The Effect of Gender Stereotypes and Congruence in Principal Evaluation of Teacher Effectiveness

Kurtis Jensen

NEE Director of Member Services

University of Missouri

jensenkc@missouri.edu

Dr. Eli Jones

NEE Research Specialist

University of Missouri

joneseli@missouri.edu

The Effect of Gender Stereotypes and Congruence in Principal Evaluation of Teacher Effectiveness

Background

With increased federal pressure on schools to raise the importance of evaluation on teacher effectiveness and personnel decisions (Sawchuk, 2013), it is vital for principals to accurately score teachers based primarily on the teacher's quality of instruction. Unfortunately, evaluation based on rater judgments can be subjective, which may lead to inaccurate scores (Bailey, Bocala, Shakman, & Zweig, 2016; Grissom, Nicholson-Crotty, & Keiser, 2012; Grissom & Loeb, 2017; Jacob & Lefgren, 2008).

Two widely recognized sources of bias that may impact teacher evaluation scores are "gender congruence" and "gender role congruence." Gender congruence theory refers to the impact same-gender relationships have within a social setting (Pedersen, 2016). In the teacher evaluation context, gender congruence refers to the principal and teacher sharing the same gender. Gender role congruence theory suggests that when the occupation is considered to be feminine or masculine, the correlating gender tends to be rated higher (Eagly, Karau, & Johnson, 1992). While studies provide evidence that gender congruence and gender role congruence impact the outcome of employee evaluations (Swim, Borgida, Maruyama, & Myers, 1989), this phenomenon has not been well investigated in a teacher evaluation context.

Purpose

The current study will extend the existing research from non-educational settings to a teacher evaluation setting to test the hypothesis that gender congruence among principals and teachers will have an impact on teacher observation scores. The study will explore the following questions:

1. How does teacher gender influence teacher observation scores?
2. Is there an interaction effect between the gender of the principal and the gender of the teacher?
3. How does the influence of teacher gender compare to other predictors of teacher ratings, including first-year status and prior-year scores?

Setting

Data are from the Network for Educator Effectiveness (NEE), an outreach unit of the University of Missouri that provides schools a structured framework for conducting teacher evaluations in Missouri. The state of Missouri consists of 538 public school districts, 38 charter schools and 660 private schools (Missouri Department of Elementary and Secondary Education, 2016; Private School Review, 2016). The districts consist of small rural to large urban with 3,611 principals and 68,189 teachers.

Sample

Participants in this study include classroom observation scores from 1,318 mathematics and science teachers who were rated by 396 principals using the NEE Classroom observation Rubric (COR, see below). Ratings of 590 elementary and 728 secondary school teachers took place during the 2015-2016 and 2016-2017 school years. Most teachers were female (75%) while most principals were male (56.9%). Principals included 194 elementary schools and 201 secondary school principals.

Instrument

The NEE Classroom Observation Rubric (COR) is a web-based program utilized by principals while observing teacher instruction. The COR uses an eight-point Likert scale ranging from "0: (not present) to "7" (perfect exemplar). The rubric provides anchor points at zero, one, three, five and seven. Principals observe teachers on average four times per school year, for 10-20 minutes each observation. Each observation typically focuses on 3-4 specific teaching practices adapted by NEE from the Interstate New Teacher Assessment and Support Consortium (INTASC, 2005) teaching standards.

Research Design

This particular paper adds to the current literature by exploring the role gender plays in how principals rate teachers during classroom observations. We present the results of a multiple regression in which we predict a teacher's classroom observation score on three specific teaching practices that were used by principals to evaluate teachers during the 2015-2016 and 2016-2017 school years: (a) use of formative assessment (FA), (b) promotion of critical thinking (CT), and (c) cognitive engagement strategies (CE).

Data Collection and Analysis

We applied hierarchical multiple regression to predict a teacher's average classroom observation score in 2017. As covariates, we include: (a) the teacher's gender, (b) the teachers' prior-year average CO score for the teaching practice in question, (c) the teacher's level of experience, (d) the principal's gender, and (e) the interaction between the teacher's gender and the principal's gender, sot that:

$$\gamma_i = \beta_0 + \beta_1 t_i + \beta_2 p_i + \beta_3 t_i p_i + \beta_4 s_i + \beta_5 x_i + \varepsilon_i \,, \qquad (1)$$

where γ_i is the 2017 average classroom observation score for teacher *i*, t_i is the gender of teacher *i*, $p_i$is the gender of the principal of teacher *i*, $t_i p_i$ is the interaction between teacher gender and principal gender, $s_i$is the prior year average classroom observation score for teacher *i*, $x_i$is a dichotomous variable indicating first- or second-year status for teacher *i*, and ε_i is the error term.

