Measuring the Effectiveness of Wisconsin Principals: A Study of Wisconsin Framework for Principal Leadership Ratings

Curtis J. Jones & Leon Gilman, University of Wisconsin Milwaukee
Steven Kimball, University of Wisconsin Madison
Katharine Rainey, Wisconsin Department of Public Instruction

March, 2018

Socially Responsible Evaluation in Education

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

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

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

SREED
Socially Responsible Evaluation in Education

Table of Contents

Measuring the Effectiveness of Wisconsin Principals:

A Study of Wisconsin Framework for Principal Leadership Ratings

Research has consistently demonstrated that teaching quality represents the largest in-school influence on student achievement (Chetty, Freidman, & Rockoff, 2014; Dee & Wykoff, 2015; Hanushek, 2012; Koedel, Mihaly, & Rockoff, 2015). Principal leadership is critical in creating conditions that attract, develop and retain teaching talent (Hallinger, & Heck, 1996; Marzano, Waters, & McNulty, 2005; Leithwood, Louis, Anderson, & Wahlstrom, 2004). Capitalizing on these findings and accelerated through federal policies such as Race to the Top, the Teacher Incentive Fund, the School Improvement Grant program and the No Child Left Behind waiver process, most states have engaged in principal evaluation reform efforts. Although there have been informative studies of principal evaluation designs and processes (Anderson, & Turnbull, 2016; Goldring, Cravens, Murphy, Porter, Elliott, & Carson, 2009; Kimball, Milanowski, & McKinney, 2009), the authors are only aware of one study that has explored the measurement quality of recent principal evaluation systems (McCullough, Lipscomb, Chiang, Gill, & Cheban, 2016). This study begins to address this gap in the literature by examining evidence of the concurrent validity of ratings assigned to Wisconsin principals as part of the Wisconsin Educator Effectiveness system and teacher perceptions of principal effectiveness.

Wisconsin's Educator Effectiveness (EE) Process

In 2011, Wisconsin passed Act 166, which required the state Department of Public Instruction (DPI) to develop an Educator Effectiveness (EE) system and school districts to implement principal and teacher evaluation systems based on the state EE System (Wisconsin, 2011). The Wisconsin DPI had already begun developing the new evaluation process prior to the legislative requirement. The primary purpose of the Wisconsin EE process is to help school districts implement local processes that improve the quality and quantity of performance feedback educators receive, which in turn improves the quality of leadership, teaching, and student learning across the state. As part of the EE process, districts are required to provide ongoing, formal feedback to principals about their practice using a standard process and leadership rubric.

Although there is some flexibility allowed within the process, at a minimum, all principals go through an evaluation process that includes formal feedback on their performance every three

years.[1] Principals who are new to the profession or new to a district go through the evaluation process in their first year and every third year thereafter. During these summative feedback years, principals engage in School Learning Objectives (SLO) and Professional Practice Goal (PPG) setting processes. The SLO process uses student data to address a school academic need. In the PPG process, principals identify an area of self-growth to focus on. This area is aligned to the framework of principal effectiveness the district is using. The PPG and SLO form a basis for school leadership focus and are intended to shape a continuous improvement process. At the end of the year, district administrators provide principals evaluations ratings, using a leadership rubric, based on leadership documents and observations of practice collected throughout the year.

Although all Wisconsin principals must complete the steps of the EE process, individual districts are permitted to choose a principal leadership framework to use to inform their EE process. With assistance from the Wisconsin Center for Education Research (WCER), DPI developed a principal effectiveness rubric called the Wisconsin Framework for Principal Leadership (WFPL). About two-thirds of Wisconsin districts have chosen the WFPL to inform their principal evaluation and feedback process. Following two pilot years, Wisconsin districts began fully implementing the principal EE process during the 2014-15 school year. This paper focuses on WFPL ratings assigned to principals at the end of the 2016-17 school year, which was the first year that WFPL ratings were collected from districts across the state.

Current Study

The current study presents WFPL ratings of 322 principals from the 2016-17 school year and measures the concurrent validity of WFPL ratings with teacher perceptions of principal effectiveness across and number of areas. This is the first study that the authors are aware of that explores how ratings assigned to principals as part of an EE process reflect the conditions experienced by teachers within schools.

[1] Some Wisconsin school districts conduct annual performance evaluations.

Study Methods

Data Sources

My Learning Plan. Most schools document their EE process and end-of-year EE ratings using My Learning Plan (MLP), an online data management system supported by Frontline Education. As part of setting up their EE process in MLP, districts indicate the type of evaluation cycle educators will follow each year, summary, supporting, or new educator (within their first three years). Both summary year and first-year educator ratings are documented within MLP at the end of the year. Teacher contact information (email) is also documented in MLP. 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.

Public Records of School and Principal Characteristics. School and principal characteristics from the 2016-17 school year were pulled from public data maintained by DPI. These data include principal race, gender, and education, the type of school (elementary, middle, or high school), student enrollment, and the percent of students in each school who are economically disadvantaged (ED), white, and have a disability.

Annual EE Evaluation Survey of Teachers. As part of the state evaluation of the EE system, all Wisconsin classroom teachers are invited each spring to participate in a survey and provide feedback about their school and district EE processes. This survey, sent directly to teachers through the Qualtrics survey program, includes questions measuring many aspects of the EE process and items that ask about school social factors related to the EE process. In the 2016-17 evaluation, over 21,000 teachers completed the survey, which represented about 42% of all Wisconsin classroom teachers.

