

Inclusive Instructional Leadership: A Quasi-Experimental Study of a Professional Development Program for Principals

Craig Howley

Aimee Howley

WordFarmers Associates LLC

Jacqueline Yahn

Ohio University

Pamela VanHorn

Deborah Telfer

University of Cincinnati

This quasi-experiment tested whether or not a statewide professional development program for principals yielded measurable changes in self-reported attitudes and practices. The two-year program combined training events at central and regional locations with on-site coaching. The program attempted (1) to change attitudes toward inclusion, broadly understood and grounded in social justice, and (2) to cultivate practices constituting inclusive instructional leadership. The program positions the practices as efforts to serve all students well (i.e., including students of color, English Language Learners, and students with disabilities). This quasi-experiment used a post-only nonequivalent group design with propensity score matching to equate treatment and control groups. Dependent measures were principals' attitudes towards inclusion and self-reported use of inclusive instructional leadership practices. Treatment group members included 56 participating principals; control group members were 56 non-participating principals matched (one-to-one matching) via the R optimal match routine. Comparison of attitude scores exhibited a statistically significant effect favoring the treatment group ($ES=0.47$). Also observed was a statistically significant effect ($ES=.38$) for one of the nine practice items—principals' work with teachers on collaborative problem solving and professional learning. None of the observed values for the 21 items (i.e., across both scales, attitudes and practices) favored the control group.

Introduction

This article describes the findings of a quasi-experiment to measure the effectiveness of a statewide professional development program for principals. Ohio Leadership for Inclusion, Implementation, and Instructional Improvement (OLi⁴) is a two-year professional development (PD) program for principals in Ohio. The program began in the fall of 2014 and has expanded each year since then. OLi⁴ recently enrolled its fifth statewide cohort and, for the first time in

2017-18, is delivering the program to a large urban district.¹ OLi⁴ grounds its work on evidence showing that effective instructional leadership, especially from principals, helps schools improve teaching, collective efficacy, and student performance (Leithwood & Seashore-Louis, 2011; Marzano, Waters, & McNulty, 2005; Riehl, 2000). It also expands the construct, *instructional leadership*, to incorporate a focus on social justice and equity (Kearney, 2011; Theoharis & Causton-Theoharis, 2008). OLi⁴ receives support from the Ohio Department of Education's special education and general education offices.

The quasi-experiment asked if there were differences in the attitudes and practices of principals participating in the program compared to nonparticipants, as follows in these research questions:

1. Do OLi⁴ principals score higher on attitudes towards inclusion as compared to nonparticipant Ohio principals?
2. Do OLi⁴ principals score higher on practices of inclusive instructional leadership as compared to nonparticipant Ohio principals?
3. Do OLi⁴ principals score higher on any particular practice of inclusive instructional leadership as compared to nonparticipant Ohio principals?

Project Background

Program features include: (1) statewide reach involving many districts; (2) annual enrollment of cohorts of at least 60 principals (with a few assistant principals); (3) two-year duration; (4) full-cohort PD sessions three times per year (“centralized training”); (5) regional PD sessions six times per year (“regional training”); (6) practical and reflective assignments in participants’ schools; (7) monthly coaching sessions at each participant’s school; and, notably, (8) superintendent engagement (e.g., via two mechanisms: three annual in-district “progress checks” and central office participation in the three annual centralized training sessions).

The two-year curriculum² helps principals (1) establish effective instructional teams (Building Leadership Teams—BLTs and Teacher-based Teams—TBTs) and (2) develop routines for observing teachers and eliciting their participation in discussions about effective instruction. Throughout the curriculum, moreover, presentations and activities emphasize inclusivity, which project leaders position in terms of three core values: (1) promoting equity and social justice, (2)

¹ Work on OLi⁴ began as a response to a federal request for application related to school leadership. The resulting proposal did not secure federal funding, but the project leadership collaborated with the Ohio Leadership Advisory Council (OLAC) and submitted a modified proposal for state funding. Funding has been provided by Ohio's Office for Exceptional Children (OEC) and Office of Innovation and Improvement (OII). Planning for statewide implementation began in 2013. Aimee participated in the initial planning of the project and became the third-party project evaluator. This study proceeds from that involvement.

² In July 2016, the project submitted a formal statement of the curriculum framework to the funders. The statement specifies the curriculum's grounding in research and standards; role definitions and learning outcomes; and curriculum model and sequence. The statement is available from the OLi⁴ first author: howleycb@gmail.com. Also see the program website: <https://www.oli-4.org/>

presuming the competence of all learners, and (3) treating access to a high-quality general education curriculum as every student's educational right. The key construct for OLi⁴ is inclusive *instructional leadership*, which the OLi⁴ core team has defined in terms of six leadership practices: (1) visioning, (2) using data well, (3) using research and evidence to guide instruction, (4) sharing leadership, (5) coaching teaching, and (6) reflecting on practice. Outcome statements provide brief descriptions of each of the practices (see Appendix A). For extensive information about OLi⁴, readers can consult the website (see footnote 2).

Literature Review

The role of principal in the United States emerged as a middle management position—but one that always paid a certain amount of attention to teaching and learning (Kafka, 2009). This attention has been known as *instructional leadership* for a long time (Ganon-Shilon & Schechter, 2017). Conceptualizations vary, but the core meaning refers to principals' efforts to improve teaching practice as “the key direction for the school” (Leithwood, Seashore-Louis, Anderson, & Wahlstrom, 2004, p. 6). This focus, indeed, characterizes OLi⁴.

