Measuring the Principal Effectiveness Equity Gap within Wisconsin School Districts

A Study from the State-Wide Evaluation of Wisconsin Educator Effectiveness

Curtis J. Jones & Leon Gilman
University of Wisconsin Milwaukee

March, 2018

Socially Responsible Evaluation in Education

This paper would not have been possible if not for the helpful feedback received from the following:

Sheila Briggs – Wisconsin Department of Public Instruction

Katharine Rainey – Wisconsin Department of Public Instruction

Elizabeth Cain – University of Wisconsin Milwaukee

Scott Davis – University of Wisconsin Milwaukee

Joseph Schmidlkofer – University of Wisconsin Milwaukee

Rachel Westrum – University of Wisconsin Milwaukee

Bradly Carl – University of Wisconsin Madison

Herb Heneman – University of Wisconsin Madison

Steve Kimball – University of Wisconsin Madison

Emily Kite – University of Wisconsin Madison

Anthony Milanowski – Education Analytics

Socially Responsible Evaluation in Education (SREed) conducts research and evaluation studies in the Office of Research at the University of Wisconsin in Milwaukee. SREed works with education agencies to conduct rigorous studies on issues that impact the quality of and equitable access to effective instruction for all Wisconsin students.

For more information about this study or about the state-wide evaluation of Wisconsin Educator Effectiveness, please contact Curtis Jones at jones554@uwm.edu or visit www.uwm.edu/sreed

Table of Contents

Measuring the Principal Effectiveness Equity Gap within Wisconsin School Districts

As part of the ongoing, state-wide evaluation of the state of Wisconsin's Educator Effectiveness (EE) process we used principal effectiveness ratings to measure the extent that effective principals are equitably distributed within Wisconsin school districts. Although there is a considerable body of research on the inequitable access to effective teachers and its effects on student learning (Boyd, Lankford, Loeb, Rockoff, & Wychoff, 2008; Hanushek, Kain, & Rivkin, 2004; Scafidi, Sjoquist, & Stinebricker, 2008), comparatively little research exists on addressing issues regarding access to effective principals. The research that does exist shows low-income and ethnically diverse schools are more likely to have principals with less experience and less education, and that principals in these schools are more likely to transfer away (Loeb, Kalogrides, & Horng, 2010; Clofelter, Ladd, Vignor, & Wheeler, 2007). However, it is not clear how these documented differences in principal credentials translate into differences in effectiveness ratings. The current study examines this by comparing the effectiveness ratings of principals in more diverse and less affluent schools to those in less diverse and more affluent schools.

The Importance of Effective Leadership in Low-income and Diverse Schools

That low-income and diverse schools are often faced with high principal turnover and less effective leadership can be particularly problematic when considering the challenges these schools face in attracting and retaining effective teachers. Effective leadership is critical in low-income and ethnically diverse schools for attracting and retaining effective and qualified teachers (Baker & Cooper, 2005; Beteille, Kalogrides, & Loeb, 2011; Clark, Martorell, & Rockoff, 2009). Additionally, more effective leaders is these schools have been found to help teachers develop into more effective teachers (Beteille, Kalogrides, & Loeb, 2011) and be more satisfied with their jobs (Grissom, 2011). Thus, the schools facing the largest challenges for educating their students, that also have the greatest difficulty finding and retaining effective leaders and teachers, are the most negatively affected by ineffective leadership.

Faced with the growing body of evidence of the inequity of access to quality educators in low-income and ethnically diverse schools, and the detrimental impact this has on these schools, the

U.S. Department of Education established the Excellent Educators for All Initiative. As part of this initiative, each state was required to develop an Educator Equity Plan. The Wisconsin Educator Equity Plan (Wisconsin Department of Public Instruction, 2015) identifies a number of root causes that lead to teacher quality equity gaps but does not specifically identify ineffective school leadership as a cause. This omission is consistent with equity plans developed for states across the U.S., as only 15 mention ineffective school leadership as a cause of teacher equity gaps. (Williams II, Adrien, Murthy, & Pietryka). Related, improving principal effectiveness was not commonly listed by states as a strategy for addressing teacher equity gaps.

The lack of consideration of principal equity in the Wisconsin plan, and in most other state plans, ignores the potential for addressing both teacher equity gaps and principal equity gaps simultaneously. As previously mentioned, effective principal leadership is an important, and perhaps necessary, strategy for attracting and retaining effective teachers in low-income and ethnically diverse schools. Even if teachers in these schools are provided professional development, and their effectiveness improves, if principals are still ineffective, many teachers will still likely leave and will be difficult to replace. By developing these teachers, it may make retention even more difficult as effective teachers may have more employment options. Thus, any effort to address teacher equity gaps should also consider principal equity.