Results

Table 1 shows the summary statistics for the sample. The results of the regression models are presented in Table 1 (CE), Table 2 (CT), and Table 3 (FA). Across all models, teacher gender explained a significant amount in the current year COR score (CE: ($F(2, 1093) = 137.34$, $p < .001$, $R^2 = .201$, $R^2_{\text{Adjusted}} = .199$; CT: ($F(2, 1093) = 152.68$, $p < .001$, $R^2 = .218$, $R^2_{\text{Adjusted}} = .217$; FA: ($F(2, 1093) = 93.15$, $p < .001$, $R^2 = .146$, $R^2_{\text{Adjusted}} = .144$), suggesting that male teachers consistently received lower scores than female teachers. Principal gender was a significant predictor of current COR score for FA (($F(2, 1093) = 44.88$, $p < .001$, $R^2 = .149$, $R^2_{\text{Adjusted}} = .147$) in the presence of teacher gender and prior year COR score. The interaction between principal gender and teacher gender was not significant in the presence of the other predictors.

Conclusion

The results suggest that gender congruence did not play a significant role in how principals rated teachers. Conversely, gender role congruence may have a significant impact on ratings. Specifically, we found that male teachers consistently received lower scores than their female counterparts across all teaching practices. One hypothesized reason for this is the implicit perception that teaching is more of a female role (Basten, 1997; Drudy, 2008; Grissom et al., 2012). Nevertheless, prior year performance was still the strongest predictor of current-year performance in the presence of gender and teacher level of experience. The full paper will further explore this phenomenon and discuss the implications for practice and potential future research.

References

Bailey, J., Bocala, C., Shakman, K., & Zweig, J. (2016). Teacher demographics and evaluation: A descriptive study in a large urban district (REL 2017–189). *Institute of Education Sciences.* Washington, DC. Retrieved from http://ies.ed.gov/ncee/edlabs.

Basten, C. (1997). A feminised profession: women in the teaching profession. *Educational Studies*, *23*(1), 55-62.

Drudy, S. (2008). Gender balance/gender bias: The teaching profession and the impact of feminisation. *Gender and Education*, *20*(4), 309-323.

Eagly, A. H., Karau, S. J., & Johnson, B. T. (1992). Gender and leadership style among school principals: A meta-analysis. *Educational Administration Quarterly*, *28*(1), 76-102.

Grissom, J.A., & Loeb, S. (2017). Assessing principal's assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy.*

Grissom, J. A., Nicholson-Crotty, J., & Keiser, L. (2012). Does my boss's gender matter? Explaining job satisfaction and employee turnover in the public sector. *Journal of Public Administration Research and Theory*.

Interstate New Teacher Assessment and Support Consortium. (2005). Interstate New Teacher Assessment and Support Consortium Standards. *Washington, DC. Council of Chief State Schools Officers. Retrieved*, *4*(18), 05.

Pedersen, M. J., & Nielsen, V. L. (2016). Manager–Employee Gender Congruence and the Bureaucratic Accountability of Public Service Employees: Evidence From Schools. *Public Personnel Management*, *45*(4), 360-381.

Sawchuk, S., & Michigan, I. (2013). Teachers' ratings still high despite new measures. *Education Week*, *32*(20), 18-19.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations?. *Psychological Bulletin*, *105*(3), 409.