Three bodies of literature bear on the pertinent issues: (1) research suggestive of the empirical limitations of attempting to link principals' instructional leadership to student learning; (2) the theoretical and practical literature relevant to principal leadership and social justice; and (3) empirical studies about efforts to foster leadership practice. The following review deals with each of these themes.

Limitations of Instructional Leadership

However it might be exerted, principals' instructional leadership is supposed to be functional: to help educators alter their teaching in the name of subsequently demonstrable learning (Leithwood & Seashore-Louis, 2011; Marzano et al., 2005). The most-sought demonstration is improved high-stakes test results, but improved equity (“closing the gap”) is another valued demonstration (Condon, 2011). To the extent that principals' practices are empirically associated with alterations in student learning, instructional leadership might be logically inferred (Leithwood & Seachore-Louis, 2011).

The logic model for such improvements generally stems from: (1) a set of valued leadership practices, (2) the influence of the leadership practices on teaching practices, and (3) demonstrably improved learning outcomes. Within the chain, each step embeds considerable nuance. As Branch, Hanushek, and Rivkin aver, “Understanding the impact of principals on learning is a particularly difficult analytical problem” (2012, p. 29). Reputable studies have nonetheless estimated effect sizes related to student outcome levels based on a wide variety of definitions and methods. The estimates range from nil (ten Bruggencate, Luyten, Scheerens, & Slegers, 2012) to moderate (Dhuey & Smith, 2014). Branch and colleagues (2012) reported an effect size of .10 for their fixed-effects model. Overall, principals' influence on student

achievement levels seems probable as the Leithwood and Marzano teams claim (largely on the basis of correlational studies).³

The connection of particular leadership practices to improved student achievement, however, remains unproven despite recent investigations. For instance, Miller, Goddard, Kim, Jacob, Goddard, and Schroeder (2016), in a randomized controlled trial (RCT), studied the effect of the widely used Balanced Leadership Program (BLD) developed by Waters, Marzano, and McNulty (2003). Though the PD was well executed and participants reported satisfaction and improved use of practices, the authors found that neither school climate nor student test scores showed differences between treatment and control schools.⁴ Camburn, Goldring, Sebastian, May, and Huff (2016) reported another RCT, with similar results. Exactly which practices can be changed, and which group of practices (six for OLi⁴ and 21 for BLD) cannot be specified with much certainty.⁵

Instructional Leadership and Social Justice

Both critical theory (Giroux, 2013; Theoharis, 2007) and historical scholarship (Kliebard, 1995) show that curriculum itself results from struggles over the purposes of schooling. Acknowledgment of this struggle has implications for the practice of instructional leadership.

For instance, Rigby (2014) empirically identified three species of instructional leadership: (1) prevailing, (2) entrepreneurial, and (3) social justice. “Prevailing” principals claimed both building management and instructional leadership roles, simply following conventional wisdom. “Entrepreneurial” principals are very similar to the prevailing type, but oriented to business models and accountability demands. Rigby’s (2014) “social justice” principals, by contrast, led faculty toward educational practices that fostered inclusion and equality, and towards achievement parity. Rigby (2014) regarded this alternative as a “full departure” (p. 630) from the other two models.

Indeed, according to Rigby (2014, p. 636), instructional leadership for social justice is itself “marginalized” within the profession. Principals who take an interest in marginalized students (students from impoverished families, students with disabilities, students of color) confront power arrangements that structure inequity throughout society as a whole (Hauser, Warren, Huang, Carter, Arrow, Bowles, & Durlauf, 2000; Isenberg, 2016; Kristal, 2013) and schooling in particular (Glass, 2007; Johnson, 2014; Tye, 2000; Theoharis & Causton-Theoharis, 2008). À

³ Both Dhuey and Smith (2014) and Branch and colleagues (2012) took a very conservative approach (“agnostic” for Branch and colleagues); they find that no particular practices define leadership quality. According to the studies, principal quality is inferred from value-added achievement measures. In other words, leadership inheres by definition to the role in these studies, and the independent variable (principal quality) rests on an interpolation of the dependent variable (achievement). The ten Bruggencate team (2012), by contrast, used structural equation modeling that parses principal influence to school culture and organization.

⁴ Miller and colleagues (2016, p. 536) describe the Marzano program as follows: [It] “is not so much a grand theory of educational administration as it is a compilation and synthesis of what works in education leadership research.” The program seeks to cultivate 21 leadership responsibilities on that basis: for instance, change agency, emphasizing clear goals, involving teachers in decision making.

⁵ This difficulty surfaces in the work of Leithwood and Seashore-Lewis (2011) who avoid “instructional leadership” in favor, simply of “leadership” (see the discussion in Leithwood et al., (2004), p. 6).

propos of this fact, Theoharis and Causton-Theoharis (2008, p. 243) have observed, “Intense resistance...comes with championing inclusive schools.”⁶

Efforts to Foster Instructional Leadership for Social Justice

University-based or agency-based leadership programs for social justice in schools also reportedly encounter substantial resistance (Hernandez & McKenzie, 2010). Thus, published accounts of professional development programs that foster instructional leadership with a focus on equity or social justice are rare (Capper & Young 2014; Frattura & Capper, 2007). We searched “professional development,” “principal leadership,” and “social justice or inclusion” and found just 17 documents in four education databases. All but one described the efforts of individual principals, and—in these accounts—the professional development utilized was that arranged by those principals for their teachers. The single work describing any program focusing on social justice for principals accounted for the longevity of a university principal preparation program (Merchant & Garza, 2015), and it examined the practices of the program, not the practices of its graduates.

