Hammill Institute on Disabilities

Remedial and Special Education
2022, Vol. 43(5) 328–343

DOI: 10.1177/07419325211054208
rase.sagepub.com
SAGE

Data Literacy Training for K–12 Teachers: A Meta-Analysis of the Effects on Teacher Outcomes

Marissa J. Filderman, PhD[1], Jessica R. Toste, PhD[2], Lisa Didion, PhD[3], and Peng Peng, PhD[2]

Abstract

This meta-analysis explores training teachers in the use of data, defined as any quantifiable information that helps teachers know more about their students for instructional decision-making. The questions addressed are as follows: (a) What are the features of data literacy training for kindergarten through 12th-grade teachers? (b) What are the effects of data literacy training on kindergarten through 12th-grade teacher outcomes? (c) Do training characteristics moderate the effects of training? A comprehensive search of research conducted between 1975 and 2019 yielded 33 studies with 163 effect sizes that met inclusion criteria. Using a random effects model, the researchers found significant positive effects on knowledge and skills, $g = .67$, 95% confidence interval (CI) = [0.40, 0.93], and beliefs, $g = .48$, 95% CI = [0.17, 0.79]. A collaborative training format significantly and positively moderated effects. Implications for teacher trainings and the design of future research are discussed.

Keywords

teacher, professional development, training, data literacy, meta-analysis

The use of data to inform instruction has become increasingly important to support students in meeting academic expectations, particularly as nationwide trends indicate that a vast majority of students are not proficient in reading or math (National Assessment of Educational Progress, 2019). Educational initiatives emphasize the use of student data as a central tenet to improving the U.S. education system and tackling these subpar outcomes (e.g., Every Student Succeeds Act, 2015; Race to the Top Act, 2011). In addition, campaigns (e.g., Data Quality Campaign, 2016), research organizations (e.g., National Center on Intensive Intervention, 2013), and professional teacher organizations (Council of Chief State School Officers, 2013; National Council for the Accreditation of Teacher Education, 2010) promote increased data use. A recent court case (*Endrew F. v. Douglas County School District*, 2017) emphasizes the importance of data use for students with disabilities in particular by legislating appropriate progress unique to each child. However, research on general and special education teachers consistently confirms that teachers have difficulty using data to guide instructional decisions (e.g., Datnow & Hubbard, 2016; Mandinach & Gummer, 2016).

To support teachers in the effective use of data, training is necessary (Mandinach & Gummer, 2016). To date, little is known about the effects of such training. This review sought to describe the features of data literacy training and to determine the effects of data literacy training on teacher outcomes (i.e., knowledge, skills, and beliefs). It was also of interest to examine the extent to which best practices of training were used, how this unique topic was addressed, and how each of these impacted teacher learning. Thus, this review provides critical information about current trends and effects of training in data literacy, with important implications for the design of future trainings and research.

Conceptual Framework

The conceptual framework guiding this review is derived from a combination of widely referenced frameworks and definitions for data (Joint Committee on Standards for Educational Evaluation [JCSEE]; Klinger et al., 2015), data-driven decision-making (Mandinach et al., 2008), and teacher training (Desimone, 2009). In the context of this

[1]The University of Alabama, Tuscaloosa, USA
[2]The University of Texas-Austin, USA
[3]The University of Iowa, Iowa City, USA

Corresponding Author:
Marissa J. Filderman, The University of Alabama, College of Education, Box 870232, Tuscaloosa, AL 35487, USA.
Email: mjfilderman@ua.edu

