June 2022

THE SUNSHINE PORTAL
Florida's Early Childhood Integrated Data System

Examining the Associations Between School Readiness Providers' CLASS Scores and Gold Seal Accreditation

The Early Childhood Policy Research Group (ECPRG)

Jing Huang, M.S., Phillip R. Sherlock, Ph.D., Herman T. Knopf, Ph.D.

DIVISION OF Early Learning
LEARN EARLY. LEARN FOR LIFE.

UF Anita Zucker Center for Excellence in Early Childhood Studies
UNIVERSITY of FLORIDA

FLORIDA Preschool Development Grant

The Sunshine State Early Childhood Information Portal is supported by the Division of Early Learning through the Preschool Development Grant Birth through Five Initiative (PDG B-5) Number 90TP0004-02-00 from the Office of Child Care, Administration for Children and Families, U.S. Department of Health and Human Services. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the United States Department of Health and Human Services, Administration for Children and Families.

Abstract

The primary aim of this study was to investigate the associations between the Florida Gold Seal Quality Care Program (GS) accreditation and CLASS scores. Additionally, we examined the associations between other provider-level characteristics and Classroom Assessment Scoring System (CLASS) scores, as well as statewide variability in CLASS scores across coalitions. We found CLASS scores with the following four GS accreditation agencies (APPLE (Accredited Professional Preschool Learning Environment), NACECEP (National Accreditation Commission for Early Care and Education Programs), and NAEYC (National Association for the Education of Young Children) for center-based providers and NAFCC (National Association for Family Child Care) for home-based providers) tend to have higher CLASS scores compared to non-GS providers. While GS accreditation is associated with higher CLASS scores, the magnitude of the CLASS score differences among APPLE, NACECEP and NAEYC, as measured by the effect size, is negligible. The effect size of CLASS scores for home-based providers accredited by NAFCC is small. Furthermore, the between-coalition variation has a significant and moderate-large effect on CLASS score variations across coalitions for both center-based and home-based providers. Regarding provider-level characteristics other than Gold Seal accreditation, increased SR enrollment was associated with decreases in composite CLASS scores for center-based providers.

Introduction

Recent changes resulting from state legislative action have prompted the transfer of the administration of the Florida Gold Seal Quality Care Program (GS) from the Florida Department of Children and Families to the Division of Early Learning (DEL) within the Florida Department of Education. The change in administration of the GS program provides an opportunity for the new administering agency to review the process for approving different early childhood accrediting bodies to increase the credibility of the high-quality status granted to providers qualifying as GS. This study by the Early Childhood Policy Research Group (ECPRG) at the University of Florida Anita Zucker Center for Excellence in Early Childhood Studies is an investigation of the relationship between GS accreditation and the Classroom Assessment Scoring System (CLASS) scores of providers contracted with the Florida School Readiness (SR) program. Given that both the CLASS and GS accreditation are measures of child care quality at the provider level, we hypothesized that GS accreditation is positively associated with CLASS scores. This study is guided by the following research questions:

(1) Is accreditation by the different state approved Gold Seal agencies associated with providers' CLASS score?
(2) What characteristics of SR providers are associated with changes in CLASS scores?

To answer the research questions, the ECPRG employed the two-level Multilevel Modeling (MLM) framework (i.e., providers as level-1 and coalitions as level-2) to detect any significant relationships between CLASS scores and the GS agency who

granted accreditation. Specifically, MLM was used to separate the variation of CLASS scores across coalitions from the effects of accreditation. This multilevel model will produce effect estimates associated with provider level predictors (e.g., accreditation, VPK) as well as an estimate of the variability of CLASS scores across the state.

Data

Variables and Data Sources

Two data sources were used in analysis. A provider level file from DEL covering fiscal year 2020-2021 (July 2020 – June 2021, hereafter referred to as FY20-21) included provider level data, including CLASS scores, for all providers contracted to provide SR services. The DEL data were supplemented by the complete list of providers from the Florida Department of Children and Families (DCF) which included data regarding provider accreditation status.