The Wisconsin EE Process for Evaluating Principal Effectiveness

In 2011, Wisconsin passed Act 166, which required the Wisconsin Department of Public Instruction (DPI) to develop standardized performance-based principal and teacher evaluation systems. DPI had already begun developing such a system with its EE process. As part of this process, districts provide ongoing, formal performance feedback to principals, using a standard process and leadership rubric. With assistance from the Wisconsin Center for Education Research, DPI developed a principal effectiveness rubric called the Wisconsin Framework for Principal Leadership (WFPL). The purpose for developing the principal evaluation process using the WFPL was to provide districts with a valid measure of principal leadership that could also guide local efforts for developing principals. Although state legislation allows for alternative models approved by DPI, about two-thirds of Wisconsin districts chose the DPI

evaluation model, which includes the WFPL, to inform their principal evaluation and feedback process.

Previous Studies of WFPL Ratings

The current study represents a continuation of previous research that found WFPL ratings had high concurrent validity with teacher perceptions of principal effectiveness (Jones, Gillman, Kimball, & Rainey, 2017). A study of WFPL ratings from the 2016-17 school year found strong relationships between WFPL ratings and a variety of measures of teacher perceptions of principal effectiveness; principals rated as more effective leaders according to the WFPL had teachers who believed they were better leaders, held greater levels of trust in them, believed they were more qualified to provide teachers performance feedback, and felt that the feedback they provided was more useful and accurate. The finding that more effective principals provide more effective performance feedback to teachers suggests a possible explanation of why previous research found that teachers in schools with more effective leaders develop into more effective educators (Beteille, Kalogrides, & Loeb, 2011). These findings also suggest that ratings assigned to principals by their administrator accurately reflect the conditions in schools and therefore provide a useful tool for measuring the quality of school leadership Wisconsin school districts. The current study capitalizes on this by using ratings to explore the issue of principal equity.

Current Study

The current study explores principal equity gaps within 61 school districts and 322 Wisconsin schools by using principal performance ratings data collected as part of the state's EE process during the 2016-17 school year. This is the first study the authors are aware of that uses a state's EE system to quantify principal effectiveness equity gaps. We use ratings assigned to principals by their administrator to identify school and principal characteristics that are related to principal effectiveness, and to measure the magnitude of principal effectiveness equity gaps. Although we predict that novice principals and principals in less affluent and more diverse schools will be rated as less effective, we make no specific predictions regarding the magnitude of these differences.

Study Methods

Data sources

My Learning Plan. Most Wisconsin schools document their EE process and final EE ratings using My Learning Plan (MLP), an online data management system supported by Frontline Education. Roughly 1,000 schools, state-wide, that chose to use the state EE model and process, used MLP during the 2016-17 school year to document their EE process.

The WFPL is organized by two main leadership domains (Effective Educators and Leadership Actions) and five subdomains (Human Resource Leadership, Instructional Leadership, Personal Behavior, Intentional and Collaborative School Culture, and School Management) (Wisconsin Department of Public Instruction, 2017). At the end of an evaluation cycle, based on evidence that includes leadership documents and observations of practice collected throughout the year, principals are assigned by their district administrator possible ratings of 1 (*Unsatisfactory*), 2 (*Basic*), 3 (*Proficient*), or 4 (*Distinguished*) on each of the 21 effectiveness components. For purposes of this study, subdomain ratings were calculated by averaging the component ratings, and an overall rating of principal effectiveness was calculated by averaging these five subdomains. In the current study, the reliability of the 21 component ratings was 0.92.

Public records of school characteristics. School characteristics from the 2016-17 school year were extracted from public data maintained by DPI. These data include the type of school (elementary, middle, or high school), student enrollment, and the percent of students in each school that are economically disadvantaged (ED), white, and with a disability.

WISEstaff. Principal characteristics were also extracted from public data maintained by DPI. The data used for this report included the 2014-15, 2015-16, and 2016-17 WISEstaff files. The most recent file (2016-17) was used to identify the demographic characteristics of each principal (race, gender, and education). The three files together were used to identify principals who were in their first year, those who had two years of experience, and those who had more than two years of experience as a principal in the State of Wisconsin.