Instruments

Administrator ratings of principal effectiveness using the Wisconsin Framework for Principal Leadership (WFPL). The WFPL includes leadership standards and rubrics that were developed in 2012-13 and tested as part of the state EE system pilot. The rubric was revised in 2014 based on feedback from school and district leaders. The WFPL is organized by two main leadership domains (Effective Educators and Leadership Actions) and five sub-domains (Human Resource Leadership, Instructional Leadership, Personal Behavior, Intentional and Collaborative

School Culture, and School Management). The domains are identified by 21 components representing leadership competencies. Each component includes a 4-level rubric with descriptions of leadership actions. Principals are assigned ratings of 1 (Unsatisfactory), 2 (Basic), 3 (Proficient), or 4 (Distinguished) on each effectiveness component.[2]

Table 1 presents the framework and descriptive ratings results for the 322 principals included in this study. For this study, subdomain ratings are calculated by averaging the component ratings and an overall rating of principal effectiveness is calculated by averaging these five subdomains. In the current study, the reliability of all 21 component WFPL ratings was 0.92, for the five Human Resources Leadership components it was .75, for the five Instructional Leadership components it was .80, for the four Personal Behavior components it was .70, for the four Intentional and Collaborative School Culture components it was .73, and for the three School Management components it was .59.

Teacher perceptions of principal effectiveness. Two subscales of the 5Essentials Survey (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010), which were included in the Annual EE Evaluation Survey of Teachers, were also included in this study. Principal Leadership was chosen as it has a theoretical connection to WFPL ratings (i.e., Hallinger, & Heck, 1996; Marzano, Waters, & McNulty, 2005; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Spillane, Halverson, & Diamond, 2001; Stronge, Richard, Catano, 2008). Principal – Teacher Trust was also included because it is an important organizational quality of effective schools (Bryk & Schneider, 2002). For both scales, teachers are asked whether they *Strongly Disagree*, *Disagree*, *Agree*, or *Strongly Agree* with statements about their principal. Responses were recoded such that *Strongly Disagree* was coded as a 1, *Disagree* as 2, *Agree* as 3, and *Strongly Agree* as a 4. Overall ratings for each scale were calculated by averaging the responses across questions. Higher scores on either of these scales reflect more effective principals. The specific questions and responses for each scale are listed in Table 2. In the current study, the reliability of responses for Principal Leadership was 0.94 and for Principal – Teacher Trust was 0.96.

[2] The full WFPL may be accessed here: https://dpi.wi.gov/sites/default/files/imce/ee/pdf/WI_Framework_for_Principal_Leadership.pdf

Table 1. Wisconsin Framework for Principal Leadership (WFPL) Ratings

Domain	Component		Unsatisfactory	Basic	Proficient	Distinguished	Mean	Std. Dev
Human Resource Leadership	1a	Recruiting and Selecting	0	5	277	40	3.11	0.36
	1b	Assignment of Teachers and Instructional Staff	0	16	256	50	3.11	0.44
	1c	Observation and Performance Evaluation	2	42	236	42	2.99	0.54
	1d	Professional Development and Learning	0	30	227	65	3.11	0.53
	1e	Distributed Leadership	1	55	201	65	3.02	0.62
Instructional Leadership	2a	Mission and Vision	0	44	229	48	3.01	0.54
	2b	Student Achievement Focus	0	63	207	50	2.96	0.59
	2c	Staff Collaboration	1	21	216	83	3.19	0.55
	2d	School-wide Use of Data	0	73	201	47	2.92	0.61
	2e	Student Learning Objectives (Teacher SLOs)	1	19	267	34	3.04	0.42
Personal Behavior	3a	Professionalism	1	9	185	126	3.36	0.55
	3b	Time Management and Priority Setting	1	42	236	42	2.99	0.52
	3c	Use of Feedback for Improvement	1	28	266	26	2.99	0.43
	3d	Initiative and Persistence	0	29	223	69	3.12	0.54
Intentional and Collaborative School Culture	4a	School Climate	2	38	217	63	3.07	0.58
	4b	Communication	2	27	254	37	3.02	0.47
	4c	Conflict Management and Resolution	1	21	243	57	3.11	0.49
	4d	Consensus Building	1	19	265	37	3.05	0.43
School Management	5a	Learning Environment Management	0	10	227	85	3.23	0.49
	5b	Financial Management	0	23	256	43	3.06	0.45
	5c	Policy Management	0	11	282	29	3.06	0.35

Table 2. 5Essentials Survey questions and responses

Principal Leadership Scale	Strongly Disagree	Disagree	Agree	Strongly Agree	Mean	Std. Dev.
The principal participates in instructional planning with teams of teachers.	386	811	1,205	519	2.64	0.92
The principal knows what's going on in my classroom.	385	693	1,308	531	2.68	0.92
The principal carefully tracks student academic progress.	170	478	1,530	724	2.97	0.81
The principal encourages teachers to implement what they have learned in professional development.	141	351	1,557	860	3.08	0.78
The principal communicates a clear vision for our school.	207	471	1,385	831	2.98	0.86
The principal makes clear to the staff his or her expectations for meeting instructional goals.	160	451	1,475	812	3.01	0.81
The principal understands how children learn.	167	395	1,489	844	3.04	0.81
The principal sets high standards for student learning	156	323	1,459	948	3.11	0.80
Principal – Teacher Trust Scale						
It's OK in this school to discuss feelings, worries, and frustrations with the principal.	325	495	1,260	829	2.89	0.94
The principal looks out for the personal welfare of the staff.	287	499	1,292	829	2.92	0.92
I trust the principal at his or her word.	281	559	1,255	806	2.89	0.92
The principal is an effective manager who makes the school run smoothly.	340	680	1,252	628	2.75	0.93
The principal places the needs of children ahead of personal and political interests.	190	417	1,355	942	3.05	0.85
The principal has confidence in the expertise of staff.	208	463	1,344	891	3.00	0.87
The principal takes a personal interest in the professional development of staff.	185	453	1,452	819	3.00	0.83
Staff feel respected by the principal.	305	596	1,265	737	2.84	0.92