The outcome variable is the SR provider CLASS composite score, provided by the DEL dataset. The independent variable, "GS accreditation of child care providers," was taken from the DCF dataset. Three additional covariates were included to explore the second research question: (1) the average SR enrollment in FY20-21 (hereafter "Avg_SR_enroll"); (2) the capacity of SR providers in FY20-21 (hereafter "Capacity"); and (3) the VPK status (hereafter "VPK", 1 for active and 0 for the inactive VPK provider).

Data Processing

The first step in the data curation process was to link the DEL and DCF datasets based on DCFID. Only providers with a CLASS score who matched across data sources were retained for the analysis.

The ECPRG then cross-tabulated the data to review the frequency distribution of GS agencies. The cross-tabulation revealed that some GS agencies only accredited a small number of providers, which if left in the model would have resulted in a less than acceptable increase in the margin of error for statistical inference. For this reason, the ECPRG chose to exclude agencies who accredited less than 20 providers. For example, there are only 2 providers accredited by AISFL (Association of Independent Schools of Florida) and FCC (Florida Catholic Conference). These providers were removed. This resulted in six GS accrediting agencies included in this study: APPLE, NECPA, NAFCC, GAACS, NACECEP, and NAEYC. To be included in the final data set used for the analysis, a provider needed to be in one of the following two groups: (1) accredited by one of the five agencies and had a CLASS score; or (2) non-GS provider with a CLASS score (i.e., baseline subgroup).

All home-based GS providers were accredited by NAFCC, an agency that only provides accreditation for this provider type. For this reason, the ECPRG analyzed home-based and center-based providers separately.

After retaining providers based on the aforementioned inclusion criteria, the sample consisted of 3,562 providers, including 2,741 center-based providers and 821

home-based providers. Figure 1 shows the distribution of the CLASS scores for center-based and home-based providers, respectively. Tables 1-4 show the descriptive statistics on the providers included in the analysis.

Figure 1. The Density Plots of CLASS Scores for Center-based and Home-based Providers.