Scholars have written extensively about the need for principal leadership on behalf of social justice (Albritton et al., 2017; Brooks, Adams, & Morita-Mullaney, 2010; Celoria, 2016; DeMatthews, Carey, Olivarez, & Saeedit, 2017; Kemp-Graham, 2015; Miller & Martin, 2014; Theoharis, 2007), and some scholars have reported the weak attention given to social justice issues of various kinds within principal preparation programs at universities (Hernandez & McKenzie, 2010; Marcellino, 2012; O’Malley & Capper, 2015).

Lessons from the Empirical Literature

Deriving lessons from the broad literature reviewed here helps to position the present study and the subsequent discussion of findings:

- Principals influence student learning through the way they shape the culture of a school, notably including how and why teachers teach (ten Bruggencate et al., 2012; Camburn et al., 2016; Leithwood & Seashore-Louis, 2011).
- Efforts to promote social justice (relevant to the key OLi⁴ construct of *inclusive instructional leadership*) are not a prevalent feature of current leadership practice (Angelle, Arlestig, & Norberg, 2016; Rigby, 2014). Overall it seems fair to assert that reported PD programs like OLi⁴ are rare.
- Evaluations of the effectiveness of PD for instructional leadership demonstrate that (1) changing the demonstrable practice of principals is difficult (Camburn et al, 2015;

⁶ Some readers may be confused by the use of “inclusive” in the context of general education. OLi⁴ applies the term *across* general and special education realms to indicate all marginalized groups: ethnic groups, English Language Learners, and students with disabilities alike.

Miller et al., 2016)—and (2) rigorous investigations of the phenomenon have only just begun.

If (1) the influence of principals is exerted through organizational culture (however measured) and (2) learning the relevant practices on site (and not only in PD events) requires assistance, then (3) programs must likewise be extensive (long-term) and intensive (job-embedded) and (4) evaluations must allow a relatively long trajectory for assessing summative effects.

This study asked relevant research questions of a program purporting to understand these conclusions of the literature on instructional leadership with a key social justice focus. The study is small-scale and, in that sense, exploratory, but there is clearly room for reports of such studies in the extant literature. Are participants' attitudes and practices at the conclusion of such a program (possibly as a result of participation) different from those of other principals?

Methods

This section begins with a description of participants. Then it describes data collection (including instrumentation). Next it provides a full description of study design, with details of the propensity-score matching techniques adopted by the study. The consideration of methods concludes by describing the data analysis methods used to compare matched treatment and control group performance on the dependent measures.

Participants

The treatment group consisted of the 56 members of OLi⁴'s Cohort 2 who completed their two-year training in June 2017. Treatment group members were matched one-to-one with 56 Ohio principals who had not participated in any OLi⁴ activities, chosen from among 201 who completed the study's two instruments. One-to-one propensity score matching (PSM) reduced selection bias for the comparison of treatment and control group means on the dependent variables.

The “propensity” in PSM is the likelihood that study participants (i.e., in both groups) will belong to the treatment group (based on similarities according to a set of independent variables). Technically speaking, one calculates a logistic regression that produces odds of treatment-group membership for each subject. In PSM, those odds are simply dubbed “propensity scores.” The actual matching in PSM takes place through matching algorithms (in this case, available in the R statistical libraries). In PSM researchers must specify both the logistic regression used to produce propensity scores and select the algorithm used to match treatment and control subjects. The choices made for this study are described subsequently, under “Study Design.”

PSM is a method for correcting differences in control-group and treatment-group membership (Holmes, 2014). A persistent problem in experimental research is the difficulty of assigning subjects randomly to treatment and control groups. The OLi⁴ program's main concern is delivery of the program, not the quasi-experiment reported here: the program had not been planned to include a randomized control trial (RCT)—the “gold standard” in experimental study. PSM is a

well-accepted alternative (Holmes, 2014) because it implements procedures to generate statistically comparable control groups (i.e., control subjects systematically matched to treatment subjects). In essence, PSM explicitly confronts the program-selection bias inherent when random assignment cannot be used.

Data Collection

All treatment-group principals ($n = 56$) completed the two survey instruments (see description below) at the conclusion of their two-year experience in June 2017. We recruited control-group members during the final months of the treatment group's participation in OLi⁴. First, the project coordinator asked Cohort 2 principals to nominate principals in nearby districts. The rationale for this approach was the likelihood that administrators in nearby districts would be more similar to project principals than those from other parts of the state. This effort yielded 190 nominations, all of whom were invited to complete the survey instruments; 44 did so. The response was insufficient for matching and we then decided to issue invitations to all Ohio principals of traditional (non-charter) public schools (except those in Cohort 2 and the 44 who had responded positively). At this second stage, 266 principals completed instruments, yielding 310 control-group candidates for matching with the 56 treatment-group members.⁷

Instrumentation

The study examined two project-relevant constructs: (1) principals' attitudes towards inclusion and (2) *inclusive instructional leadership* practices (see Appendix A). The former was measured with a 12-item 5-point Likert scale instrument and the latter with a 9-item 5-point Likert scale instrument ($\alpha = .78$ and $\alpha = .79$, respectively, in this administration).⁸ Neither the treatment group nor the control group had previously been administered the instruments. Tables 3 and 4 contain the full text of items for both instruments, together with scale anchors.

Through successive iterations in various contexts, the researchers had adapted the attitudes instrument and developed the practices instrument *de novo*.⁹ The attitudes instrument was adapted with authors' permission from the *Attitudes Towards Teaching All Students* instrument (Gregory & Noto, 2012). The practice instrument was developed *de novo* by the researchers in consultation with the <program> Core Team; initial versions were piloted with other <program> cohorts (i.e., Cohorts 1 and 3).