Teacher Perceptions of the Feedback Process. The Examining Evaluator Feedback Survey (Cherasaro, Brodersen, Yanoski, Welp, & Reale, 2015) was used to measure teacher perceptions of Feedback Quality, Feedback Accuracy, and Evaluator Qualifications to Provide Feedback. Although in some schools individuals other than principals may provide formal feedback to teachers, in most schools formal feedback would be provided to teachers by their principal. As such, these three aspects of the feedback process are used in this study as another measure of principal effectiveness. Only teachers who indicated that they had received feedback in the past year were asked to answer these questions.

In the survey, teachers are asked whether they *Disagree*, *Somewhat Disagree*, *Somewhat Agree*, or *Agree* with statements about the feedback process. Responses were recoded such that *Disagree* was coded as a 1, *Somewhat Disagree* as a 2, *Somewhat Agree* as a 3, and *Agree* as a 4. Overall ratings for each scale were calculated by averaging responses across questions. Higher scores reflect more effective feedback. The specific questions and responses for the teachers included in this study are listed in Table 3. In the current study, the internal reliability of these scales was 0.95 for Feedback Quality, 0.86 for Feedback Accuracy, and 0.94 for Evaluator Qualifications to Provide Feedback.

Table 3. Examining Evaluator Feedback Survey Questions and Responses

Feedback Quality (My evaluator's feedback…)	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Mean	Std. Dev.
included specific improvement suggestions.	278	286	740	636	2.89	1.02
included specific suggestions to improve my content/subject knowledge.	407	362	660	503	2.65	1.08
included specific instructional strategies that I could use to improve my teaching.	329	320	716	562	2.78	1.05
included specific classroom management strategies that I could use to improve my teaching.	382	359	671	512	2.68	1.07
included recommendations for finding resources or professional development to improve my teaching.	446	384	632	459	2.57	1.09
was provided as frequently as I needed it.	263	313	698	653	2.90	1.02
was provided in time for me to use it to inform my practice.	245	229	742	711	3.00	1.00
Feedback Accuracy						
The feedback I received was an accurate portrayal of my teaching.	194	194	715	827	3.13	0.96
The classroom observations or walkthroughs that informed the feedback I received represented a typical day in my classroom.	167	220	708	837	3.15	0.93
In our evaluation system, different evaluators reviewing the same evidence would likely give the same ratings.	275	274	740	627	2.90	1.02
Evaluator Qualifications (In my opinion, my evaluator had sufficient…)						
knowledge of my subject/content to effectively evaluate me.	211	241	643	831	3.09	0.99
knowledge of how my students learn to effectively evaluate me.	178	256	681	805	3.10	0.96
knowledge of effective teaching practices to effectively evaluate me.	118	187	652	965	3.28	0.88
understanding of the curriculum being observed to effectively evaluate me.	188	269	664	797	3.08	0.97
understanding of the established teacher evaluation system to effectively evaluate me.	105	136	628	1,061	3.37	0.84

Schools and Principals Included in the Study

Three hundred twenty-two WFPL principal ratings across 61 school districts were documented in MLP at the end of the 2016-17 school year. One hundred seventy-seven (55.0%) of these principals were female, 250 (77.6%) were white, 51 (15.8%) were African American, and 284 (88.2%) had a master's degree as their highest level of education.

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 school principals with ratings was 400. This assumes that 10% of the 1000 principals received ratings as novice principals, and 300 received "summative" ratings, which occur every three years. Based on this, the 322 principal ratings included in this study are estimated to reflect 81% of the population of schools using the WFPL and documenting ratings in MLP.

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 71 schools (22.0%) the principal was new (in their first three years). Since principal turnover is common in the Milwaukee Public Schools (MPS) (Simonaitis, 2013), a disproportionate number (36; 50.0%) of the new principals included in the study were MPS principals. Table 4 presents the descriptive characteristics of the 322 schools included in this study.

Table 4. Descriptive characteristics of 322 study schools with WFPL ratings

	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%

One hundred sixty-two of these schools had at least 40% of all classroom teachers complete the Annual EE Evaluation Survey of Teachers. Within these schools, 2,873 classroom teachers completed surveys, representing a 62% response rate.

Among these 162 schools, 114 (70.3%) were elementary schools, 20 (12.3 %) were middle schools, 22 (13.6%) were high schools, and 6 (3.7%) were combined elementary and secondary schools. In 56 schools (34.6%) the principal was new (in their first three years). Table 5 presents the descriptive characteristics of the 162 schools that also participated in the annual teacher survey.