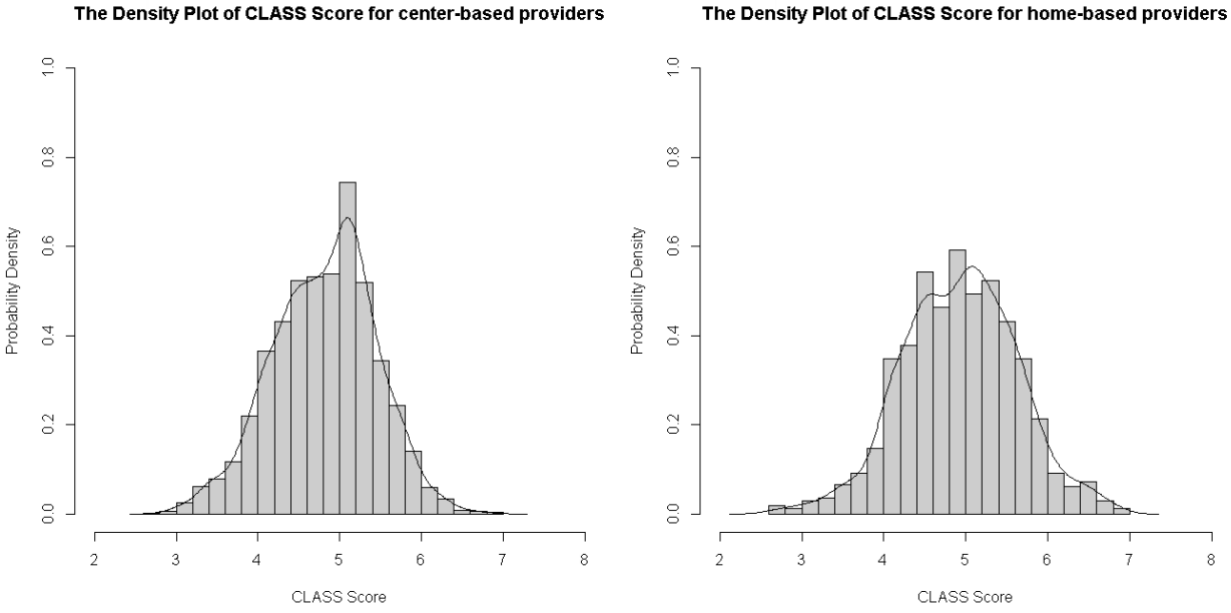


Table 1. Center-based provider descriptive statistics, continuous variables (total: 2,741)

Continuous Variables	Mean (standard deviation)	Range	# Missing
CLASS score	4.82 (0.63)	[2.78, 6.94]	
Capacity	100.23 (63.30)	[10, 595]	26
Average SR enrollment	25.78 (17.97)	[0, 206.25]	

Table 2. Center-based provider descriptive statistics, continuous variables (total: 2,741)

Binary Variables	Count (proportion)
VPK	2412 (88.00%)
GS - APPLE	475 (17.33%)
GS - NECPA	157 (5.73%)
GS - GAACS	24 (0.88%)
GS - NACECEP	163 (5.95%)
GS - NAEYC	74 (2.70%)
Non-GS	1848 (67.42%)

Table 3. Home-based provider descriptive statistics, continuous variables (total: 821)

Continuous Variables	Mean (standard deviation)	Range	# Missing
CLASS score	4.93 (0.70)	[2.61, 6.86]	
Capacity	9.81 (1.86)	[2, 12]	1
Average SR enrollment	4.18 (2.63)	[0, 19.5]	

Table 4. Home-based provider descriptive statistics, binary variables (total: 821)

Binary Variables	Count (proportion)
VPK	168 (20.46%)
GS - NAFCC	131 (15.96%)
Non-GS	690 (84.04%)

Centering the Capacity

Since the multilevel modeling method used in this study has a random effect on the intercept, which means the intercept is the outcome variable at level 2, the interpretation of the intercept should be as follows: the intercept is the expected CLASS score when all predictors are zero (e.g., when capacity is 0). However, models using raw data would cause a meaningless intercept since the capacity of providers cannot be zero. Centering is a commonly used method to make the intercept more interpretable. In this study, we chose to use grand-mean centering for the raw capacity.

Grand-mean centering is calculated separately for center-based and home-based providers by extracting the mean capacity across all coalitions (i.e., 100.231 for center-based and 9.810 for home-based providers) from the raw capacity. Hereafter we refer to this grand-mean-centered capacity as "GMC Capacity".

Missing Capacity

Tables 1-4 show that some providers were missing capacity information. For the center-based providers, although the percent of missing was not high (1%, 26 out of 2,741), the missingness was not randomly distributed across coalitions with a majority of missing capacities in the Early Learning Coalitions of Hillsborough and Palm Beach Counties. We conducted multiple imputation (MI) for the missing capacity to avoid the bias that may be caused by directly excluding those providers. Similarly, the percentage of home-based providers with missing data was low (i.e., 0.12%, 1 out of 821 providers); however, we conducted the MI procedure for this provider as well to keep the resulting interpretation consistent with that of center-based providers.

