reported as odds ratios; odds ratios below 1.0 indicate lower odds of closure associated with higher scores on that predictor. Item-level models were run separately for each of the 22 rubric items and should be treated as exploratory given the number of tests conducted.

Table C9. Mixed Effects Logistic Regression: Model 1 Overall Rubric Score

Predictor	OR	Standard Error	95% CI Lower	95% CI Upper	*p*
Overall rubric score (standardized)	0.506	0.136	0.299	0.858	.011*
% FRPL eligible (standardized)	0.586	0.176	0.325	1.055	.075
% White students (standardized)	0.655	0.233	0.326	1.313	.233
Log enrollment (standardized)	0.349	0.087	0.214	0.57	< .001*
Locale: City (ref = Suburb)	1.562	0.841	0.544	4.488	.407
Locale: Rural (ref = Suburb)	0.268	0.318	0.026	2.745	.267

Note. OR = odds ratio. All continuous predictors standardized (mean = 0, SD = 1). CI = Confidence Interval. Reference category for locale = Suburb. *N* = 174 schools with complete data on all controls. Random intercept by school district included in all models. FRPL = free or reduced-price lunch program.

* = $p < .05$.

Table C10. Mixed Effects Logistic Regression: Model 2 Exploratory Factor Analysis Factors Predicting School Closure

Predictor	OR	Standard Error	95% CI Lower	95% CI Upper	*p*
Factor 1: Governance and Operations	0.571	0.17	0.319	1.022	.059
Factor 2: Curriculum and Design	1.001	0.273	0.586	1.71	.997
Factor 3: Culture and Support	0.853	0.25	0.48	1.516	.587
% FRPL eligible (standardized)	0.577	0.176	0.317	1.05	.072
% White students (standardized)	0.721	0.248	0.367	1.417	.343

Predictor	OR	Standard Error	95% CI Lower	95% CI Upper	*p*
Log enrollment (standardized)	0.351	0.087	0.216	0.569	< .001*
Locale: City (ref = Suburb)	1.607	0.859	0.564	4.582	.375
Locale: Rural (ref = Suburb)	0.223	0.275	0.02	2.512	.224

Note. OR = odds ratio. All continuous predictors standardized (mean = 0, SD = 1). CI = Confidence Interval. Reference category for locale = Suburb. *N* = 174 schools with complete data on all controls. Random intercept by school district included in all models. FRPL = free or reduced-price lunch program.

* = $p < .05$.

Table C11. Item-Level Mixed Effects Logistic Regression: Predicting School Closure (With Demographic Controls)

Item	OR	Standard Error	95% CI Lower	95% CI Upper	Unadjusted *p*-Value	B-H Adjusted *p*-Value
English Language Learners	0.563	0.135	0.352	0.902	.017*	.187
Governance	0.567	0.138	0.352	0.914	.020*	.187
Human Resources and Employment	0.543	0.149	0.317	0.931	.026*	.187
Startup Plan	0.639	0.135	0.423	0.966	.034*	.187
Financial Management and Oversight	0.648	0.162	0.397	1.058	.083	.348
Facilities	0.681	0.156	0.435	1.069	.095	.348
Budget	0.719	0.169	0.454	1.138	.159	.389
Student Recruitment and Enrollment	0.712	0.186	0.426	1.189	.194	.389
Curriculum Plan	0.733	0.176	0.458	1.175	.198	.389

Item	OR	Standard Error	95% CI Lower	95% CI Upper	Unadjusted *p*-Value	B-H Adjusted *p*-Value
Transportation	0.730	0.179	0.452	1.181	.200	.389
Management and Staffing	0.704	0.195	0.409	1.212	.205	.389
Exceptional Students	0.723	0.188	0.434	1.203	.212	.389
Student Performance Assessment	0.779	0.187	0.487	1.247	.298	.488
Parent and Community Involvement	0.548	0.336	0.165	1.821	.327	.488
Supplemental Programming	0.547	0.346	0.158	1.893	.341	.488
Food Service	1.476	0.631	0.638	3.412	.363	.488
Target Population and Student Body	0.813	0.190	0.514	1.286	.377	.488
School Safety and Security	1.43	0.738	0.520	3.934	.489	.581
School Culture and Discipline	1.214	0.350	0.689	2.138	.502	.581
Professional Development	0.778	0.361	0.313	1.933	.589	.648
Mission and Guiding Principles	0.934	0.201	0.612	1.425	.751	.814
Educational Program Design	0.940	0.248	0.561	1.575	.814	.814

Note. OR = odds ratio. CI = Confidence Interval. B-H Adjusted *p*-Value = Benjamini-Hochberg false discovery rate-adjusted *p*-value. Because 22 separate models were estimated, each testing a different rubric item, unadjusted *p*-values may overstate the evidence for any single item. The B-H procedure controls the expected proportion of false discoveries across the full set of tests. No items met the adjusted significance threshold of $p < .05$.

* = $p < .05$ (unadjusted).

District Scoring Stringency

To examine whether district-level scoring variation was associated with outcomes, the research team constructed a school-level measure of relative rubric score defined as each school's score minus its district's mean score. This measure was modeled as a predictor of both closure and accountability outcomes within the same mixed effects framework. When both the absolute and relative rubric scores were entered simultaneously, high collinearity between the two measures precluded interpretation of independent effects. Results are therefore reported for each separately and compared.

Table C12. Mixed Effects Logistic Regression: District Scoring Stringency Predicting School Closure

Predictor	OR	SE	95% CI Lower	95% CI Upper	*p*
Relative rubric score (standardized)	0.521	0.127	0.323	0.839	.007*
% FRPL eligible (standardized)	0.59	0.178	0.327	1.066	.081
% White students (standardized)	0.681	0.242	0.34	1.365	.279
Log enrollment (standardized)	0.354	0.089	0.217	0.578	< .001*
Locale: City (ref = Suburb)	1.599	0.864	0.555	4.612	.385
Locale: Rural (ref = Suburb)	0.309	0.366	0.03	3.155	.322

Note. OR = odds ratio. Relative rubric score = school score minus district mean score. CI = Confidence Interval. All continuous predictors standardized (mean = 0, SD = 1). Reference category for locale = Suburb. *N* = 174 schools with complete data. Random intercept by school district included. FRPL = free or reduced-price lunch program.

* = $p < .05$.

Limitations

Several limitations should be considered when interpreting these findings. First, the analytic sample consisted only of schools that were approved through the FACSA process, which likely excluded the weakest applicants and restricted the range of rubric scores available for analysis. This selection effect may attenuate observed associations between rubric scores and outcomes. Second, the relatively small sample size—particularly for closed schools (n = 45)—limits

statistical power, and item-level findings should be treated as exploratory. A related concern is multiple comparisons: The item-level models tested 22 rubric items against each outcome without correction, and a Benjamini-Hochberg false discovery rate adjustment (see Table C11) shows that none of the individually significant items survived correction, so item-level results should be read as hypothesis generating rather than confirmatory. Third, demographic controls were drawn from CCD records for the most recent year each school was observed rather than from the time of application, which may introduce measurement timing differences between open and closed schools.