⁷ Following selection by PSM, 14 of the 44 nominated cases (37.8% of those with complete scores) and 42 of the other available control cases (20.2% of those with complete scores) comprised the control group. Propensity scores were higher for the nominated group than for the other available control principals. Mean propensity scores for the selected groups ranged from .35 (treatment) to .33 (nominated) to .31 (others).

⁸ Odd-numbered Likert scales embed the possibility that the middle option is "non-substantive," i.e., that it masks lack of knowledge or lack of opinion (Blasius & Thiessen, 2001). In our instruments, the middle option is "neither agree nor disagree" and the stem items are either statements about students (ATTITUDES) or characterizations of practice (PRACTICES). The substance of the statements on both instruments relates to conditions of work for responding principals; they know what they do (PRACTICES) and they have the experience to judge their views of students (ATTITUDES). The likelihood that the middle choice is a non-substantive response category seems low in this study.

⁹ Development details are available from the authors.

Study Design

The study used a post-only nonequivalent group design with propensity-score matching to equate treatment and control groups. That is, the test was administered one time and once matched, scores of the instruments in the treatment and control groups were compared and tested for statistically significant differences.

The 56 treatment group members completed all items on both instruments; of the 310 potential control group members, only 201 completed all items on both instruments. 109 cases were deleted listwise.¹⁰ These 201 principals were the cases used for propensity score matching with members of the treatment group.

The researchers opted for a parsimonious propensity score regression designed to theoretically model the selection of OLi⁴ members after the treatment group members (Holmes, 2014, p. 31). Researchers conferred with program staff to help specify these variables. We first learned that selection for the program had been based on the nomination of superintendents. On this basis (according to program staff and available data on participants), superintendents would likely (1) select younger principals (years as principal) at (2) schools challenged to improve (accountability rating); moreover, (3) district enrollment would likely be lower than statewide averages, and (4) school locale would be more often rural.

Propensity scores were produced with a logistic regression predicting treatment-group membership (coded 0 and 1) on four independent variables: (1) years as an educator, (2) Ohio school-level performance index, (3) district enrollment, and (4) rural (flagged as categorical, 0 and 1). The first three variables were statistically significant ($p < .05$). Rural entered the equation at $p = .053$. This four-variable model correctly classified 27% of treatment-group members and 97% of control-group members.¹¹ Predicted values were saved for use in propensity score matching via the R statistical analysis package.¹² Table 1 provides the equation used to create the propensity scores.

¹⁰ Listwise deletion of cases with missing data on dependent variables is conventional in PSM analyses (Kupzyk & Beal, 2017). The threat of listwise deletion is parameter estimate bias (Jacovidis, Foelber, & Horst, 2017). Bias is more probable if the group of deleted cases differs substantially from those retained, so we compared the deleted cases ($n = 109$) and the retained cases ($n = 201$). The cases retained for use in the R optimal match draw did not differ on 3 of the 4 PSM variables; only rural proved significantly different, and, as noted in the narrative, RURAL was marginally significant in the PSM logistic regression ($p = .053$). Moreover, in the set of $n=56$ control cases, the percentage of rural cases (30%) was similar to that of the set of $n=56$ matched treatment cases (27%). We believe that listwise deletion of cases with missing data was unlikely to have exerted systematic bias on parameter estimates.

¹¹ Variables chosen theoretically need not achieve statistical significance (Holmes, 2014).

¹² MatchIt and Optmatch modules (Ho, Imai, King, & Stuart, 2014).

Table 1
Logistic Regression for Creating Propensity Scores

	B	S.E.	Wald	<i>p</i>	Exp (<i>B</i>)
Years as educator	-.087	.027	10.022	.002	.917
School level performance index	-.101	.022	20.813	.000	.904
District enrollment	.000	.000	10.837	.001	1.000
Rural	.718	.371	3.752	.053	2.051
Constant	10.051	2.288	19.304	.000	23189.414

Note: *N* = 257 (*N* = 56 treatment cases and *N* = 201 control-eligible cases). *B* = log odds.

To create matched samples, the evaluators prepared a file with the propensity scores and a case identification number, imported the file (with 257 cases) into R, and drew two sets of one-to-one matched cases: (1) nearest match and (2) optimal match. Summaries of the two draws were saved and compared. The optimal-match draw rendered treatment and control groups that were more similar by reducing mean distance between scores overall. The percentage improvement in balance was greater with the optimal-match as compared to the nearest-match draw.

Post-hoc analysis further demonstrated that the optimal-match draws had reduced the significant differences on all predictor variables (years, performance index, enrollment, and rural) to non-significant. Table 2 reports the differences on the predictor variables before and after optimal matching. The optimal-match draw became the basis of comparing the dependent variables across treatment and control groups for the reasons given.

Table 2
Effect of Optimal Matching on Group Differences

	Groups	Before		After	
		Mean	<i>p</i>	Mean	<i>p</i>
Years as educator	Control	22.69	.002	19.68	.490
	Treatment	18.93		18.93	
District enrollment	Control	7235.90	.090	4903.02	.577
	Treatment	6058.05		6058.05	
Rural	Control	.42	.000	.29	.835
	Treatment	.27		.27	
School level performance index	Control	97.61	.021	93.55	.327
	Treatment	91.67		91.67	

^a. *N* = 201 control, 56 treatment; unequal variances for years as educator, rural, and school level performance index.

^b. *N* = 56 in each group; variances for all variables equalized.