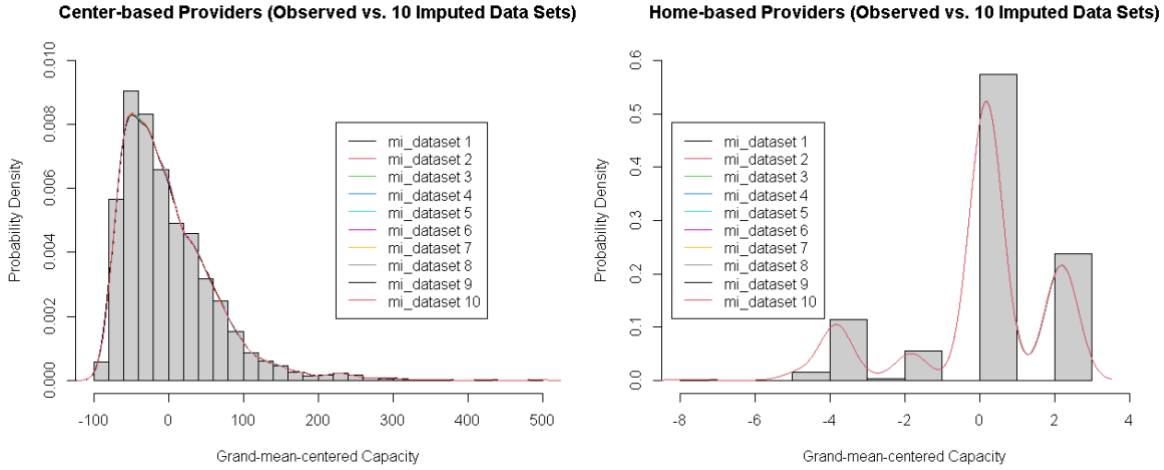
Considering the capacity missingness are non-negative values, the imputation method used in this study is the predictive meaning matching (PMM), which imputes the missingness from the observed capacities. The imputed capacities are not outside the range of the observed capacities (Morris et al., 2014). The MI process was conducted in Stata (StataCorp, 2021) and all variables were included in the imputation procedure. The model settings used for the MI procedure according to Morris et al. (2014) and UCLA Statistical Consulting Group (1b, n.d.)

are: random seed number is 53,421, the number of imputed data sets (m) is 10, and the number of nearest neighbor matches (knn) is 5.

Validated Process for Imputation Data

Figure 2 compares the distribution of the original observed capacity to the multiple imputed capacities. The histogram represents the density distribution of the observed capacities, and the lines represent the densities of the 10 imputed data sets. We can see the imputed capacities have similar distributions to the observed capacity.

Figure 2. Density Plots for the Grand-mean-centered Capacity ("GMC_capacity") of Center-based and Home-based Providers (Observed Data vs. 10 Imputed Data Sets).



We also inspected the performance metrics Relative Variance Increase (RVI) and Fraction of Missing Information (FMI) for the multiple imputation. The average RVI across all variables is .0001 for center-based providers and .0002 for home-based providers. This indicates that the average estimated sampling variance of center-based/home-based providers is 0.01%/0.02% larger than the sampling variance of the model using complete data only. FMI represents the proportion of the total sampling variance that is caused by missing data. The largest FMI across all variables is the variable "grand-mean-centered capacity" for both center-based and home-based providers (with values of .0015 and .0013, respectively). This indicates that the largest proportion of the total sampling variance can be related to missing data is 0.15%/0.13% for center-based/home-based providers. The small RVI and FMI values indicate that the imputed datasets are reliable.