Table 5. Descriptive characteristics of 162 schools with both WFPL ratings and teacher surveys

	Min	Max	Mean	Std. Dev.
Enrollment	42	1,903	475	254
Percent Students with Disability	5.2%	77.8%	17.5%	7.7%
Percent Economic Disadvantaged Students	2.1%	99.0%	60.7%	26.5%
Percent White Students	0.6%	95.2%	44.9%	33.1%

Analysis Procedures

Hierarchical linear modeling (HLM) (Raudenbush & Bryk, 2002), was used to test the agreement between WFPL ratings and teacher perceptions of principal effectiveness (Principal Leadership, Principal – Teacher Trust, Feedback Quality, Feedback Accuracy, and Evaluator Qualifications) in the 162 schools with both teacher survey and WFPL results. HLM was used because the data have a nested structure, with teachers nested within schools. Three modeling steps were used for each analysis.

First, an unconditional model was used to determine the degree that teacher perceptions of principal effectiveness were related to school factors. The first, or unconditional model, for teacher (i) in school (j) is written as,

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (1)$$

where Y_{ij} represents the Principal Leadership, Principal – Teacher Trust, Feedback Quality, Feedback Accuracy, or Evaluator Qualifications Ratings and is a function of the grand mean of all schools (γ_{00}), differences between schools (u_{0j}), and differences across teachers (r_{ij}).

Second, WFPL ratings were included in the model, to determine its bivariate relationship with each aspect of teacher perceptions of principal effectiveness. This is written as,

$$Y_{ij} = \gamma_{00} + \gamma_{01}(WFPL) + u_{0j} + r_{ij} \quad (2)$$

Third, school and principal characteristics were added to the model to clarify the relationship of WFPL ratings with teacher perceptions of principal effectiveness. This model was expressed as,

$$Y_{ij} = \gamma_{00} + \gamma_{01}(WFPL) + \gamma_{02}(\%White) + \gamma_{03}(HS) + \gamma_{04}(New\ Principal) + u_{0j} + r_{ij} \quad (3)$$

where γ_{02}(White) is the school percentage of white students, γ_{03}(HS) is 1 if the school is a high school, and γ_{04}(New Principal) is 1 if the principal is new to the school.

Intraclass correlations (ICC) were computed to calculate the amount that variability in teacher perceptions of principal effectiveness is associated with schools factors. ICC was defined as,

$$\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2} \quad (4)$$

where τ^2 represents the variance between schools and σ^2 is the variance within schools.

Pseudo R^2 was computed to determine the percent of school-level variance in teacher perceptions of their principal that is explained by each model. *Pseudo* R^2 was expressed as,

$$R^2 = \frac{{\tau^2}_{Model\ 1} - {\tau^2}_{Model\ 2}}{{\tau^2}_{Model\ 1}} \quad (5)$$

Results

WFPL Descriptive Results

Although the majority of principals were rated as proficient on each of the WFPL components, there was more variability on components in the Instructional Leadership domain, with more principals rated as Basic than was observed across the other domains (Table 1). With some exceptions, the individual components were all correlated with each other at between .3 and .5 (Table 6). Taken together, most principals were rated as Proficient overall, with overall WFPL ratings averaging 3.08 (*SD* = .30), and with domain ratings ranging from a low of 3.02 (Instructional Leadership) to a high of 3.10 for both School Management and Personal Behavior (Table 7).

Table 6. Wisconsin Framework for Principal Leadership component correlations (all correlations significant $p < .01$)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	Recruiting and Selecting	1																			
2	Assignment of Teachers and Instructional Staff	.460	1																		
3	Observation and Performance Evaluation	.267	.362	1																	
4	Professional Development and Learning	.281	.309	.343	1																
5	Distributed Leadership	.366	.457	.498	.415	1															
6	Mission and Vision	.350	.390	.370	.410	.413	1														
7	Student Achievement Focus	.344	.425	.480	.508	.521	.424	1													
8	Staff Collaboration	.245	.381	.358	.452	.492	.363	.514	1												
9	School-wide Use of Data	.328	.382	.516	.490	.495	.464	.660	.457	1											
10	Student Learning Objectives (Teacher SLOs)	.323	.364	.335	.370	.297	.345	.375	.373	.344	1										
11	Professionalism	.212	.343	.384	.238	.351	.301	.264	.283	.301	.206	1									
12	Time Management and Priority Setting	.336	.408	.556	.326	.404	.323	.472	.329	.421	.370	.342	1								
13	Use of Feedback for Improvement	.296	.340	.397	.267	.451	.384	.308	.384	.347	.283	.311	.350	1							
14	Initiative and Persistence	.398	.430	.481	.289	.478	.362	.438	.353	.375	.323	.374	.489	.388	1						
15	School Climate	.330	.339	.317	.330	.433	.399	.355	.402	.319	.313	.356	.340	.400	.356	1					
16	Communication	.300	.394	.347	.339	.392	.466	.381	.490	.364	.373	.404	.378	.423	.431	.408	1				
17	Conflict Management and Resolution	.287	.449	.299	.287	.418	.278	.281	.340	.257	.28	.398	.376	.303	.313	.411	.444	1			
18	Consensus Building	.390	.449	.396	.412	.544	.421	.452	.482	.449	.321	.326	.378	.451	.396	.404	.444	.371	1		
19	Learning Environment Management	.334	.389	.248	.260	.318	.333	.301	.381	.252	.347	.254	.321	.253	.385	.364	.356	.283	.284	1	
20	Financial Management	.364	.391	.211	.245	.285	.334	.291	.319	.270	.317	.212	.333	.298	.264	.224	.275	.237	.259	.400	1
21	Policy Management	.201	.225	.187	.135	.224	.115	.148	.129	.185	.244	.244	.261	.196	.300	.157	.281	.219	.210	.342	.236

