

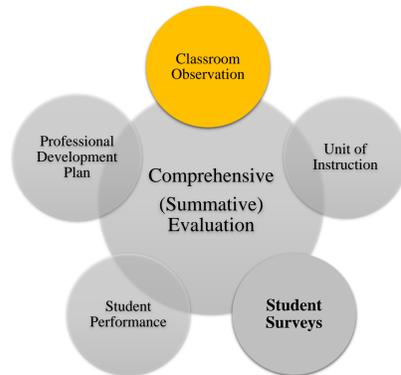
INTRODUCTION

Objective

To test the hypothesis that gender congruence among principals and teachers will have an impact on teacher observation scores.

Setting

Data are from the Network for Educator Effectiveness (NEE), an outreach unit of the University of Missouri that provide Missouri schools a structured framework for conducting teacher evaluations in over 270 rural, urban, and suburban districts.



Sample

- Data are from 5,791 teachers rated by 584 principals using the NEE Classroom Observation Rubric.
- This included ratings of 2,962 elementary and 2,829 secondary teachers during the 2016-2017 school year.
- Teachers were 78% female and principals were 52% male.

RESULTS

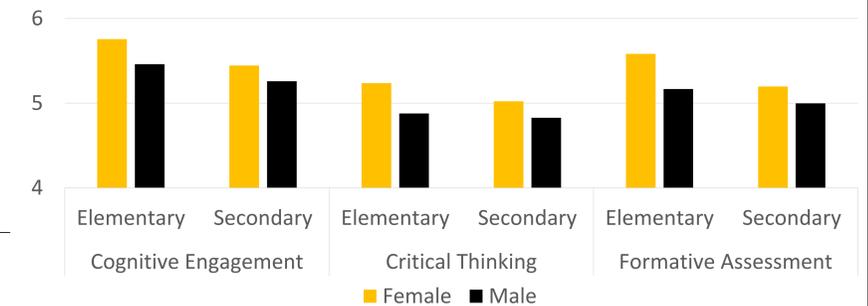
Multiple Regression

- Across all models, teacher gender explained a significant amount of the average Classroom Observation Score (COR) suggesting that male teachers consistently received lower scores than female teachers.
- Principal gender was a significant predictor of current COR average score, as was teacher experience and school type.
- The interaction between principal gender and teacher gender was not significant in the presence of the other predictors.

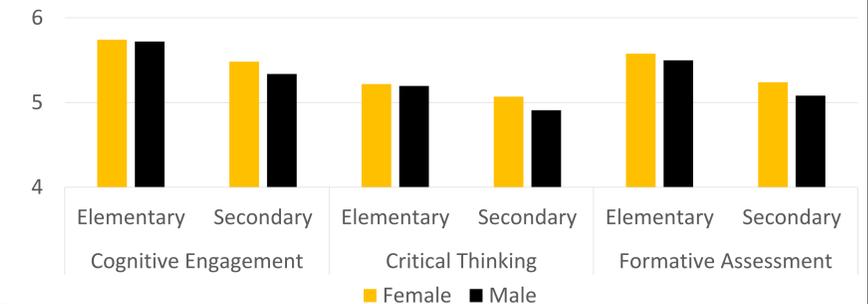
	Model 1	Model 2	Model 3	Model 4
Constant	5.71 (0.02)**	5.77 (0.02)**	5.77 (0.02)**	5.49 (0.03)**
Male (teacher)	-0.30 (0.03)**	-0.20 (0.03)**	-0.20 (0.05)**	-0.18 (0.05)**
Male (principal)	-0.17 (0.02)**	-0.06 (0.02)**	-0.06 (0.03)*	-0.06 (0.03)*
Is secondary school		-0.27 (0.03)**	-0.27 (0.03)**	-0.28 (0.03)**
Male _t X Male _p			0.00 (0.06)	-0.02 (0.06)
Teacher Experience ^a				
2 to 3 years				0.21 (0.04)**
4 to 9 years				0.30 (0.04)**
10 to 14 years				0.35 (0.04)**
>15 years				0.40 (0.03)**
R ²	0.04	0.06	0.06	0.08
Adj. R ²	0.04	0.06	0.06	0.08

	Teachers		Principals	
	Elementary	Secondary	Elementary	Secondary
Female	93%	63%	69%	30%
Male	7%	27%	31%	70%

Average Score by Teacher Gender



Average Score by Principal Gender



METHODS

NEE Classroom Observation Rubric

- The NEE Classroom Observation Rubric 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 receive annual calibration training in the use of the rubric.
- 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.

Analysis Strategy

We applied hierarchical multiple regression to predict a teacher’s average classroom observation score in 2017 across three teaching practices: (a) formative assessment (FA), (b) critical thinking (CT), (c) cognitive engagement strategies (CE), so that:

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

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 providing the observation for teacher i , $t_i p_i$ is the interaction between teacher gender and principal gender, s_i is a dummy variable for secondary school assignment, x_i is a vector of dummy-coded variables for years of experience for teacher i , and ε_i is the error term.

DISCUSSION

- 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: male teachers consistently received lower scores than their female counterparts across all teaching practices. However, this effect was quite small ($R^2 < .10$).
- 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). Implications are that principal OTP scores may be minimally influenced by gender role biases, although not enough to cause significant differences.
- Further research is needed to test the competing hypothesis that the male teachers in the sample were in reality less effective. Additional data not hypothesized to be affected by gender, such as value-added scores, may be a potential next step.