Associate Editor: Emily Solari

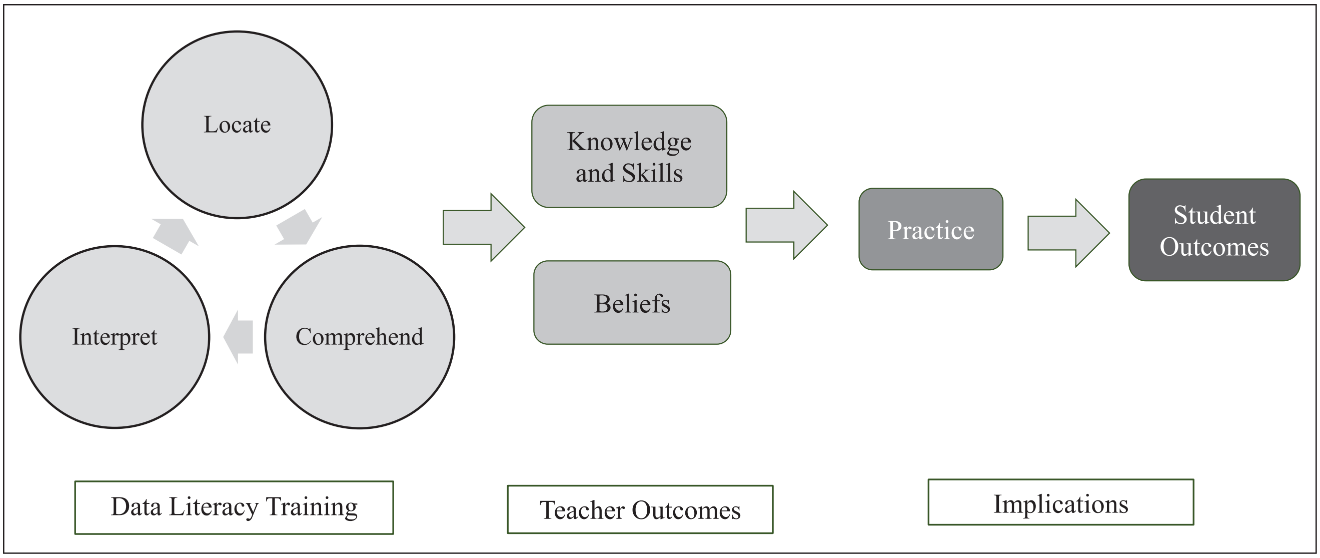

Figure 1. Data literacy training model.
Note. This model was adapted from the general training model in Desimone (2009). The initial framework for data literacy training was adapted from the framework for data-driven decision-making seen in Mandinach et al. (2008).

review, the term *data* refers to any quantifiable information that helps educators know more about their students for instructional decision-making. To meet the latter part of this definition, data include assessments *for* learning, rather than assessments *of* learning—that is, assessments from which data may be used to improve teaching and learning (e.g., curriculum-based measurement, mastery measures) rather than assessments that evaluate the extent to which content was learned (e.g., high-stakes assessments; Klinger et al., 2015; Mandinach, 2012). The data-driven decision-making framework suggests that data literacy consists of three key components: (a) location (e.g., design, select, administer, and collect data), (b) comprehension (e.g., read and understand data), and (c) interpretation (e.g., understand the implications of data to adjust instruction; Gummer & Mandinach, 2015; Mandinach et al., 2008; Mandinach & Gummer, 2016). To promote data literacy, teachers require training. A widely accepted framework (Desimone, 2009) suggests that training teachers will increase their (a) *knowledge and skills* (defined in this study as having an understanding of and ability to apply data literacy components), and (b) *beliefs* (defined as self-efficacy and perceived value of data use). In turn, *teacher practices* (defined as sustained data use in the classroom) will be enhanced, resulting in improved student outcomes. There is evidence to support the latter part of this framework, that is, when teachers use data, academic outcomes improve (Filderman et al., 2018; Jung et al., 2018). As such, this review focuses on immediate outcomes of training (i.e., knowledge, skills, and beliefs; see Figure 1).

Data Literacy Teacher Training

Data Literacy Knowledge and Skills. Many studies have pointed to gaps in teacher knowledge and skills related to the data literacy components (i.e., location, comprehension, and interpretation). For instance, teachers struggle to *locate* data (e.g., to develop diagnostic assessments: Gallagher et al., 2008; to use systems to access available data: Means et al., 2009). Teachers also struggle to *comprehend* data (e.g., to understand test scores: Means et al., 2011; to describe progress monitoring graphs: Wagner et al., 2017; or to extract relevant information from graphs and understand slope: Espin et al., 2017). Finally, teachers struggle to *interpret* data (e.g., to link data to instructional changes: van den Bosch et al., 2017; to use data systematically to make decisions: Espin et al., 2017; Wagner et al., 2017). Given the breadth of knowledge and skills required for each of the data literacy components, it is unsurprising that research suggests that a majority of general and special education teachers struggle with data use (Datnow & Hubbard, 2016; Means et al., 2011).

Beliefs Surrounding Data Use. Teacher beliefs in the value and utility of data influence their data use. For instance, teachers who held the belief that assessments were valuable for guiding instruction (Remesal, 2011) and who believed collecting and analyzing data to identify and meet student needs was foundational to their role as a teacher (Jimerson & Wayman, 2012) were found to use data significantly more than teachers who did not hold these beliefs. Other foundational beliefs for data use include belief in the collective responsibility of data collection, the importance of using data to monitor progress, and teachers' own autonomy to make decisions based on data (Hoogland et al., 2016). Teachers' belief in their abilities to use data, known as *self-efficacy* (Bandura, 1997), is also important. Evidence suggests teachers have low self-efficacy related to data use; moreover, teachers' data use practices are shaped by these self-efficacy beliefs, with lower self-efficacy associated with less frequent use of data (Wayman & Jimerson, 2014).

High-quality training features. To support teacher capacity in each of these areas, training must be delivered. There are several features of training that are demonstrated to increase the effects on teacher outcomes, including (a) content focus (i.e., information presented on subject matter content and how students learn such content), (b) active learning (i.e., providing opportunities for interaction with and collaboration about content presented), (c) collective participation (i.e., opportunities for interaction with teachers in close proximity to practice), (d) duration (i.e., providing a sufficient amount of training), and (e) coherence (i.e., aligning content with teacher beliefs and/or school and district policy; Desimone, 2009). In the following paragraphs, we describe how these features may be interpreted within the context of this study.

Content focus. First, content focus may be of particular interest for data literacy training. Desimone's (2009) definition of content focus includes information about the subject matter and how it works to improve student learning. When considering the three components of data literacy (i.e., location, comprehension, and interpretation), there are two features that could be considered to provide information about the subject matter (information about student learning: location and comprehension) and one that could be considered to provide teachers with ways to improve student learning (how to use data to inform instruction: interpretation). In her 2016 review of the effects of professional development on student outcomes, Kennedy found that trainings that included ways to improve student learning were associated with larger effects than trainings that focused on knowledge-building alone. It was therefore of interest to determine whether data interpretation may moderate effects and whether more than one component of data literacy may impact outcomes.