Table 7. WFPL overall and domain rating descriptive statistics

	Min	Max	Mean	Std. Dev.
Human Resource Leadership	2	4	3.07	0.36
Instructional Leadership	1.8	4	3.02	0.40
Personal Behavior	1.75	4	3.12	0.37
Intentional and Collaborative School Culture	1.5	4	3.06	0.37
School Management	2	4	3.12	0.32
Overall WFPL Rating	1.96	4	3.08	0.30

The distribution of overall WFPL ratings suggests that there was a narrow range of ratings, with the majority rated close to 3.0 (Figure 1). Even considering the narrow range of WFPL ratings, there were still differences in WFPL ratings across types of schools and for new principals compared to experienced principals (Table 8). New principals were rated 0.28 points lower on average, nearly a full standard deviation ($z = .93$), than experienced principals ($F = 42.87$, $df = 1$, $p < 0.001$). No differences in ratings were found between principals in different types of schools ($F = 1.75$, $df = 3$, $p = 0.16$).

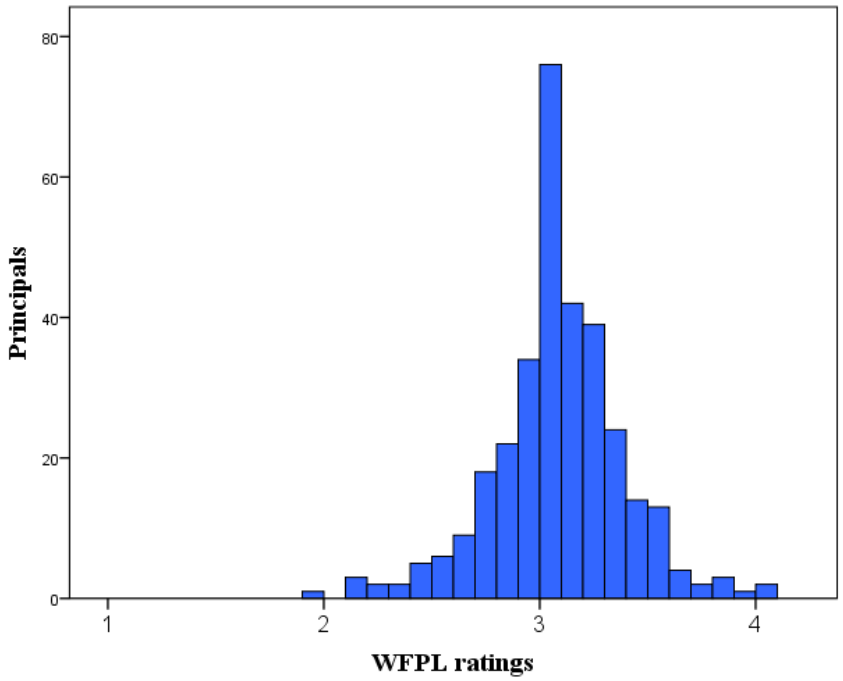

Figure 1: Distribution of Overall WFPL Ratings

Table 8. WFPL ratings comparisons by school type and principal tenure

	Mean	Std. Dev.	*n*
Principal tenure			
Experienced	3.13	0.28	251
New (First 3 years)	2.88	0.31	71
School Type			
Combined Elementary/Secondary School	3.07	0.41	13
Elementary School	3.08	0.30	210
High School	3.12	0.28	56
Middle School	2.99	0.33	43

Correlations of Administrator Ratings (WFPL) with Teacher Ratings of Principal Effectiveness

As discussed previously, teachers rated the effectiveness of principals across five areas. The descriptive statistics for these scales are presented in Table 9. Correlations of the five measures of teacher perceptions of principal effectiveness with WFPL Ratings (Table 10) suggest that principals with higher overall WFPL ratings were rated as more effective by teachers on Principal Leadership ($r = 0.36$, $p < 0.001$), Principal – Teacher Trust ($r = 0.34$, $p < 0.001$), Feedback Accuracy ($r = 0.26$, $p = 0.001$) and Evaluator Qualifications ($r = 0.26$, $p = 0.001$). However, overall WFPL ratings were not related with Feedback Quality ($r = 0.15$, $p = 0.05$). Except for School Management, WFPL domain ratings were also related with measures of teacher perceptions of principal effectiveness.

Table 9. Teacher ratings of principal effectiveness – descriptive statistics

	Mean	Std. Dev.	n
Principal Leadership	2.93	0.70	2,873
Principal – Teacher Trust	2.91	0.79	2,872
Feedback Quality	2.78	0.92	1,942
Feedback Accuracy	3.05	0.87	1,924
Evaluator Qualifications	3.18	0.84	1,915

Table 10. School and principal correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Human Resource Leadership Rating	-													
2. Instructional Leadership Rating	.80**	-												
3. Personal Behavior Rating	.75**	.65**	-											
4. Intentional & Collaborative School Culture Rating	.72**	.71**	.65**	-										
5. School Management Rating	.46**	.49**	.51**	.50**	-									
6. Overall EE Evaluation Rating	.90**	.88**	.85**	.86**	.69**	-								
7. Principal Leadership	.39**	.39**	.26**	.36**	.08	.36**	-							
8. Principal – Teacher Trust	.36**	.36**	.22**	.36**	.10	.34**	.91**	-						
9. Feedback Quality	.17*	.16*	.09	.16*	.08	.15	.57**	.50**	-					
10. Feedback Accuracy	.27**	.30**	.14	.26**	.11	.26**	.60**	.64**	.62**	-				
11. Evaluator Qualifications	.25**	.26**	.17*	.28**	.13	.26**	.71**	.71**	.71**	.76**	-			
12. Percent students with disabilities	-.09	-.25**	-.03	-.01	.10	-.08	-.23**	-.20**	.10	-.11	-.07	-		
13. Percent students Econ. Disadv.	-.27**	-.36**	-.29**	-.18*	-.20*	-.32**	-.20*	-.25**	.10	-.20*	-.08	.42**	-	
14. Percent White students	.30**	.41**	.31**	.20*	.23**	.36**	.27**	.34**	-.02	.27**	.18*	-.34**	-.88**	-
15. Enrollment	-.01	-.03	.02	.02	.01	.003	-.13	-.12	.03	-.08	-.02	-.06	.01	-.13