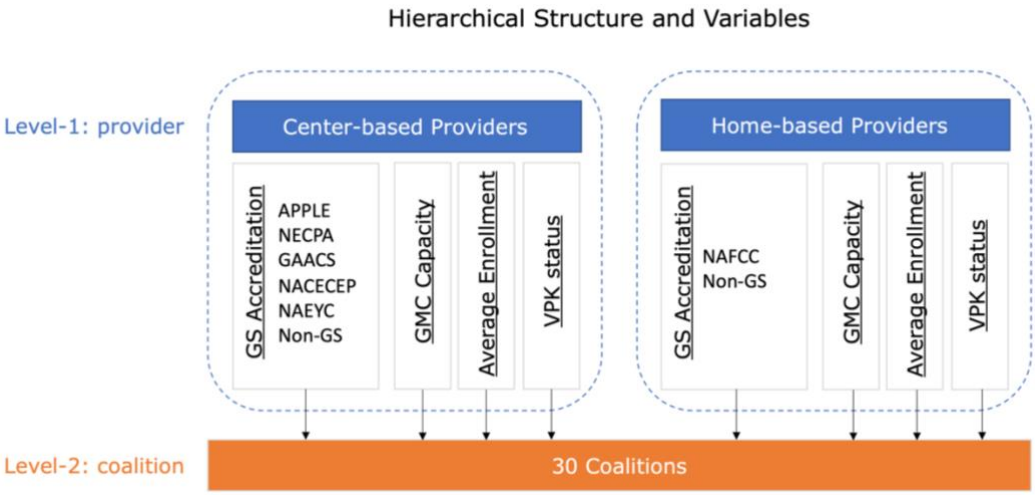
Methodology

Multilevel Modeling

This study explores the relationships between CLASS scores and GS accreditations using the multilevel modeling (MLM) framework. We choose the MLM method because the data structure is nested – providers are nested within coalitions, meaning the spatial, social, and economic differences between coalitions may have affected providers' quality ratings in ways that were more similar within coalitions. Although some researchers argue that MLM is not necessary for models with low Intraclass Correlation Coefficients (ICCs; e.g., ICCs < 0.05) (Maas & Hox, 2005; Hayes, 2006), we believe it is best practice to not ignore the clustering effect because even low ICCs may cause large bias, especially when the within-cluster sample size is large in a particular level-2 unit (i.e., coalition) (Musca et al., 2011; Huang, 2018).

Figure 3 shows the structure and variables included in the MLM models. The samples used for MLM models include 821 home-based providers nested within 30 coalitions (the average observations per cluster is 27.4) and 2,741 center-based providers nested within 30 coalitions (the average observations per cluster is 91.4). All variables (i.e., GS accreditation types, GMC Capacity, average enrollment, and VPK status) are level-1 predictors. Coalition ID is the only variable used at level 2, and no other coalition level predictors are included in this study. To simplify the model, we only included a random effect on the intercept and fixed all slopes because we expected the average CLASS score to vary across coalitions. We did not, however, expect the value-added by specific accreditations or other provider level predictors to vary across coalitions. The MLM models were fit in Stata (StataCorp, 2021) with maximum likelihood estimation (MLE).

Figure 3. The Hierarchical Structure and Variables



Evaluation of Each Variable

We used significance testing to explore whether there is an effect of each predictor on CLASS scores and reported the standardized effect sizes as complements to describe the magnitude of effects (Maher et al., 2013).

As for the effect size measures, we use Cohen's f^2 index to calculate the effect size for fixed effects and ICC to represent the effect size for the random effect, according to the suggestion by Lorah (2018) for multilevel mixed-effect models. Cohen's f^2 is the proportion of variance explained by a variable (e.g., a GS accreditation type) to the unexplained proportion of variance in the outcome variable (CLASS score) (Lorah, 2018). According to Cohen's guidelines (Cohen, 1988), f^2=.02, .15, and .35 represent small, medium, and large effect sizes, respectively. ICC is the proportion of variance in the outcome variable (CLASS score) that can be explained by the level 2 random effect (coalitions), and LeBreton and Senter (2008) suggested adopting the traditional cutoffs that are used to interpret effect sizes, namely .01 for a small effect, .1 for a medium effect, and .25 for a large effect. We followed guidelines described by Selya et al. (2012) and UCLA Statistical Consulting Group (1a, n.d.), which we calculated in Stata.

Results

Center-based Providers

The results for center-based providers can be found in Table 5. The fixed effect of intercept γ_{00} is 4.893 with a standard error of .057, t = 85.170, p <.001, which indicates the average CLASS score for non-GS providers ("non-GS") is 4.893 when all other variables are set to zero (i.e., the average SR enrollment=0, the capacity equals to 100.231, and non-VPK provider).