Active learning and collective participation. Active learning and collective participation both involve interaction, the former with content and individuals and the latter with individuals within a local community of practice. Because both features involve collaboration, and because of the challenge in identifying whether teachers participated in training with school-based teams, we encompass both of these categories within a larger category of *collaboration*. There is evidence that collaborative training helps teachers to effectively learn content (Wayman et al., 2017). Prior qualitative studies have found this collaboration to be particularly important for teacher learning related to data literacy (Jimerson & Wayman, 2012). Additional research has found collaboration to be central to teachers' ongoing use of data (Means et al., 2011; Piro et al., 2014). As such, we explored whether trainings that included collaboration were associated with increased effects of training.

Duration. There is no consensus yet on a sufficient duration for training, with researchers reporting 14 hr (Yoon et al., 2007) 20 hr (Desimone, 2009), 30 hr (Guskey & Yoon, 2009), and 49 hr (Darling-Hammond et al., 2009) to be necessary for training to improve learning. Other researchers have not found a direct link between duration and outcomes (Kennedy, 2016; Kraft et al., 2018). Given the complexities of data literacy, it could be posited that a more intensive duration may be particularly important for training. However, recent research has found that as little as one session of in-service training may be related to increased data use by teachers (Filderman et al., 2020). Because of the variety of findings and the potential differences with data literacy training, we sought to include duration as a moderator.

Coherence. Like other reviews of teacher training (e.g., Didion et al., 2020; Kennedy et al., 2016; Yoon et al., 2007), coherence was not considered for several reasons. First, as the use of data is emphasized in U.S. legislation, it can be argued that all data literacy content is aligned with what teachers need to do. Despite this emphasis, we note that there is substantial variation in the amount to which districts and schools emphasize data use, which has been demonstrated to be an important foundation for sustained data use in practice (Hoogland et al., 2016). However, a second reason why we did not include coherence is that it is often not measured in training literature. Specifically, the training literature often does not include contextual information, such as school schedules, curricula, or administrator interviews, that would allow us to evaluate the extent to which school and district policies align with federal legislation. Third, teacher beliefs were often not measured in conjunction with their knowledge and skills; thus, we are also not able to draw conclusions about whether teacher beliefs were aligned with the content covered and whether this in turn had an impact on teacher knowledge and skills related to data use.

Coaching. In addition to these key features, we consider an additional feature not identified by Desimone (2009) but that may be particularly important for data use, namely, coaching (Datnow & Hubbard, 2016). Reviews of the literature have found coaching ($d = .49$; Kraft et al., 2018) and performance feedback (Fallon et al., 2015; Solomon et al., 2012) to significantly improve instruction. However, a recent review of teacher training found that coaching did not significantly moderate the effects of training (Brock & Carter, 2017). Due to the challenging content, coaching may provide teachers with important affirmative and corrective feedback, as well as an opportunity to clarify the content learned. Thus, we sought to explore whether coaching may impact the effects of training.

Purpose of the Present Review

Previous reviews have investigated prerequisite skills required for data use (Datnow & Hubbard, 2016; Hoogland et al., 2016; Xu & Brown, 2016), the ways in which teachers use data to make decisions (Datnow & Hubbard, 2015; Little, 2012), and state and district initiatives to support data use (Marsh, 2012). To add to the existing literature, we sought to examine the effects of data literacy training on teacher outcomes. As such, we posed the following three research questions:

Research Question 1 (RQ1): What are the features of data literacy training for kindergarten through 12th-grade teachers?
Research Question 2 (RQ2): What are the effects of data literacy training on kindergarten through 12th-grade teacher outcomes (i.e., data literacy knowledge and skills, beliefs)?
Research Question 3 (RQ3): Do training characteristics (i.e., coaching, delivery format, inclusion of multiple data literacy skills, and inclusion of data interpretation skills) moderate the effects of data literacy training on teacher outcomes?

Method

Search Procedures and Study Identification

An overview of search procedures is presented in Figure 2. We conducted a systematic search for studies published in English between the years 1975 and 2019. We chose 1975 as a start date based on a conservative estimate for when data literacy training may have begun for K–12 teachers as this is when the first federal legislation to directly call for the use of data to support instruction was passed (P.L. 94-142, now the Individuals With Disabilities Education Act).

Search terms. Search terms were developed with the aid of a librarian to capture (a) participants (teacher, educator, and instructor), (b) training (training, professional development, career development, inservice, in service, teacher education, teacher preparation, learning communit*, communit* of practice, professional education, teacher learning, [teacher n2 mentor*], [teacher n2 coach*], educational coach*, and instructional coach*), and (c) the data focus of the training (data literacy, data us*, [us* n2 data], dbi, data based, evidence base*, data driven, curriculum based measur*, [progress* n2 monitor*], DBDM, CBM, and formative assessment). All search terms were entered into a single search. Within the three subject areas (i.e., participants, training, and data focus), terms were separated with the term "or" so that a study would be included if it contained any of the terms. Between the three subject areas, the terms were separated with "and" so that a study had to include at least one term within each subject area.