Schools included in the study

The sample of schools included in this study was taken from the previously referenced study of WFPL ratings and their validity (Jones, et. al., 2017). All 322 schools across 61 school districts

included in this study had a principal who received WFPL ratings at the end of the 2016-17 school year. There were roughly 1,000 schools state-wide using MLP and following the state model of EE that includes the WFPL. We estimate that the population of schools scheduled to have principal ratings in 2016-17 was 400. This assumes that 10% of the 1000 principals would receive ratings as new principals, and 300 would receive "summative" ratings, which occur every three years. Based on this, the 322 schools included in this study reflect about 81% of the potential population of schools.

Among these 322 schools, a disproportionate number (64; 20%) were in the Milwaukee Public Schools (MPS) district. Kenosha, the next largest district, had 34 schools in the sample. Table 1 presents the districts with the most study schools. Eighteen of the 61 districts included in this study had only one school in this study.

Table 1. Districts with more than 10 schools included in study

District	Schools with WFPL ratings
Milwaukee Public Schools	64
Kenosha School District	34
Madison Metropolitan School District	20
Green Bay Area Public School District	19
Waukesha School District	15
Racine Unified School District	13
Appleton Area School District	12
Oshkosh Area School District	11

Among these 322 schools, 210 (65.2%) were elementary schools, 43 (13.4 %) were middle schools, 56 (17.4%) were high schools, and 13 (4.0%) were combined elementary and secondary schools. In 57 schools (18%), the principal was a first-year principal, in 31 (10%) schools the principal was in their second year, and in 234 (73%) schools the principal had three or more years of experience. Since principal turnover is more common in MPS (Simonaitis, 2013), a disproportionate number of the first-year (16; 28%) and second-year principals (8; 26%) were MPS principals. One hundred seventy-seven (55%) principals were female, 250 (78%) were white, and 51 (16%) were African American. Most principals (284, 88%) held a master's degree as their highest education. Table 2 presents the descriptive characteristics of schools in this study.

Table 2. Descriptive statistics of schools in study

	Min	Max	Mean	Std. Dev.
Enrollment	24	2287	476	321
Percent Students with Disability	3.5%	77.8%	15.7%	6.7%
Percent Economic Disadvantaged Students	0.6%	100%	52.7%	26.2%
Percent White Students	0.6%	100%	55.4%	31.4%
WFPL Ratings	1.96	4	3.08	0.30

Analytic methods

Bivariate statistical analysis and multivariate modeling were used to analyze the equitable access to effective principals with Wisconsin school districts. Generalized Linear Modeling (GLM), with Robust Standard Errors and fixed district effects, was then used to identify the school and principal factors that most related to WFPL ratings within districts. All of the study factors listed in Table 2, along with fixed school effect, school type (high school versus not) and principal experience, were initially included in the models. A parsimonious model was built by removing non-significant factors one at a time. The resulting model included just three factors, the fixed factors of district, principal tenure, and percent white students. The functional form of the relationship between school racial composition with WFPL ratings was tested by including the quadratic of percent white students in a school, and was found to be significant. The resulting model presented below was then used to examine the relationship of WFPL ratings with school student racial composition and principal experience within school districts:

$$\text{WFPL rating}_{ij} = \gamma_{0j} + \gamma_{01j}(\%\text{white}) + \gamma_{02j}(\text{First} - \text{year principal}) + \gamma_{03j}(\text{Second} - \text{year principal}) + \gamma_{04j}(\%\text{white} * \%\text{white}) + \sum_{j-1}^{j} \gamma_{05j} D_j + r_{ij}$$

Where the WFPL rating of principal (i) in district (j) is a function of the average rating of all principals (γ_{0j}) the percentage of a school that is white (γ_{01j}), whether they are in their first year as a principal (γ_{02j}), whether they are in their second year as a principal (γ_{03j}), the quadratic relationship of the percent of a school that is white (γ_{04j}), where *Dj* is the school *j* dummy with coefficient γ_j ($j = 1,\ldots, J$), and an error term r_{ij}. The results of this model tell us estimate the extent there is equity in access to effective principals within school districts.

Results

Bivariate relationships

Across all schools, WFPL ratings were unrelated to school size (Table 3). WFPL ratings were found to be positively correlated with the percent of the school that was white ($r = .292$, $p < 0.01$) and negatively correlated with the percent of the school that was economically disadvantaged ($r = -.242$, $p < 0.01$). Among the subdomains of the WFPL, Instructional Leadership was found to be the most correlated both with the percent of the school that was white ($r = .294$, $p < 0.01$) and with the percent of the school that was economically disadvantaged ($r = -.249$, $p < 0.01$).