APPLE, NACECEP, and NAEYC accreditations have significant positive relationships with the CLASS score, with slopes equal to .203, .219, and .284, respectively. This indicates that on average, the CLASS score of providers with APPLE/NACECEP/NAEYC accreditation is .203/.219/.284 points higher than that of providers without any GS accreditation, when controlling for all other variables. However, the values of Cohen's f^2 indicate that all these effects are negligible (f^2 <.02).

The average SR enrollment has a negative relationship with CLASS scores when controlling for GS accreditation types. With one unit of increase in the average SR enrollment, the CLASS score will decrease by .002 units. However, the effect is also negligible according to the effect size measure (f^2=.004<.02).

The random effect residual is an estimate of the variability of model errors. It has a standard deviation, *sd(Residual),* of .574. We can see that after controlling for GS accreditations, provider capacity, SR enrollment, and VPK status at the coalition level, the random variance of the residuals is .329. As for the random effect on intercept across coalitions, the standard deviation *sd(Intercept)* is .233, with a standard error of .033, which indicates the between-coalition variation of intercept (i.e., non-GS providers with no SR enrollment, a capacity of 100.231, and not VPK)

is .054 with a moderate-large effect (ICC=.141>.1). We can conclude that there is substantial cluster-level effect in the CLASS scores of the center-based providers.

Table 5. MLM-1 Results for Center-based Providers

	MLM-1 (Center-based)			
Fixed Effects	**Coefficient**	**Std.err.**	**95% CI**	**Cohen's f^2**
Intercept (γ_{00})	4.893***	.057	[4.780,5.005]	
GMC_capacity (γ_{10})	.000	.000	[-.000, .000]	.000
Avg_SR_enroll (γ_{20})	-.002***	.001	[-.004, -.001]	.004
VPK (γ_{30})	-.023	.035	[-.093, .046]	.000
APPLE (γ_{40})	.203***	.034	[.136, .270]	.013
NECPA (γ_{50})	-.033	.051	[-.133, .067]	.000
GAACS (γ_{60})	.101	.119	[-.133, .335]	.000
NACECEP (γ_{70})	.219***	.049	[.123, .315]	.007
NAEYC (γ_{80})	.284***	.070	[.146, .423]	.006
Random Effects	**Estimate**	**Std.err.**	**95% CI**	**ICC**
sd(Intercept) (τ)	.233	.033	[.176, .308]	.141
sd(Residual) (σ)	.574	.008	[.559, .590]	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

Home-based providers

The results for home-based providers can be found in Table 6. NAFCC accreditation for home-based providers has a significant positive relationship with the CLASS score. On average, the CLASS score of home-based providers with NAFCC accreditation is .404 points higher than that of home-based providers without any Gold Seal accreditation, when controlling for all other variables. Also, the effect is small-moderate according to the value of Cohen's f^2 (.02<.47<.15).
The GMC capacity, average SR enrollment, and VPK status are not significant predictors for home-based providers.

As for the random effect residual has a standard deviation of .635, with a standard error of .016 and 95% CI [.604, .667]. We can see that after controlling GS accreditations, provider capacity, average SR enrollment, and VPK status, the variation of CLASS scores is .403 (.635 squared) and significant. The standard deviation of the intercept across coalitions is .299, with a standard error of .054 and 95% CI [.209, .427]. This indicates the between-coalition variation of intercept (i.e., non-GS providers) is significant with a moderate-large effect size (ICC=.181 > .1). We can conclude that there is substantial clustering in the CLASS scores of the home-based providers.