Inclusion criteria. To be included in the initial search, studies had to (a) involve either preservice or in-service teachers of students in kindergarten through 12th grade of all subject areas, (b) have data literacy training as an independent variable, (c) report at least one dependent measure of teacher outcomes, and (d) utilize an experimental, quasi-experimental, or repeated measures design (i.e., pre–post single group design). Studies were excluded if they were of non-experimental (e.g., action research, descriptive, survey) or single-case design. For the purposes of this meta-analysis, *data literacy training* referred to direct instruction related to the data literacy components (i.e., location, comprehension, and interpretation; Mandinach & Gummer, 2016). *Data* referred to any quantifiable information that helps educators know more about their students academically for the purpose of instructional decision-making (Klinger et al., 2015).

Study screening

Database search. We conducted a search of electronic databases, including PsycINFO, Education Resources Information Center (ERIC), Education Source, and Academic Search Complete within the EBSCO electronic library, using the search terms described above to locate peer-reviewed studies. We used the same search terms and databases, with the addition of ProQuest, to locate unpublished dissertations and reports. After duplicates were removed, the search yielded 5,411 published studies (i.e., peer-reviewed) and 3,895 unpublished studies (i.e., non-peer-reviewed studies, dissertations, and reports). The primary author conducted abstract screening and eliminated 5,043 published studies and 3,857 unpublished studies that did not meet inclusion criteria based on abstract review. Three hundred sixty-eight published studies and 38 unpublished studies were screened by the primary author in full to determine eligibility. Most commonly, studies were excluded because they did not meet the criteria for study design (e.g., action research: Dillon

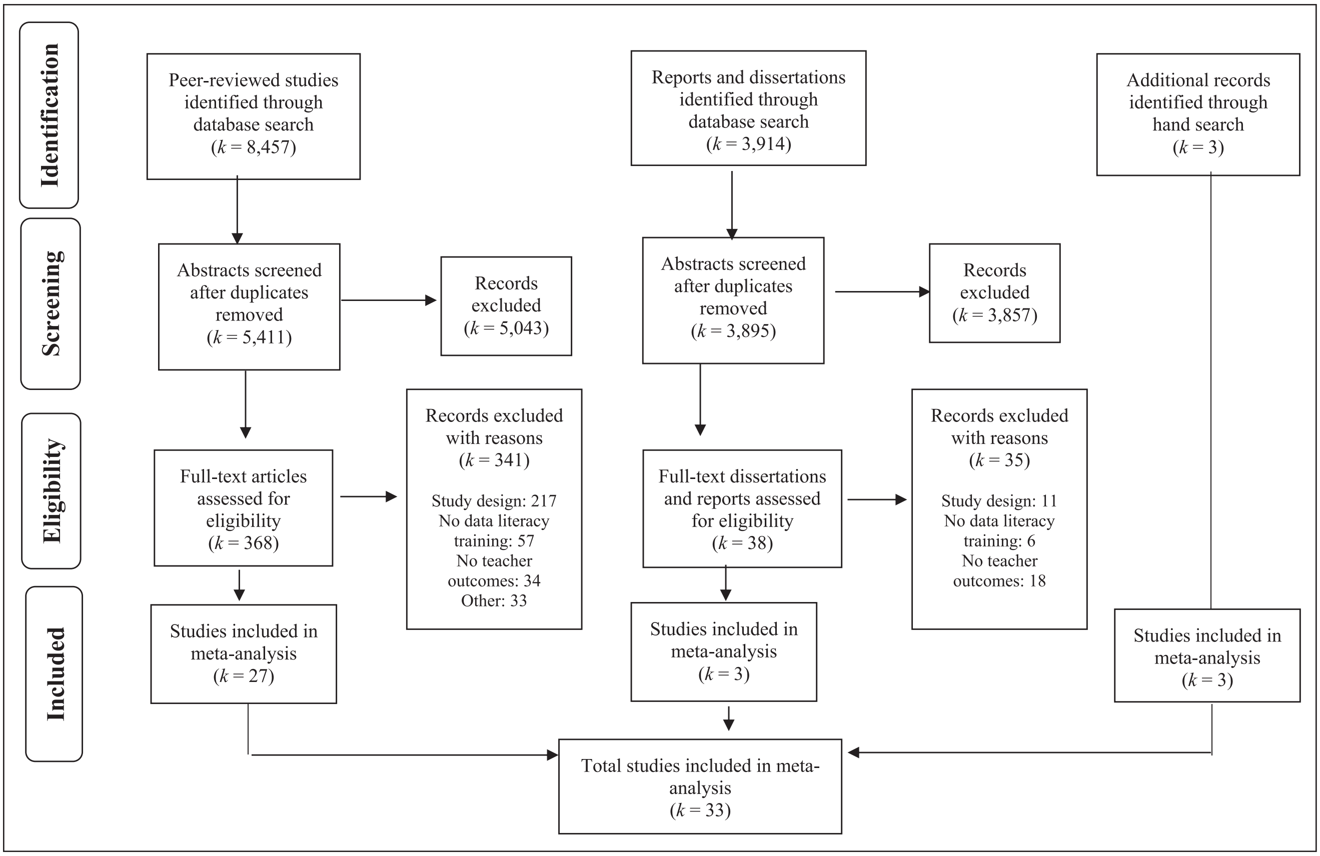

Figure 2. Search procedures for studies on data literacy teacher training.

et al., 2015; qualitative: Breiter & Light, 2006; case study: Lifter et al., 2005; $k = 228$). Other primary reasons for exclusion included not meeting the operational definition for data training (e.g., Boulton, 2010; Graney, 2008; $k = 63$), not including teacher outcomes (e.g., Griffin et al., 2010; Rule et al., 1990; $k = 52$), and other (e.g., duplicates not removed by software program, $k = 30$; not available in English, $k = 3$). If studies qualified but did not provide necessary information to calculate effect sizes ($k = 5$), the authors were contacted through email. In four cases, the data requested were either not available or a response was not received, resulting in their disqualification. For one study, a previous report was referred to for the missing data. The final number of studies identified through this search was 27 published and three unpublished studies.