Note: * = $p < 0.05$, ** = $p < 0.0$

HLM Results

Although the correlation results provide evidence that WFPL ratings are associated with teacher perceptions of principal effectiveness, the nested nature of the data and the covariance of ratings with school and principal characteristics cloud the relationship between them. Hierarchical linear modeling was used to clarify their relationships. As outlined previously, each of the following analyses follow three steps: (1) an unconditional model that establishes the amount of variance explained by school factors, (2) a simple model that establishes the amount of shared variance in teacher perceptions of principal effectiveness with WFPL ratings, and (3) a full model that adjusts the relationship of WFPL ratings and teachers perceptions by including principal and school characteristics.

Principal Leadership. The results of the unconditional model (Model 1; Table 11) suggest that 26.0% of the total variation in teacher perceptions of Principal Leadership was attributed to differences between schools ($\chi^2 = 1174.3$, $df = 158$, $p < 0.001$). Model 2 results suggest that higher overall WFPL ratings were related with higher teacher perceptions of Principal Leadership ($t = 5.28$, $df = 157$, $p < 0.001$); a one-point increase in WFPL rating corresponded with a 0.47 point increase in teacher perceptions of Principal Leadership. WFPL ratings explained 16.4% of the school variance in teacher perceptions of Principal Leadership (*Pseudo* $R^2 = 0.164$). This positive relationship remained in Model 3 after adjusting for school characteristics and principal experience ($t = 3.79$, $df = 154$, $p < 0.001$).

Principal – Teacher Trust. The results of the unconditional model (Model 1; Table 12), suggest that 26.4% of the total variation in teacher perceptions of Principal – Teacher Trust was attributed to the differences between schools ($\chi^2 = 1182.5$, $df = 158$, $p < 0.001$). Model 2 results suggest that higher overall WFPL ratings were related with higher teacher perceptions of Principal – Teacher Trust ($t = 4.97$, $df = 157$, $p < 0.001$); a one point increase in WFPL rating corresponded with a 0.50 point increase in teacher perceptions of Principal – Teacher Trust. WFPL ratings explained 10.1% of the school variance in teacher perceptions of Principal – Teacher Trust (*Pseudo* $R^2 = .101$). This positive relationship remained in Model 3 after adjusting for school characteristics and principal experience ($t = 3.09$, $df = 154$, $p = 0.002$).

Feedback Quality. The results of the unconditional model (Model 1; Table 13) suggest that only 8.7% of the total variation in teacher perceptions of Feedback Quality was attributed to

differences between schools ($\chi^2 = 338.8$, $df = 158$, $p < 0.001$). Model 2 results suggest that higher overall WFPL ratings were related with higher teacher perceptions of Feedback Quality ($t = 2.41$, $df = 157$, $p = 0.017$); a one-point increase in the WFPL ratings corresponded with a 0.21 point increase in teacher perceptions of Feedback Quality. WFPL ratings explained only 2.7% of the school variance in teacher perceptions of Feedback Quality (*Pseudo* $R^2 = 0.027$), suggesting that ratings were only weakly related with teacher perceptions of feedback quality. However, this positive relationship remained in Model 3 after adjusting for school characteristics and principal experience ($t = 3.32$, $df = 154$, $p = 0.001$).

Feedback Accuracy. The results of the unconditional model (Model 1; Table 14) suggest that 10.4% of the total variation of teacher perceptions of Feedback Accuracy was attributed to differences between schools ($\chi^2 = 370.7$, $df = 158$, $p < 0.001$). Model 2 results suggest that higher WFPL ratings were related to higher teacher perceptions of Feedback Accuracy ($t = 3.94$, $df = 157$, $p < 0.001$); a one-point increase in WFPL ratings corresponded with a 0.31-point increase in the teacher perceptions of Feedback Accuracy. WFPL ratings explained 10.9% of the school variance in teacher perceptions of Feedback Accuracy (*Pseudo* $R^2 = 0.109$). This positive relationship remained in Model 3 after adjusting for school characteristics and principal experience ($t = 2.36$, $df = 154$, $p = 0.02$).

Evaluator Qualifications to Provide Feedback. The results of the unconditional model (Model 1; Table 15) suggest that 12.6% of the total variation of teacher perceptions of Evaluator Qualifications was attributed to differences between schools ($\chi^2 = 411.7$, $df = 158$, $p < 0.001$). Model 2 results suggest that there was a positive relationship between overall WFPL ratings and teacher perceptions of Evaluators Qualifications to Provide Feedback ($t = 3.43$, $df = 157$, $p < 0.001$); a one point increase in WFPL ratings corresponded with a 0.35 point increase in teacher ratings of Evaluator Qualifications. Overall WFPL Ratings explained 12.2% of the school variance in teacher perceptions of Evaluator Qualifications (*Pseudo* $R^2 = 0.122$). This positive relationship remained in Model 3 after adjusting for school characteristics and principal experience (Model 3).