Table 6. MLM-2 Results for Home-based Providers

	MLM-2 (Home-based)			
Fixed Effects	**Coefficient**	**Std.err.**	**95% CI**	**Cohen's f^2**
Intercept (γ_{00})	4.967***	.076	[4.818,5.115]	
GMC_capacity (γ_{10})	.017	.017	[-.015, .050]	.001
Avg_SR_enroll (γ_{20})	.011	.009	[-.007, .028]	.002
VPK (γ_{30})	-.088	.057	[-.200, .024]	.003

NAFCC (γ_{90})	.404***	.066	[.275, .533]	.047
Random Effects	**Estimate**	**Std.err.**	**95% CI**	**ICC**
sd(Intercept) (τ)	.299	.054	[.209, .427]	.181
sd(Residual) (σ)	.635	.016	[.604, .667]	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

Discussion

Results from this analysis show positive and significant relationships between CLASS scores and four GS accreditation agencies (APPLE, NACECEP, and NAEYC for center-based providers and NAFCC for home-based providers). This indicates that providers with a GS accreditation from one of these agencies tend to have higher CLASS scores than non-GS providers. No significant relationships were detected for the NECPA or the GAACS accrediting organizations.

While GS accreditation is associated with higher CLASS scores, the magnitude of the CLASS score differences among APPLE, NACECEP and NAEYC, as measured by the effect size, is negligible. The effect size of CLASS scores for home-based providers accredited by NAFCC is small.

In contrast, the between-coalition variation had a significant and moderate-large effect on CLASS score variations for both center-based and home-based providers. However, it is not clear why this variability exits—coalition level variation could be explained by real differences in CLASS scores resulting from local quality initiatives or it could be attributed to differences in observer ratings. Further research that includes measures of inter-rater reliability and clear standards for ongoing observer calibration would inform our understanding of variance currently attributed to the coalition level.

As for the characteristics of the SR providers that may influence the CLASS score, capacity and VPK are not significantly associated with differences in CLASS scores. Interestingly, increasing SR enrollment at center-based providers is significantly associated with lower CLASS scores at a significance level of .001. This indicates that increased SR enrollment is associated with decreased composite CLASS scores for center-based providers. This finding raises questions regarding how classroom composition might influence CLASS scores. Future research exploring this is warranted, particularly considering the interest of the Florida SR program supporting access to high-quality experiences for children and families.

Limitations

A key limitation of this study is that we only included providers who received accreditation from GS organizations serving more than 20 providers. This resulted in the exclusion of 11 out of 17 GS accreditation agencies. Additionally, due to data availability, the ECPRG only included three provider attributes in the MLM models: capacity, average SR enrollment, and VPK status. Future research can address this

limitation by accessing and including other provider characteristics such as price for care, overall provider occupancy, and teacher to student ratios.

References

Cohen, J. E. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, *32*, 385–410. https://doi.org/10.1111/j.1468-2958.2006.00281.x

Huang, F.L. (2018). Multilevel modeling myths. *School Psychology Quarterly,* 33(3), 492.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods*, 11(4), 815-852.

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, 6(1), 1-11.

Maas, C., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92.

Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: effect size analysis in quantitative research. *CBE—Life Sciences Education*, *12*(3), 345-351.

Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology*, *14*(1), 1-13.

Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A., & Brauer, M. (2011). Data with hierarchical structure: Impact of intraclass correlation and sample size on type-I error. *Frontiers in Psychology*, *2*, 1–6. https://doi.org/10.3389/fpsyg.2011.00074

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen'sf 2, a measure of local effect size, from PROC MIXED. *Frontiers in psychology*, 3, 111.

StataCorp. (2021). *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC.

UCLA: Statistical Consulting Group (1a, n.d.). How Can I Estimate Effect Size For Mixed Models?. https://stats.oarc.ucla.edu/stata/faq/how-can-i-estimate-effect-size-for-mixed/ (accessed June 23, 2022).

UCLA: Statistical Consulting Group (1b, n.d.). Multiple Imputation in Stata. https://stats.oarc.ucla.edu/stata/seminars/mi_in_stata_pt1_new/ (accessed May 31, 2022).