Hand search. Finally, the primary author conducted a hand search of articles published in the past 20 years across eight major journals that frequently publish studies on general and special education teacher training: *Remedial and Special Education*, *Journal of Educational Psychology*, *Teaching and Teacher Education*, *The Journal of Special Education*, *Exceptional Children*, *Journal of Teacher Education*, *American Educational Research Journal*, and *Teacher Education and Special Education.* This hand search yielded three additional studies that met inclusion criteria.

Data Analysis Procedures

Coding features. To align with Desimone's (2009) conceptual framework of data literacy, we coded outcomes under two broad areas: knowledge and skills, or beliefs (Desimone, 2009). *Knowledge and skills* were measured by direct assessments of teacher conceptual understanding (e.g., What is curriculum-based measurement?) and/or application of that knowledge (e.g., accurate scoring of a curriculum-based measurement). *Beliefs* were tapped by measures of teacher perceptions of both their ability to use data (i.e., self-efficacy) and the value of data use.

Variables of interest, explored through descriptive and moderation analysis, were selected based on Desimone's (2009) framework, subsequent evidence on the framework (i.e., Kennedy, 2016), and evidence surrounding the importance of coaching (e.g., Kraft et al., 2018). We interpreted data as falling into two broad categories: informal and formal data. *Informal* data refers to classroom-oriented sources of data (e.g., quizzes, error analysis, and mastery measures), whereas *formal* data refers to normed data sources (e.g., curriculum-based measurement, benchmark assessment; Klinger et al., 2015). *Coaching* refers to an internal (e.g., support staff) or external (e.g., researcher) individual who provided supplemental training in a small group or 1:1. The *delivery format* was divided into *collaborative*, interactive with teachers working through all or part of the training in teams (includes both active learning and collective participation; e.g., workshops, professional learning communities), and *non-collaborative* or large group training delivered in lecture format (e.g., coursework, sage-on-stage). The content focus was divided into the components of data literacy, including *location* (e.g., assessment design, selection, and administration), *comprehension* (e.g., data analysis), and *interpretation* (e.g., making decisions based on data). *Duration* was divided into high intensity (14 hr or more) or low intensity (13 hr or less; Yoon et al., 2007). Online supplemental Table S1 presents the full coding scheme for each variable of interest.

Coding procedures. A coding protocol for educational intervention research (Vaughn et al., 2014) was adapted to identify (a) design features, (b) participant information, (c) training characteristics, (d) teacher outcome measures, and (e) effect sizes. The code sheet used a combination of forced-choice items (e.g., training format), open-ended items (e.g., number of participants), and written descriptions. Two researchers with prior training and experience with the coding protocol, a doctoral student (primary author) and a doctoral candidate at the time of the study, first independently coded two articles. They then compared coding and discussed discrepancies, adding detail to the coding manual for future reference. Next, all of the remaining articles were independently coded and double-coded by the researchers. Interrater agreement was calculated as a coefficient of the percentage of agreement by dividing the number of agreements by the total number of cells, resulting in 96.46% reliability across the remaining 31 articles (94.53% design features, 97.40% participant information, 94.47% training characteristics, 100% teacher outcome measures, and 96.17% effect sizes). Meetings were held to reach consensus on all disagreements.

Effect size calculation. To account for the small sample sizes in many studies, we used Hedges' g as a measure of effect size, where $g = \frac{M_{PostT} - M_{PostC}}{S_{pooled}}$ and $s_p^2 = \frac{(df_1)s_1^2 + (df_2)s_2^2}{(df_1 + df_2)}$ for independent group designs, and $g = \frac{M_{post} - M_{pre}}{S_{pre}}$ for repeated measures designs. These formulas produce a common metric when using a combination of independent group and repeated measures designs (Morris & DeShon, 2002). Effect sizes were then corrected for small sample bias by multiplying each effect size by $J = 1 - \frac{3}{4m-1}$. For studies of independent group design, variance was calculated where $v = \sqrt{s_p^2}$. For studies of repeated measures design, variance was calculated as $v = \left(\frac{1}{n} + \frac{g^2}{2n}\right)\left[2(1-r)\right]$, where values for the correlations were imputed as $r = \frac{s^2 pre + s^2 post - s^2 D}{2 s_{pre} s_{post}}$.