Table 3. Correlation matrix of school factors (n = 322)

	1	2	3	4	5	6	7	8	9
1. WFPL effectiveness rating	1								
2. Human Resource Leadership Subdomain	.891**	1							
3. Instructional Leadership Subdomain	.864**	.777**	1						
4. Personal Behavior Subdomain	.853**	.717**	.630**	1					
5. Intentional and Collaborative School Culture Subdomain	.849**	.698**	.661**	.682**	1				
6. School Management Subdomain	.699**	.504**	.485**	.511**	.477**	1			
7. Enrollment	.040	.013	.020	.054	.043	.036	1		
8. Percent students with disabilities	-.135*	-.164**	-.253**	-.077	-.083	.045	-.099	1	
9. Percent students economic disadvantaged	-.242**	-.213**	-.249**	-.229**	-.175**	-.128*	-.104	.459**	1
10. Percent White students	.292**	.262**	.294**	.260**	.193**	.197**	-.119*	-.356**	-.844**

Note: *$p < 0.05$, **$p < 0.01$

The relationships between principal experience with school and principal characteristics were explored through a series of mean comparisons. Table 4 presents mean comparisons of principal tenure and WFPL ratings. New principals received lower WFPL ratings. First-year principals were rated approximately a standard deviation less effective than principals with three or more years of experience ($z = 94$). A comparison of WFPL subdomain ratings (Table 5) suggests that more experienced principals were rated as more effective on all aspects of their leadership. The differences in effectiveness among the different WFPL subdomains did not clearly suggest that more experienced principals were differentially more effective in specific areas of leadership. Since the correlations did not suggest that domain ratings were differentially correlated with school characteristics (Table 3), and these results suggest that principal experience was also not differentially related to WFPL domain ratings, all subsequent analysis presented in this paper focus exclusively on overall WFPL ratings.

Table 4. WFPL ratings - comparisons of new and experienced principals

New	Mean	Std. Dev	ANOVA results	Sum of Squares	*df*	Mean Square	*F*	Sig.
First-year	2.88	0.31	Between Groups	3.499	2	1.749	21.208	<.001
Second-year	2.98	0.29	Within Groups	26.314	319	0.082		
Three or more years	3.14	0.28						

Table 5. WFPL subdomain ratings - comparisons of new and experienced principals

	First-year		Second-year		Three or more years		Difference	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Scale	Effect size (z)
Human Resource Leadership	2.87	0.36	2.92	0.33	3.13	0.34	0.26**	0.73
Instructional Leadership	2.75	0.40	2.90	0.37	3.11	0.38	0.35**	0.87
Personal Behavior	2.91	0.36	3.06	0.44	3.17	0.35	0.26**	0.70
Intentional and Collaborative School Culture	2.88	0.39	2.94	0.36	3.12	0.35	0.24**	0.65
School Management	2.96	0.33	3.05	0.33	3.16	0.31	0.20**	0.61

Note: * = $p < 0.05$, ** = $p < 0.01$

Mean comparisons of principal experience and school racial composition are presented in Table 6. Across all participating Wisconsin schools, and within non-MPS schools, there was no clear relationship between the racial composition of schools led by new or experienced principals. Since there was no relationship found between principal experience and school racial or

Table 6. Comparison of school racial composition (% white) by principal tenure

New	Mean	Std. Dev	ANOVA results	Sum of Squares	*df*	Mean Square	*F*	Sig.
All schools								
First-year	48.1%	33.6%	Between Groups	4083	2	2042	2.09	0.126
Second-year	53.4%	33.5%	Within Groups	312436	319	979		
Three or more years	57.4%	30.4%						

Table 7 presents mean comparisons of principal experience and the student economic composition of schools. There was no difference in the economic composition of schools led by new or experienced principals. Since there was no relationship found between principal experience and school racial or economic composition, we did not provide comparisons of these school characteristics broken down by WFPL subdomains.