Table 1. Summary Statistics by Gender Grouping and Teaching Practice

		Current Year (2016-2017)						Prior Year (2015-2016)					
		Cognitive Engagement		Critical Thinking		Formative Assessment		Cognitive Engagement		Critical Thinking		Formative Assessment	
	N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Teacher													
Male	284	5.28	0.89	4.87	1.04	5.03	0.94	4.92	1.17	4.63	1.31	4.68	1.48
Female	812	5.57	0.81	5.17	0.91	5.36	0.88	5.31	1.14	4.89	1.18	5.07	1.34
Principal													
Male	212	5.42	0.84	5.04	0.96	5.19	0.90	5.16	1.10	4.72	1.29	4.88	1.32
Female	183	5.59	0.82	5.17	0.94	5.39	0.91	5.37	1.21	4.98	1.10	5.08	1.47

Table 2. Hierarchical Regression Summary for Predictors of Teachers' Average Classroom Observation score on Cognitive Engagement (*N* = 1,096)

	Model 1			Model 2			Model 3			Model 4		
Predictor	B	SE	*β*	B	SE	*β*	B	SE	*β*	B	SE	*β*
Prior Year Score	0.308	0.020	0.424**	0.306	0.020	0.421**	0.307	0.020	0.427**	0.297	0.020	0.408**
Male (*teacher)*	-0.208	0.052	-0.109**	-0.193	0.053	-0.101**	-0.240	0.100	-0.126**	-0.240	0.099	-0.125**
Male (*principal*)				-0.063	0.047	-0.037	-0.076	0.053	-0.045	-0.091	0.052	-0.053
$Male_T$ x $Male_P$							0.065	0.118	0.031	0.065	0.117	0.031
Years of Experience												
2-3 vs. 1-2										0.158	0.096	0.058
4-9 vs. 1-2										0.206	0.080	0.109*
10-15 vs. 1-2										0.283	0.084	0.134**
15+ vs. 1-2										0.294	0.078	0.163**
R2		0.201			0.202			0.202			0.214	
Adj. R2		0.199			0.200			0.199			0.208	
FΔ for R2		16.11**			1.77			0.30			.012**	

Note: ** $p < .01$, * $p < .05$.

Table 3. Hierarchical Regression Summary for Predictors of Teachers' Average Classroom Observation score on Critical Thinking (N = 1,096)

	Model 1			Model 2			Model 3			Model 4		
Predictor	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Prior Year Score	0.350	0.021	0.449**	0.351	0.021	0.449**	0.351	0.021	0.449**	0.342	0.021	0.438**
Male (*teacher)*	-0.208	0.058	-0.096**	-0.210	0.060	-0.096**	-0.239	0.113	-0.110**	-0.243	0.112	-0.111**
Male (*principal*)				0.006	0.053	0.003	-0.002	0.059	-0.001	-0.016	0.059	-0.008
$Male_T$ x $Male_P$							0.041	0.133	0.017	0.065	0.117	0.031
Years of Experience												
2-3 vs. 1-2										0.004	0.109	0.001
4-9 vs. 1-2										0.192	0.091	0.089*
10-15 vs. 1-2										0.227	0.095	0.095*
15+ vs. 1-2										0.274	0.089	0.134**
R2		0.218			0.218			0.218			0.230	
Adj. R2		0.217			0.216			0.216			0.224	
$F\Delta$ for R2		12.68**			0.01			0.09			3.93**	

Note: ** $p < .01$, * $p < .05$.

Table 4. Hierarchical Regression Summary for Predictors of Teachers' Average Classroom Observation score on Formative Assessment (N = 1,096)

	Model 1			Model 2			Model 3			Model 4		
Predictor	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Prior Year Score	0.238	0.018	0.363**	0.226	0.018	0.345**	0.228	0.018	0.348**	0.221	0.019	0.337**
Male (*teacher)*	-0.249	0.058	-0.120**	-0.222	0.060	-0.107**	-0.349	0.112	-0.168**	-0.354	0.111	-0.171**
Male (*principal*)				0.114	0.053	-0.062*	-0.149	0.059	-0.081*	-0.158	0.059	-0.086**
$Male_T$ x $Male_P$							0.178	0.132	0.078	0.187	0.132	0.082
Years of Experience												
2-3 vs. 1-2										0.028	0.108	0.009
4-9 vs. 1-2										0.114	0.090	0.056
10-15 vs. 1-2										0.093	0.095	0.041
15+ vs. 1-2										0.250	0.088	0.128**
R2		0.146			0.149			0.151			0.160	
Adj. R2		0.144			0.147			0.148			0.154	
$F\Delta$ for R2		18.16**			4.63*			1.81			3.08*	

Note: ** $p < .01$, * $p < .05$.