Data Analysis

Dependent variables for the matched samples produced by the opti-match algorithm were compared using independent-sample *t*-tests with effect sizes (Cohen's *d*) calculated for significant differences. Significance level was set at $p < .05$. Data analysis also included calculation of descriptive statistics (means and standard deviations) for individual items.

Ancillary analysis (unplanned and subordinate) included an item-by-item comparison for the *practices* scale, which showed non-significant results for the total score. Given the concern raised by Leithwood and colleagues (2004, p. 6) about "leadership by adjective," we were curious if any of the OLi⁴-specified practices might exhibit a statistically significant difference in the treatment as compared to the control group.

Findings

Descriptive statistics for all items on the two instruments appear in Tables 3 (*attitudes*) and 4 (*practices*). The results of independent samples *t*-tests comparing the means of the dependent measures for the treatment and control groups appear in Table 5.

Attitudes exhibits a statistically significant difference favoring the treatment group. The difference is equivalent to a moderate effect size of $d = .47$. Although mean differences for the *practices* scale (sum of nine items) did not reach statistical significance, the observed values (28.48 and 29.52) favored the treatment group by about one-quarter of a standard deviation, as did all items but one.

Given this pattern, and as an ancillary analysis, the researchers examined the nine individual practices items for statistical significance. One item (reverse coded) proved statistically significant (P1: "How difficult is it for you to engage teachers in collaborative problem solving and collaborative professional learning [e.g., in TBT and BLT meetings]?"). Table 5 also includes the results of this ancillary analysis.

The *attitudes* and *practices* instruments comprised 21 total items. Researchers found it noteworthy that none of these items exhibited group means that favored the control group (see Tables 3 and 4). For 20 of the item pairs (control group as compared to treatment group), the observed item means favored the treatment group. The remaining item, "How clearly can teachers see that your actions demonstrate your own habits of self-reflection?" exhibited the same mean for both groups.

Table 3
Descriptive Statistics for Attitudes Individual Items

Items	Condition			
	Treatment		Control	
	Mean	SD	Mean	SD
Special classrooms and schools are not the best place to provide services to students who have complex needs.	2.96	1.29	2.75	1.00
Students who do not have special needs get very little benefit from interactions involving students with deafness, blindness, or other such conditions. [R]	4.48	.76	4.41	.73
General education classrooms are usually the best placement for students with mild to moderate disabilities.	4.11	.89	4.04	.99
The placement of students with low-incidence disabilities in general education classrooms provides them with little academic benefit. [R]	4.15	1.01	3.91	.88
The placement of students with low-incidence disabilities in general education classrooms benefits their social development.	4.43	.66	4.21	.78
Standards-based instruction is academically productive for students with cognitive disabilities.	3.71	.96	3.30	1.13
A child's low-incidence disability label does not tell us where he or she should be placed.	3.95	1.08	3.80	.92
Students with mild to moderate disabilities should not be trusted with responsibilities in the classroom. [R]	4.22	1.23	4.13	1.21
Instruction in functional skills (e.g., activities of daily living) should precede academic instruction for students with low-incidence disabilities. [R]	3.43	.82	3.04	.87
Many students with low-incidence disabilities (e.g., blindness, deafness) can handle the academic focus of the general education program.	4.14	.82	3.93	.85
Helpful instructional modifications (larger print, study guides, advance organizers) for students with low-incidence disabilities are not something general education teachers can make. [R]	4.68	.69	4.02	1.10
Students with mild to moderate disabilities can be educated more effectively in special education classrooms as opposed to general education classrooms. [R]	4.05	.94	3.86	.94

Note. R= reverse scored. Anchors: strongly disagree, disagree, neither disagree nor agree, agree, strongly agree

Table 4
Descriptive Statistics for Practices Individual Items

	Condition			
	Treatment		Control	
	Mean	SD	Mean	SD
How difficult is it for you to engage teachers in collaborative problem solving and collaborative professional learning (e.g., in TBT and BLT meetings)? [very difficult, difficult, neither difficult nor easy, easy, very easy]	3.43	.83	3.11	.85
How challenging do you find the use of data for making school-wide improvement plans? [very challenging ... very easy]	3.09	1.10	2.96	.99
How often do teachers in your school see you using data? [very often, often, sometimes, occasionally, rarely] [R]	3.84	.90	3.70	.66
To what extent are you engaged with helping teachers use data well? [very much engaged, engaged, neither engaged nor disengaged, disengaged, very much disengaged] [R]	3.91	.64	3.86	.70
How challenging is it to guide teachers in their use of evidence in planning for diverse students? [very challenging... very easy]	2.38	.93	2.29	.80
How challenging is it for you to help teachers think about their instructional practices? [very challenging... very easy]	2.88	.94	2.84	.93
How clearly can teachers see that your actions demonstrate your own habits of self-reflection? [very clearly, clearly, neither clearly nor unclearly, somewhat unclearly, not at all] [R]	3.84	.63	3.84	.76
To what extent do other activities keep you from spending time learning about effective instructional practices? [completely, quite a bit, somewhat, not so much, not at all]	2.36	.70	2.21	.78
To what extent do core values of equity, social justice, and inclusiveness serve as a basis for planning activities at your school? [completely...not at all]	3.82	.77	3.68	.90

Note. R= reverse scored.

Table 5
Group Mean Differences for Attitudes and Practices

	Group	Mean	SD	<i>p</i>	Effect Size
Attitudes	Control	45.39	6.07	.016	.47
	Treatment	48.36	6.54		
Practices (all items)	Control	28.48	4.41	.232(ns)	NA
	Treatment	29.52	4.65		
Practice item 1	Control	3.11	.85	.045	.38
	Treatment	3.43	.83		

Note. N = 56 for each group; practice items 2-9 all proved statistically non-significant (.232 < *p* < 1.0).