Table 7. Comparison of school economic composition (% Free/reduced lunch) by principal tenure

New	Mean	Std. Dev	ANOVA results	Sum of Squares	*df*	Mean Square	*F*	Sig.
All schools								
First-year	59.8%	25.6%	Between Groups	3905	2	1952	2.876	0.058
Second-year	54.5%	26.8%	Within Groups	216539	319	679		
Three or more years	50.7%	26.1%						

Modeling results

Statistical modeling was then used to predict WFPL ratings, conditioned on district, principal experience, and school racial composition. School racial composition was chosen in lieu of

school economic conditions because the two factors suppressed each other in the models and school racial composition had a stronger relationship with WFPL ratings. The results show that, state-wide, district, school racial composition and principal experience all independently explained WFPL ratings (Table 8). The quadratic form of school racial composition was also predictive of WFPL ratings, suggesting that the relationship of school racial composition with WFPL ratings was not linear.

Table 8. Results of generalized linear models of WFPL ratings

	Type III Wald Chi-Square	*df*	*p*-value
(Intercept)	979.595	1	< .001
District	27531634.12	47	< .001
Principal experience	17.767	2	< .001
Percent White Students	11.183	1	0.001
Percent White Students (Quadratic)	19.156	1	< .001

An analysis of the beta coefficients (Table 9) provides more detail about how principal experience and school racial composition related with WFPL ratings within districts. Within school districts, first-year principals were rated an adjusted 0.19 points less effective (z = .63) than experienced principals. The negative coefficient for the quadratic term of racial composition suggests that WFPL ratings are more related to school racial composition in the most racially diverse schools within districts. Using the resulting model to predict principal effectiveness ratings demonstrates the size of the equity gap found within school districts. A principal in a school with 50% ethnically diverse students would be predicted to be rated .34 scale points higher than a principal in a school with 100% diverse students. This difference corresponds to a 1.1 standard deviation difference in their effectiveness rating.

Table 9. Modeled relationships between school and principal characteristics and WFPL ratings

	B	*Std. Error*
First-year	-0.190	0.0471
Second-year	-0.094	0.0603
Three or more years	-	-
Percent White Students	0.01199	0.00274
Percent White Students (Quadratic)	-0.000105	0.0000314

A scatter-plot (Figure 1) of school racial composition and WFPL ratings (Figure 1) shows a clustering of schools with nearly 100% students of color, and with, generally speaking, less effective principals and a number of novice principals. The opposite seems to be true in schools comprised of nearly 100% white students. Considering that novice principals were also rated less effective, novice principals in schools comprised of mostly students of color would be predicted to be rated .53 scale points lower on the WFPL rubric, which corresponds to a 1.5 standard deviation difference, than a more experienced principal in a less diverse school comprised of only 50% students of color.

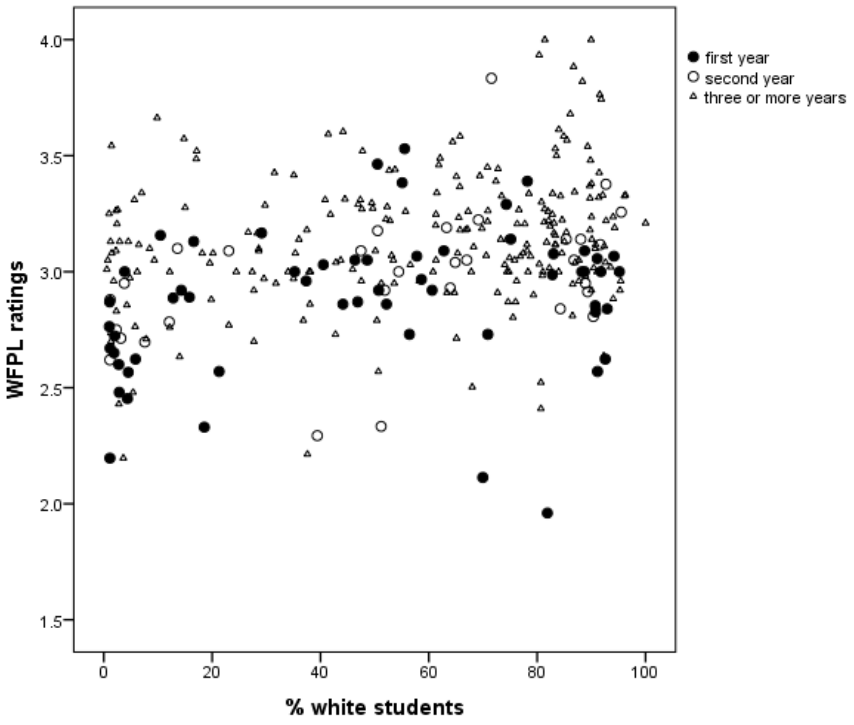

Figure 1. Scatterplot of school student racial composition and WFPL ratings. This figures illustrates the relationship between WFPL ratings and school racial composition.