Accounting for study design. This meta-analysis includes independent group design (i.e., experimental, quasi-experimental) and repeated measures design (i.e., single group pre–post design) study designs. Although some statisticians argue that studies of repeated measures design provide unreliable effect size estimates (Cuijpers et al., 2017) and they are unquestionably less rigorous study designs, other statisticians have supported their inclusion in meta-analyses, particularly when there are many studies of the type for the variable of interest (Koesters, 2017). We ultimately included studies of repeated measures design because these designs are often utilized in educational settings where it is a logistical challenge to conduct studies with a control group, and the data they provide paint a more holistic picture of teacher training (Marsden & Torgerson, 2012). We took steps correspondingly to account for the study design. First, we calculated effects so that they were on the same scale. Second, we conducted a moderator analysis to determine whether there were significant differences based on study design warranting their separate consideration. Independent group studies (g = .58), and repeated measures design (g = .80), were not found to be significantly different (p = .26). As such, we included all studies in our analyses, with study design as a covariate in the full meta-regression model (Morris & DeShon, 2002).

Meta-analytic procedures

Main effects on outcomes. Analyses were run using ROBUMETA in Stata (Hedberg, 2011). We used weighted, random-effects meta-analysis to investigate the main effect size for each outcome of interest (i.e., knowledge and skills, beliefs) due to a range of effects and between-study variance reported for teacher outcomes across training literature (Lipsey & Wilson, 2001). For studies that reported multiple effect sizes from the same sample, we accounted for the statistical dependencies by using the random effects robust standard error estimation technique developed by Hedges and colleagues (2010). This analysis allowed for the clustered data (i.e., effect sizes nested within samples) by correcting studies' standard errors to take into account correlations between effect sizes from the same sample. In all analyses, we used ρ of .80 as sensitivity analyses showed that the findings were robust across different reasonable estimates of $\rho = .80$. The robust standard error technique requires that an estimate of the mean correlation between all pairs of effect sizes within a cluster be estimated for calculating the between-study sampling variance estimate, τ^2. We also calculated heterogeneity with I^2, which measures the percentage of variability in treatment effect estimates that is due to between-study heterogeneity rather than chance. For example, I^2 of 71% suggests 71% of the variability in treatment effect estimates is due to real study differences (heterogeneity) and only 29% due to chance. In contrast, τ^2 is interpreted as the absolute between-study heterogeneity variance (Borenstein et al., 2017).

Moderation analysis. Next, we used weighted, random-effects meta-regression analysis to investigate which moderators might explain the between-study heterogeneity on the main effect, with study design included as a covariate. Moderator analyses were conducted on the following training characteristics pertaining to teacher knowledge and skills: content focus (i.e., inclusion of multiple data literacy skills, inclusion of data interpretation), active learning and collective participation (i.e., collaborative vs. non-collaborative), and coaching (i.e., coach vs. no coach). Due to fewer studies and less variability, moderator analyses could not be conducted for the teacher beliefs outcome. In addition, within the knowledge and skills outcome, we were unable to conduct planned moderator analyses for teacher characteristics and one training characteristic (i.e., high vs. low duration).

Publication bias. Finally, publication bias was examined using Egger et al.'s (1997) regression statistic. This approach tests for asymmetry in the effect sizes as a function of the standard errors reported. Asymmetry of effect sizes may indicate, among other potential factors (i.e., true heterogeneity, chance, sampling variation, and methodological quality), publication bias (e.g., selective reporting of outcomes to promote publication; Sterne et al., 2011). First, we conducted a visual analysis of the funnel plot to evaluate whether there was any asymmetry. Effect sizes were plotted with effect size on the x-axis and standard error on the y-axis. There was a reasonable amount of symmetry with the exception of several positively skewed outliers. We conducted sensitivity analysis by removing the outliers and comparing the results; as the results remained the same without the presence of the outliers, this was an indicator that publication bias did not impact our findings. Finally, we calculated Egger et al.'s (1997) regression statistic and did not find significant asymmetry in effect sizes. As significant asymmetry was not found in the gathered data set, and sensitivity analysis yielded similar results, publication bias likely did not influence the results of this data set (Cooper et al., 2019).

Results

Results are summarized across studies by (a) participant characteristics, (b) a descriptive review of dependent variables and training characteristics, (c) effects of data literacy training on teacher outcomes, and (d) moderator analyses of training characteristics on knowledge and skills outcomes. Tables include a descriptive analysis of study features (see Table 1), moderator analyses (see Table 2), and descriptions of each study, along with measures and effect sizes (see online supplemental Table S2). Forest plots for each outcome of interest can be found in online supplemental Figures S1 and S2.

Participant Characteristics

Across all 33 studies, there was a total of 4,844 participants. These included 3,361 elementary teachers, 728 secondary teachers, and 265 teachers from a range of elementary and secondary grades, with many studies not reporting the grade levels taught ($k = 12$). Participants included 463 preservice teachers and 4,305 in-service teachers, and a combination of preservice and in-service teachers ($n = 76$). There were 246 special education teachers, 2,067 general education teachers, and 2,264 participants were either special education or general education teachers, with several studies not reporting on certification ($k = 3$).

For studies that reported outcomes of teachers' knowledge and skills ($k = 27$), participants taught elementary ($k = 6$, $n = 825$), secondary ($k = 5$, $n = 728$), a range from elementary to secondary levels ($k = 5$, $n = 227$), and not reported ($k = 11$). Participants taught general education ($k = 15$, $n = 1{,}924$), then special education ($k = 8$, $n = 246$), a combination ($k = 2$, $n = 103$), and not reported ($k = 2$). Finally, the majority of participants were in-service teachers ($k = 21$, $n = 2{,}001$), with fewer preservice teachers ($k = 5$, $n = 434$), and one study including preservice and in-service teachers ($n = 76$).