Limitations

The use of quasi-experimental methods in place of randomized assignment limits the capacity of this study to assert causality. The use of self-reported data to assess practices rather than more direct measures (e.g., observation or teachers’ reports of principals’ practices, or possible defensible proxy measures such as collective efficacy surveys of teachers) is also a limitation. Miller and colleagues (2016) did find self-reported practice change, but did not find substantive changes in schools, for instance. To address this limitation in this study, the researchers are in fact now conducting a pre- and post- study of the collective efficacy of teachers and the schools of OLi⁴ principals in Cohort 5.

Attrition was not an issue because of the study design (post-only), but attrition naturally did occur among the treatment group. Attrition occurred principally because changes in district superintendents resulted in the decision to withdraw districts from participation; some attrition resulted from individual-level job changes. A few attritions stemmed from challenges to participants’ personal circumstances. OLi⁴ Cohort 2 started with 87 members and concluded with 57, of whom data was collected from 56.

Finally, effects and experiences probably varied among those participants who did complete the program. For instance, effects might be stronger among those who attended more of the training events or among those who engaged the blogging activity more extensively. The project collects these data and they would be available in future research efforts about the effectiveness of OLi⁴.

Discussion

We designed this study to address the research questions prudently. We conclude, on the basis of the completed study, that differences do exist in attitudes between participants and non-participants. A quasi-experiment is one alternative to the randomized controlled trials conventionally understood to establish causality (Holmes, 2014). This small-scale study is one example.¹³

¹³ Did OLi⁴ cause a change in attitudes? We are skeptical of any such claim—whether warranted by an RCT or

Causality?

The pattern of observed responses indicates the need for further study and the authors are engaged in such a study (mentioned in the previous section). On the basis of the results, it seems reasonable to regard this previously untested program as a promising practice documented by initial Tier II evidence (USDE, 2016) as well as being suitable for further development and further study. The findings at least demonstrate that principals in the second cohort trained by OLi⁴ differ in their self-assessment of attitudes toward inclusion. Self-reported practices of *inclusive instructional leadership* did not exhibit an overall similar difference.

Difficulties

The non-significance of *practices* is a disappointment for program leaders. OLi⁴ is a PD program with the exact aim of changing these studied practices. The item that did exhibit a significant difference, principals' work with teachers on collaborative problem solving and professional learning, though, is a focus of OLi⁴'s effort because collaborative teams are part of Ohio's school improvement provisions. Across the nation, educators in all roles are poorly prepared for this sort of work (Mandinach & Gummer, 2016).

We cannot know, from this study, if collaborative teams in the schools of OLi⁴'s principals *are* functioning better, but that question could be addressed in the future. Leithwood and Seashore-Louis (2011)—in a thoughtful and often cited literature review—conclude that principals are the largest influence on student learning after a teacher's own influence. Apparently, schools need direction and they founder without it, especially without thoughtful, active, and collaborative direction (Leithwood & Seashore-Louis, 2011).

OLi⁴ struggles to cultivate just such direction. OLi⁴ provides accessible tools and routines and cultivates principals in the use of those routines and tools, in order to create better functioning schools. The extant literature, however, is not auspicious for the causal efficacy of PD programs for principals' instructional leadership (Miller et al., 2016). The cultivation of such capacity makes sense to many actors, even if it lacks an affirmation of causality.

Self-reported data are cost-effective for use in a study such as this, but they cannot substitute for measurement independent of program participants. We think the present results are promising, and the OLi⁴ Core Team has authorized a longitudinal study of changes in the *collective teacher efficacy* of the schools of OLi⁴ Cohort 5. This choice is a nod to the prominence of organizational culture in the school leadership literature, and to the possibility that an exact toolkit of leadership skills (or specific practices) might be a distraction. Whatever the practical and research difficulties, this is surely interesting work for program leaders and for investigators.

a quasi-experiment using PSM (see Koopmans, 2014). Despite such misgivings, the construct of causality is understandable in the context of high-stakes summative evaluation of products and programs (such as OLi⁴) which are part of the terms of engagement (see United States Department of Education (USDE), 2016). On such terms, this study would be a candidate for providing "Tier II" evidence of program effectiveness (USDE, 2016, p. 7).

Recommendations for PD Organizers

Findings from this study do not in themselves suggest recommendations for how to conduct PD for principals. We think the findings do hint at the program's success and do suggest the value of steady work on the difficult issues that occupy OLi⁴. So what might the authors, based on their experience of the program, recommend to state or regional leaders interested in or committed to providing similar PD? Our perspectives proceed from many years of involvement with OLi⁴ and, indeed, from full careers of involvement with school leadership programs and related scholarship:

Recommendation 1. Secure funding and support from state-level actors. Still, this recommendation is far, far easier to give than to carry out. Its reasonableness depends on the “political” context: state, federal, or local.¹⁴ In Ohio, the originators were exceptionally fortunate.

Recommendation 2. Design the curriculum: avoid simply “delivering PD.” Curriculum design relies on the construction of a conceptual framework—it is the first and perhaps most critical step in the design work. Of course, fully specifying the curriculum, developing procedures, activities, session formats, and plans takes years. It can be done as the program unfolds, but only with concerted attention from program leadership and with the involvement of third-party evaluators.

Recommendation 3. Conduct rigorous evaluation, preferably client- and participant-oriented evaluation. This provision is essential to the program's own level of thoughtfulness and intentional action. For instance, ongoing curriculum development is less likely without such evaluation.