Conclusions

With the development and application of the WFPL as part of the EE process, there is now a common principal effectiveness metric that is used across a large number of Wisconsin schools. To capitalize on this, we used WFPL ratings to measure the extent that low-income and ethnically diverse schools had unequal access to effective principals within school districts. As predicted, more effective principals were found in less diverse and more affluent schools. These findings provide clear evidence of principal equity gaps.

The results from statistical modeling indicate that the equity gaps of school leadership in Wisconsin are large. Less experienced principals in diverse schools had effectiveness ratings 1.5 standard deviations lower than did more experienced principals in schools with fewer students of color. These differences were found within school districts. Regardless of experience, principals in the most ethnically diverse schools, within the same school district, were rated as over a standard deviation less effective than principals in less diverse schools. This effectiveness gap is compounded when the school is led by an inexperienced principal.

Discussion

This study informs a growing body of evidence that low-income and ethnically diverse schools disproportionally face challenges attracting and retaining effective principals (Loeb, et. al., 2010; Clofelter, et. al., 2007; Horng, Kalogrides, & Loeb, 2009), by quantifying, for the first time, how these challenges translate into ratings of principal effectiveness. That principal effectiveness equity gaps within school districts were found was not surprising. What was surprising was the magnitude of these gaps. Considering the greater importance of effective leadership for attracting and retaining effective teachers in diverse schools (Beteille, Kalogrides, & Loeb, 2011), these findings are particularly troubling.

Our previous research found that less effective Wisconsin principals, with lower WFPL ratings, provide teachers with lower quality feedback and have teachers who are less likely to trust them or believe they are qualified. Given the large impact principals have on teacher employment experiences, and the size of the effectiveness gaps measured in this study, many of the most diverse Wisconsin schools are likely to continue to face difficulties attracting and retaining effective teachers. Any efforts to improve their access to effective teachers must also address their access to effective school leadership.

While addressing principal effectiveness is important for a number of reasons, it is also important to consider the magnitude of the challenge. Given the size of the achievement gaps measured in the current study, it is not surprising that it takes a long time, as long as seven years, to develop an effective principal (Fullan, 2001). It is therefore critical that the retention of effective principals be a central component of districts and state efforts to address teacher and principal equity gaps. A recent review of the literature on principal turnover (Rangel, 2018) indicated that effective professional development may improve principal retention. Considering this, the EE process, using the WFPL rubric, has the potential to improve principal retention in diverse schools and therefore, indirectly, reduce principal inequity. The WFPL rubric can be used by districts to inform their efforts to develop principals, and ratings can be used to track progress for developing principals. The WFPL, by breaking down and defining the components of principal leadership, provides districts with a roadmap for addressing the specific development needs of school leaders. However, considering that the recent review of principal turnover (Rangel, 2018) also concluded that our understanding of the reasons for principal turnover is weak, any efforts to retain and develop effective Wisconsin principals in diverse schools will likely need to explore and utilize a number of strategies. The state-wide evaluation of the Wisconsin Educator Effectiveness process will continue to track and monitor the impact of any coordinated efforts to do so.

Study Limitations

One limitation of this study is that the sample reflects a relatively small number of schools across the state. It is therefore unknown how well these results generalize to the whole state. All new principals, defined as new to the job of principal or new to the district, undergo a summary year evaluation, while principals beyond their first year are evaluated every three years. Thus, the sample of schools in this study includes a disproportionate number of new principals and therefore is not representative of the state. However, it is not important that the current sample of schools represent the state. The purpose of this study was to estimate effectiveness equity gaps. There is no reason to believe that more effective principals in more diverse schools were less likely to be evaluated in the 2016-17 school year. Thus, it seems likely that the equity gaps measured in this study at least approximate the actual gaps that exist within Wisconsin districts.

Another potential limitation is that principals in schools with underperforming students or ineffective teachers may be rated lower as a function of the conditions of the school rather than because of their actual performance. Although it is unknown how independent ratings are from school or student characteristics, the results presented in this study are consistent with previous research. Future research of the Wisconsin EE system will allow us to test this issue as we track the effectiveness of principals as they change schools over time.

Even with these limitations, the findings presented in this report have significant implications for the Wisconsin education system and suggest that districts in Wisconsin might leverage the EE process to begin to address equity issues in schools with large proportions of low-income and ethnically diverse students. This study also informs the literature on principal effectiveness by quantifying effectiveness gaps for the first time.