For studies that reported outcomes examining teacher beliefs ($k = 11$), participants taught elementary ($k = 5$, $n = 2{,}286$), secondary ($k = 1$, $n = 90$), a range of elementary and secondary ($k = 2$, $n = 114$), and not reported ($k = 3$). Participants taught special education ($k = 1$, $n = 20$), general education ($k = 5$, $n = 249$), and a combination of general and special education ($k = 3$, $n = 2{,}237$), with several studies not reporting ($k = 2$). A majority of participants were at the in-service level ($k = 7$, $n = 2{,}414$), with fewer participants at the preservice level ($k = 3$, $n = 135$) and one study reporting teachers at the preservice and in-service level ($n = 76$).

Descriptive Features of Data Literacy Training

Dependent variables. Tools used to measure teacher outcomes can be divided into tools that measure the constructs

Table 1. Summary of Training Components Reviewed.

Study name	Study design	Outcome	*n*	In service or preservice	Educator type	Grades	Subject	Training focus	Data source	Intensive duration	Collaborative format	Coach
Albritton and Truscott (2014)	RM	B	18	IS	GE, SE	3 and 5	M	L, C, I	FD	Y	Y	Y
Beesley et al. (2018)	RM	K&S	47	IS	GE	6 to 8	M	C	ID	Y	Y	Y
Castillo et al. (2016)	IG, Q	B	2,143	IS	GE, SE	K to 5	G	L, C, I	FD	Y	N	Y
Chatterji et al. (2009)	IG, Q	B	38	IS	GE	5 to 6	M	L, C	ID	N	Y	Y
Cole (2010)[a]	IG, Q	B, K&S	90	IS	NR	9 to 12	G	NR	ID	N	N	N
Fan et al. (2011)	RM	K&S	47	IS	GE	NR	M	L	ID	Y	N	N
Fayadh (2017)	IG	K&S	33	IS	GE	K to 5	S	L, I	ID	NR	N	N
Förster and Souvignier (2015)	IG	B	43	IS	GE	3	R	L, C	FD	N	Y	N
L. S. Fuchs et al. (1989)	IG	K&S	30	IS	SE	2 to 9	R	L, C, I	FD	N	Y	Y
L. S. Fuchs et al. (1990)	IG	K&S	27	IS	SE	NR	R	L, C, I	FD	N	Y	Y
L. S. Fuchs et al. (1991)	IG	K&S	33	IS	SE	2 to 8	M	C	FD	N	Y	Y
L. S. Fuchs et al. (1992)	IG	K&S	33	IS	SE	1 to 9	R	C	FD	N	Y	Y
L. S. Fuchs and Fuchs (1993)	IG	K&S	22	IS	SE	NR	M	L, C, I	FD	N	Y	Y
L. S. Fuchs et al. (1994)	IG	K&S	40	IS	GE	2 to 5	M	C, I	FD	N	Y	Y
L. S. Fuchs et al. (1999)	IG	K&S	16	IS	GE	2 to 4	M	L, C, I	ID	N	Y	N
Jiminez et al. (2016)	IG	K&S	26	IS	SE	NR	G	L, C, I	FD	N	N	N
Kennedy et al. (2016)	IG	K&S	270	PS	GE	NR	G	C	FD	N	N	N
Kippers et al. (2018)	RM	K&S	27	IS	GE, SE	NR	G	L, C, I	FD	Y	Y	Y
Lembke et al. (2018)	IG	B, K&S	20	IS	SE	1 to 3	W	L, C, I	FD	Y	Y	Y
Martin et al. (2015)	RM	K&S	148	IS	NR	NR	M	L, C, I	ID	Y	Y	N
Newman-Thomas et al. (2012)	IG	K&S	18	PS	GE	NR	R	L	FD	N	N	N
Polly et al. (2018)	IG, Q	K&S	307	IS	GE	K to 1	M	C	ID	Y	Y	N
Randel et al. (2016)	IG	K&S	409	IS	GE	4 to 5	M	L, C	ID	NR	Y	N
Reeves and Honig (2015)	RM	B, K&S	54	PS	GE	NR	G	L, C, I	FD	N	Y	N
Reeves and Chiang (2017)	RM	B, K&S	52	PS	GE	NR	G	C, I	FD	N	Y	N
Reeves and Chiang (2019)	IG	B, K&S	76	PS, IS	GE, SE	PK–12	G	L, C, I	FD	N	Y	N
Riccomini and Stecker (2005)	IG	K&S	40	PS	GE	NR	R	L	FD	N	N	N
Rogers (2015)[a]	RM	B	29	PS	NR	NR	NR	C, I	ID	N	N	N
Schneider and Meyer (2012)	IG	K&S	368	IS	GE	6 to 8	M	L, C	ID	Y	Y	Y
Van der Scheer and Visscher (2016)	IG	B	62	IS	GE	3 to 5	M	L, C, I	FD	Y	Y	Y
Vendlinksi and Phelan (2011)[a]	IG	K&S	60	IS	GE	6 to 8	M	L, C, I	ID	Y	Y	Y
Wesson (1991)	IG	K&S	55	IS	SE	2 to 7	R	L, I	FD	N	Y	Y
Wylie and Lyon (2015)	RM	K&S	163	IS	GE	8 to 12	M	L	ID	Y	Y	Y

Note. RM = repeated measures; B = beliefs; IS = in service; GE = general education; SE = special education; M = math; L = location; C = comprehension; I = interpretation; FD = formal data; Y = yes; K&S = knowledge and skills; ID = informal data; IG = independent group; Q = quasi-experimental; G = general; *N* = no; NR = not reported; S = science; R = reading; PS = preservice.
[a]Unpublished study.