Table 11. Principal Leadership HLM Results

	Model 1		Model 2		Model 3	
Fixed Effects	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept γ_{00}	2.98**	0.03	2.98**	0.03	2.99**	0.03
WFPL Ratings γ_{01}			0.47**	0.09	0.40**	0.11
% White students γ_{02}					0.002*	0.001
School Type (1 = high school) γ_{03}					-0.28**	0.09
New Principal (1 = new) γ_{04}					-0.04	0.07
Random Effects	Component	χ^2 (df)	Component	χ^2 (df)	Component	χ^2 (df)
u_{0j}	0.12644	1,174.3 (158)**	0.10576	1,021.2 (157)**	0.09378	861.9 (154)**
r_{ij}	0.35985		0.35990		0.36001	

Note: * = p < .05, ** = p < 0.01

Table 12. Principal – Teacher Trust HLM Results

	Model 1		Model 2		Model 3	
Fixed Effects	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept γ_{00}	2.97**	0.04	2.97**	0.03	2.97**	0.03
WFPL Ratings γ_{01}			0.50**	0.10	0.37**	0.12
% White students γ_{02}					0.003**	0.001
School Type (1 = high school) γ_{03}					-0.27*	0.11
New Principal (1 = new) γ_{04}					-0.06	0.08
Random Effects	Component	χ^2 (df)	Component	χ^2 (df)	Component	χ^2 (df)
u_{0j}	0.16523	1,182.5 (158)**	0.14185	1,044.9 (157)**	0.12430	885.8 (154)**
r_{ij}	0.46049		0.46055		0.46063	

Note: * = p < .05, ** = p < 0.01

Table 13. Feedback Quality HLM Results

	Model 1		Model 2		Model 3	
Fixed Effects	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept γ_{00}	2.78**	0.03	2.78**	0.03	2.78**	0.03
WFPL Ratings γ_{01}			0.21*	0.10	0.33**	0.10
% White students γ_{02}					-0.001	0.001
School Type (1 = high school) γ_{03}					-0.22*	0.09
New Principal (1 = new) γ_{04}					0.07	0.07
Random Effects	Component	χ^2 (df)	Component	χ^2 (df)	Component	χ^2 (df)
u_{0j}	0.07368	338.8 (158)**	0.07173	334.6 (157)**	0.06422	310.0 (154)**
r_{ij}	0.76950		0.76886		0.76969	

Note: * = p < .05, ** = p < 0.01

Table 14. Feedback Accuracy HLM Results

	Model 1		Model 2		Model 3	
Fixed Effects	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept γ_{00}	3.06**	0.03	3.07**	0.03	3.08**	0.03
WFPL Ratings γ_{01}			0.31**	0.08	0.22*	0.09
Percent White students γ_{02}					0.003**	0.001
School Type (1 = high school) γ_{03}					-0.26**	0.09
New Principal (1 = new) γ_{04}					0.01	0.07
Random Effects	Component	χ^2 (df)	Component	χ^2 (df)	Component	χ^2 (df)
u_{0j}	0.07931	370.7 (158)**	0.07137	350.7 (157)**	0.05158	292.1 (154)**
r_{ij}	0.68042		0.68006		0.68151	

Note: * = p < .05, ** = p < 0.01

Table 15. Evaluator Qualifications HLM Results

	Model 1		Model 2		Model 3	
Fixed Effects	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept γ_{00}	3.20**	0.03	3.20**	0.03	3.20	0.03
WFPL Ratings γ_{01}			0.35**	0.10	0.36**	0.12
Percent White students γ_{02}					0.002*	0.001
School Type (1 = high school) γ_{03}					-0.31**	0.09
New Principal (1 = new) γ_{04}					0.05	0.07
Random Effects	Component	χ^2 (df)	Component	χ^2 (df)	Component	χ^2 (df)
u_{0j}	0.09041	411.7 (158)**	0.07935	386. 6 (157)	0.06700	344.1 (154)**
r_{ij}	0.62606		0.62598		0.62626	

Note: * = p < .05, ** = p < 0.0

Table 16 summarizes the results of the HLM models. Although all tested aspects of teacher perceptions of principal effectiveness proved to be related with WFPL ratings, the two scales from the 5Essentials Survey, Principal Leadership and Principal-Teacher trust, proved to be more variable across schools, and the most closely associated with WFPL ratings.

Table 16. Summary results for models predicting teacher perceptions of principal effectiveness conditioned on WFPL ratings

Outcome	*ICC (Percent of variance attributed to school)*	*B (WFPL ratings)*	School *Pseudo* R^2*(after including WFPL ratings)*
Principal Leadership	26.0%	0.47	16.5%
Principal – Teacher Trust	26.4%	0.50	10.1%
Feedback Quality	8.7%	0.21	2.7%
Feedback Accuracy	10.4%	0.31	10.9%
Evaluator Qualifications	12.6%	0.35	12.2%

Note: ICC = Intraclass Correlation Coefficient; Pseudo R^2 was calculated between Model 2 and Model 1 for all analyses; B is the corresponding increase in teacher perceptions for every 1 point increase in Overall WFPL Ratings.