Recommendation 4. PD of this sort (i.e., confronting challenging issues in school leadership) requires extensive (several years) and intensive (job-embedded) provisions. Such provisions, for the moment, represent a settled view of what good PD requires. The job-embedding seems key to us; OLi⁴ participants consistently report their appreciation of both the provocations and support provided by OLi⁴ coaches.

OLi⁴ is unusual among PD programs for principals because it has the support of the State Education Agency (SEA) and treats special and general education as a single system. In this context, OLi⁴ uses the leadership teams recommended by the Ohio Improvement Process (OIP). It also notably articulates a social justice focus (inclusive instructional leadership) as its key construct (and, in fact, its conceptual framework). OLi⁴ clearly has more work to do along with

¹⁴ “Political” in the sense of who is allocating what resources for which valued purposes. Such coalitions are notoriously unstable. Readers might note that OLi⁴ is not sectarian or dogmatic in its articulation of social justice issues, and it has incorporated and adapted the Ohio Improvement Process within its curriculum. At this writing, the SEA is updating the Ohio Improvement Process, persuaded by OLi⁴'s elaboration of the process in its curriculum.

colleagues everywhere who are involved with school leadership, the on-going professional development of school leaders, and social justice in the United States.

Author Notes

Craig Howley is a senior researcher as WordFarmers Associates, LLC.

Aimee Howley is president and principal investigator at WordFarmers.

Jacqueline Yahn is an assistant professor of education at Ohio University.

Pamela VanHorn coordinates the OLi⁴ project at the University of Cincinnati (UC).

Deborah Telfer directs the Systems Development and Improvement Center at UC.

Correspondence regarding this article should be directed to Craig Howley at howleycb@gmail.com.

References

- Albritton, S., Huffman, S., & McClellan, R. (2017). A study of rural high school principals' perceptions as social justice leaders. *Administrative Issues Journal: Connecting Education, Practice, and Research*, 7(1), 19-38.
- Angelle, P. S., Arlestig, H., & Norberg, K. (2015). The practice of socially just leadership: Contextual differences between US and Swedish principals. *International Studies in Educational Administration*, 43(2), 21-37.
- Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (CALDER Working Paper No. 66). Retrieved from <http://files.eric.ed.gov/fulltext/ED529175.pdf>
- Brooks, K., Adams, S. R., & Morita-Mullaney, T. (2010). Creating inclusive learning communities for ELL students: Transforming school principals' perspectives. *Theory into Practice*, 49(2), 145-151.
- Camburn, E. M., Goldring, E., Sebastian, J., May, H., & Huff, J. (2016). An examination of the benefits, limitations, and challenges of conducting randomized experiments with principals. *Educational Administration Quarterly*, 52(2), 187-230.
- Capper, C. A., & Young, M. D. (2014). Ironies and limitations of educational leadership for social justice: A call to social justice educators. *Theory into Practice*, 53(2), 158-164.
- Celoria, D. (2016). The preparation of inclusive social justice education leaders. *Educational Leadership and Administration: Teaching and Program Development*, 27, 199-219.
- Condron, D. (2011). Egalitarianism and educational excellence: Compatible goals for affluent societies? *Educational Researcher*, 40(2), 47-55.
- DeMatthews, D. E., Carey, R. L., Olivarez, A., & Saedit, K. M. (2017). Guilty as charged? Principals' perspectives on disciplinary practices and the racial discipline gap. *Educational Administration Quarterly*, 53(4), 519-555.
- Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Education*, 47(2), 634-663.
- Frattura, E. M., & Capper, C. A. (2007). *Leading for social justice: Transforming schools for all learners*. Thousand Oaks, CA: Sage Publications.
- Ganon-Shilon, S., & Schechter, C. (2017). Making sense of school leaders' sense-making. *Educational Management Administration and Leadership*, 45(4), 682-698.

- Giroux, H. A. (2013). *America's education deficit and the war on youth*. New York, NY: Monthly Review Press.
- Glass, G. (2007). *Fertilizers, pills, and magnetic strips: The fate of public education in America*. Charlotte, NC: Information Age Publishing.
- Gregory, J. L., & Noto, L. A. (2012, November). *Technical manual for Attitudes towards Teaching All Students (ATTAS-mm) instrument*. Paper presented at the Annual Conference of the Council for Exceptional Children, Teacher Education Division, Grand Rapids, MI.
- Hauser, R., Warren, J. R., Huang, M.-H., & Carter, W. (2000). Occupational status, education, and social mobility in the meritocracy. In J. Arrow, S. Bowles, & S. Durlauf (Eds.), *Meritocracy and economic inequality* (pp. 179–229). Princeton, NJ: Princeton University Press.
- Hernandez, F., & McKenzie, K. B. (2010). Resisting social justice in leadership preparation programs: Mechanisms that subvert. *Journal of Research on Leadership Education*, 5(3), 48-72.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 8(42), 1–28. Retrieved from <https://gking.harvard.edu/atons/matchit-nonparametric-preprocessing-parametric-causal-inference>
- Holmes, W. M. (2014). *Using propensity scores in quasi-experimental designs*. Thousand Oaks, CA: Sage Publications.
- Isenberg, N. (2016). *White trash: The 400-year untold history of class in America*. New York, NY: Viking.
- Jacovidis, J. N., Foelber, K. J., & Horst, S. J. (2017). The effect of propensity score matching method on the quantity and quality of matches. *The Journal of Experimental Education*, 85(4), 535-558.
- Johnson, H. B. (2014). *The American dream and the power of wealth: Choosing schools and inheriting inequality in the land of opportunity* (2nd ed.). New York, NY: Routledge.
- Kafka, J. (2009). The principalship in historical perspective. *Peabody Journal of Education*, 84(3), 318-330.
- Kemp-Graham, K. Y. (2015). Missed opportunities: Preparing aspiring school leaders for bold social justice school leadership needed for 21st century schools. *International Journal of Educational Leadership Preparation*, 10(21), 99-129.

- Kliebard, H. (1995). *The struggle for the American curriculum, 1893-1958*. New York, NY: Routledge.
- Koopmans, M. (2014). Change, self-organization and the search for causality in educational research and practice. *Complicity: An International Journal of Complexity & Education, 11*(1), 20–39.
- Kristal, T. (2013). The capitalist machine: Computerization, workers' power, and the decline in labor's share within U.S. industries. *American Sociological Review, 78*(3), 361-389.
- Kupzyk, K. A., & Beal, S. J. (2017). Advanced issues in propensity scores: Longitudinal and missing data. *Journal of Early Adolescence, 37*(1), 59-84.
- Leithwood, K., & Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research, 52*(3), 309-339.
- Leithwood, K., & Seashore-Louis, K. (2011). *Linking leadership to student learning*. San Francisco, CA: Jossey-Bass.
- Leithwood, K., Seashore-Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. Retrieved from <https://www.wallacefoundation.org/knowledge-center/Documents/How-Leadership-Influences-Student-Learning.pdf>
- Mandinach, E. B., & Gummer, E. S. (2016). *Data literacy for educators: Making it count in teacher preparation and practice*. San Francisco, CA: WestEd.
- Marcellino, P. A. (2012). Preparing educational leaders for social justice, action-learning, and democratic activism. *International Journal of Educational Leadership Preparation, 7*(3), 1-11.
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Merchant, B., & Garza, E. (2015). The Urban School Leaders Collaborative: Twelve years of promoting leadership for social justice. *Journal of Research on Leadership Education, 10*(1), 39-62.
- Miller, J. R., Goddard, R. D., Kim, M., Jacob, R., Goddard, Y., & Schroeder, P. (2016). Can professional development improve school leadership? Results from a randomized control trial assessing the impact of McREL's Balanced Leadership Program on principals in rural Michigan schools. *Educational Administration Quarterly, 52*(4), 531-566.

- Miller, C .M., & Martin, B. N. (2014). Principal preparedness for leading in demographically changing schools: Where is the social justice training? *Educational Management, Administration, and Leadership*, 43(1), 129-151.
- Ohio Leadership for Inclusion, Implementation, and Instructional Improvement. (2017). *Purposes and priorities: Overview*. Retrieved from the Ohio Leadership for Inclusion, Implementation, & Instructional Improvement (OLi4) website: <https://www.oli-4.org/about-oli4/purpose-priorities>
- O'Malley, M. P., & Capper, C. (2015). A measure of the quality of educational leadership programs for social justice: Integrating LGBTIQ identities into principal preparation. *Educational Administration Quarterly*, 51(2), 290-330.
- Riehl, C. (2000). The principal's role in creating inclusive schools for diverse students: A review of normative, empirical, and critical literature on the practice of educational administration. *Review of Educational Research*, 70(1), 55-81.
- Rigby, J. G. (2014). Three logics of instructional leadership. *Educational Administration Quarterly*, 50(4), 610-644.
- ten Bruggencate, G., Luyten, H., Scheerens, J., & Slegers, P. (2012). Modeling the influence of school leaders on achievement: How can school leaders make a difference? *Educational Administration Quarterly*, 48(4), 699-732.
- Theoharis, G. (2007). Social justice, educational leaders, and resistance: Toward a theory of social justice leadership. *Educational Administration Quarterly*, 43(2), 221-258.
- Theoharis, G., & Causton-Theoharis, J. (2008). Oppressors or emancipators: Critical dispositions for preparing inclusive school leaders. *Equity and Excellence in Education*, 41(2), 230-246.
- Tye, B. B. (2000). *Hard truths: Uncovering the deep structure of schooling*. New York, NY: Teachers College Press.
- United States Department of Education. (2016). *Using evidence to strengthen education investments*. Retrieved from <https://ed.gov/policy/elsec/leg/essa/guidanceuseseseinvestment.pdf>
- Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement: A working paper*. Retrieved from: <http://files.eric.ed.gov/fulltext/ED481972.pdf>

Appendix A

OLi⁴ Outcome Statements

- **Visioning:** OLi⁴ principals will use district and school non-negotiables to set specific achievement targets for all classrooms and students, and ensure the consistent use of evidence-based instructional strategies in all classrooms to reach those targets.
- **Using Data Well:** (1) OLi⁴ principals will use data to make effective decisions and (2) OLi⁴ principals will help teachers use data to make more effective decisions.
- **Using Research and Evidence to Guide Instruction:** OLi⁴ principals will guide teachers in their selection of evidence-based instructional practices for diverse learners.
- **Sharing Leadership:** (1) OLi⁴ principals will share leadership with teachers based on their expertise and (2) OLi⁴ principals will engage teachers, through the TBT and BLT structures, in collaborative problem-solving and other collaborative learning.
- **Coaching Teaching:** (1) OLi⁴ principals will monitor teaching for effectiveness and (2) OLi⁴ principals will call into question teaching practices that appear to be ineffective.
- **Reflecting on Practice:** (1) OLi⁴ principals will reflect on their own practice and (2) OLi⁴ principals will model and encourage self-reflection, active learning, and application of that learning.