Summary Findings

The purpose of this study was to examine the characteristics of Wisconsin Framework for Principal Leadership (WFPL) ratings assigned to principals as part of the Wisconsin EE process, and to determine the extent that ratings demonstrated concurrent validity with teacher perceptions of principal effectiveness. The results suggest a number of important findings. First, there was little variability in principal effectiveness ratings, with most principals rated as Proficient. Second, component and domain ratings were found to be related to each other, mostly, as expected. Third, the results of HLM models suggest that WFPL ratings were consistent with teacher perceptions of principal effectiveness. Teachers in schools where their principal was rated as more effective by their administrator, also rated their principal as a more effective leader, had more trust in their principal, felt that their principal was more qualified to provide them feedback on their instructional practice, and felt that the feedback they provided was better and more accurate. Thus, these results suggest that WFPL ratings, assigned by

administrators, have a high degree of concurrent validity with teacher perceptions of principal effectiveness.

Discussion and Implications of Findings

The DPI's primary purpose for developing the principal evaluation process using the WFPL was to provide districts with a valid measure of school leadership that also could provide guidance for how to support the development and growth of principals. This study presents promising findings that suggest DPI is making progress toward this goal. Given the strong relationship between WFPL ratings and teacher perceptions of principal effectiveness, the study findings suggest that the WFPL ratings measure critical aspects of principal leadership as experienced by teachers.

Given the close relationship of WFPL ratings to these culture and climate factors, it is reasonable to expect that schools with principals who improve according to the WFPL, will experience improvements in school culture and climate, which may, in turn, result in improvements to school performance. Research on school climate and culture bears this out (Bryk et. al., 2010; Bryk & Schneider, 2002). Future longitudinal research will explore this in Wisconsin schools.

One potential use of WFPL ratings in Wisconsin, could be to inform efforts to improve educator equity across Wisconsin. Although not all districts use the WFPL, most do, and all of the large urban schools districts do. Ratings could therefore be used to quantify the extent that low-income and diverse schools have unequal access to effective principals and may help inform Wisconsin's efforts to improve access to quality principals to low-income and diverse schools. Future research will explore this potential use of WFPL ratings for this purpose.

Limitations

There are several limitations to this study worth considering. One limitation is that it was unclear if principals were always the evaluators of teachers who completed the survey. Thus, some survey responses may have been directed at assistant or associate principals. This could help explain why WFPL ratings were less related with perceptions of feedback than they were with perceptions of principals. Another limitation is that the study schools reflect a relatively small sample of schools across the state, and it is unknown how well these results generalize to the whole state. However, it was not expected that the data presented would be representative of the

state. As part of the EE process, new principals receive ratings at the end of their first year, while experienced principals are rated once every three years. Thus, this study includes a disproportionate number of new principals. However, it was not the goal of this study to establish norms of principal effectiveness ratings for the state. The goal was to test the relationships of WFPL ratings with teacher ratings of principal effectiveness. Even with these limitations, the findings have significant implications for the Wisconsin education system and provide promising evidence that the DPI has developed a rubric and process that helps assess principal effectiveness and supports efforts to develop principals.

References

Anderson, L. M., & Turnbull, B. J. (2016). *Evaluating and Supporting Principals.* New York: The Wallace Foundation.

Bryk, A. S., Sebring, P., Allensworth, E., Luppescu, S, & Easton, J. (2010). *Organizing Schools for Improvement: Lessons from Chicago*. University of Chicago Press.

Bryk, A.S., & Schneider, B. (2002). *Trust in Schools: A Core Resource for Improvement.* New York, NY: Russell Sage Foundation.

Cherasaro, T. L., Brodersen, R. M., Yanoski, D. C., Welp, L. C., & Reale, M. L. (2015). *The Examining Evaluator Feedback Survey* (REL 2016–100). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from http:// ies.ed.gov/ncee/edlabs.

Chetty, R., Freidman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 263-79.

Dee, T., & Wykoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management, 34*(2). 267-297.

Goldring, E., Cravens, X. C., Murphy, J., Porter, A. C., Elliott, S. N., & Carson, B. (2009). The evaluation of principals: What and how do states and urban districts assess leadership? *The Elementary School Journal, 110*, 19-39.

Hallinger, P., & Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly, 32*(1), 5-44.

Hanushek, E. (2012). *The Economic Value of Higher Teacher Quality.* Working Paper No.

16606). Cambridge, MA: National Bureau of Economic Research.

Kimball, S. M., Milanowski, A., & McKinney, S. A. (2009). Assessing the promise of standards-based performance evaluation for principals: Results from a randomized trial. *Leadership and Policy in Schools, 8*, 233-263.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). *Value-added modeling: A Review. Economics of Education Review, 47*, 180-195.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K., (2004). *Review of Research: How Leadership Influences Student Learning.* New York: The Wallace Foundation.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School Leadership that Works: From Research to Results.* Alexandria, VA: ASCD.

McCullough, M., Lipscomb, S., Chiang, H., Gill, B., & Cheban, I. (2016). *Measuring School Leaders' Effectiveness: Final Report from a Multiyear Pilot of Pennsylvania's Framework for Leadership*.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Vol. 1*). Sage.

Simonaitis, P. (August 14, 2013). MPS heads into another school year with high principal turnover. *Milwaukee Journal Sentinel.* Retrieved from http://archive.jsonline.com/news/education/mps-heads-into-another-school-year-with-high-principal-turnover-b9972521z1-219658581.html/

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher, 30*(3), 23-28.

Stronge, J. H., Richard, H. B., & Catano, N. (2008). *Qualities of Effective Principals*. Alexandria, VA: Association for Supervision & Curriculum Development.

Wisconsin (2012). *2011 Wisconsin Act 166*. Retrieved from https://docs.legis.wisconsin.gov/2011/related/acts/166.pdf