Table 2. Moderators of Data Literacy Training.

Training characteristics	β	*SE*	95% CI	*p* value
Coach[a]	<.00	.21	[–.46, .46]	.98
Format—Collaborative[a,b]	1.31	.39	[.41, 2.21]	.01
Content—Interpretation[a]	.01	.34	[–.80, .82]	.99
Content—Multiple[a]	–.48	.39	[–1.39, .44]	.26
Study Design	.05	.21	[–.41, .51]	.74

Note. All moderators were included in one meta-regression model. For coach, the reference group was studies without a coach. For format, the reference group was studies that did not contain a collaborative workshop element. For content—interpretation, the reference group was studies that did not include the interpretation skill in training. For content—multiple, the reference group was studies with training that did not cover multiple data literacy skills. CI = confidence interval.

[a]Study design was included as a covariate along with the moderators in this model. The between-study sampling variance (τ^2) for the full model is .31 (I^2 = 82.86%). [b]Significant moderator.

of knowledge (k = 12, n = 17), accuracy (k = 9, n = 52), frequency (k = 9, n = 21), self-efficacy (k = 9, n = 38), and teacher beliefs (k = 8, n = 34). Twenty-two studies measured only one construct, whereas 11 measured multiple constructs. For knowledge and skills outcomes, measures of accuracy included direct observations (L. S. Fuchs et al., 1992, 1994; L. S. Fuchs & Fuchs, 1993; Wesson, 1991), evaluation of work samples (Beesley et al., 2018; Vendlinksi & Phelan, 2011), assessment scoring accuracy (Newman-Thomas et al., 2012; Riccomini & Stecker, 2005), and questionnaires (L. S. Fuchs et al., 1989). All measures of frequency were self-reported and evaluated for frequency of assessment (Cole, 2010; Martin et al., 2015; Polly et al., 2018; Wylie & Lyon, 2015), program adjustments (L. S. Fuchs et al., 1992, 1991), skills taught (L. S. Fuchs et al., 1994), and data use practices (Reeves & Chiang, 2019). Finally, knowledge was evaluated with direct tests. Seven studies included only multiple choice (Fan et al., 2011; Fayadh, 2017; Lembke et al., 2018; Newman-Thomas et al., 2012; Randel et al., 2016; Reeves & Chiang, 2017; Schneider & Meyer, 2012), three used a combination of multiple choice and open response (Jiminez et al., 2016; Kennedy et al., 2016; Reeves & Honig, 2015), and two used only open-ended responses (L. S. Fuchs et al., 1999; Kippers et al., 2018). Tests included concepts related to reliability and validity of measurement tools (Fan et al., 2011), curriculum-based measurement (Kennedy et al., 2016; Lembke et al. 2018; Newman-Thomas et al., 2012), design of formative assessment (Fayadh, 2017; Randel et al., 2016; Schneider & Meyer, 2012), and how assessment might be used to inform instruction (L. S. Fuchs et al., 1999; Jiminez et al., 2016; Kennedy et al., 2016; Kippers et al., 2018; Lembke et al., 2018; Reeves & Chiang, 2017; Reeves & Honig, 2015). Several studies based their tests on preexisting measurements (Fan et al., 2011; Kippers et al., 2018; Reeves & Honig, 2015), whereas the remaining developed their own.

Self-efficacy and teacher beliefs measures relied on rating scales and, correspondingly, were self-reported. For self-efficacy, six studies included previously validated scales (Albritton & Truscott, 2014; Lembke et al., 2018; Reeves & Chiang, 2017, 2019; van der Scheer & Visscher, 2016), whereas three developed their own measures for the study (Chatterji et al., 2009; Cole, 2010; Rogers, 2015). The aspects of self-efficacy measured related to problem-solving (Albritton & Truscott, 2014), personal and general teaching strategies (Chatterji et al., 2009; Lembke et al., 2018; van der Scheer & Visscher, 2016), data use to inform instruction (Reeves & Chiang, 2017; Reeves & Chiang, 2019; Reeves & Honig, 2015), and measurement design (Cole, 2010; Rogers, 2015). For beliefs, four studies used previously validated scales (Castillo et al., 2016; Reeves & Chiang, 2017, 2019; Reeves & Honig, 2015), whereas four used researcher-developed measures (Chatterji et al., 2009; Cole, 2010; Förster & Souvignier, 2015; Rogers, 2015). Beliefs scales measured teacher anxiety surrounding data-based decisions (Reeves & Chiang, 2017; Reeves & Chiang, 2019), attitudes toward the effectiveness of data use (Castillo et al., 2016; Chatterji et al., 2009; Cole, 2010; Förster & Souvignier, 2015; Reeves & Honig, 2015), perceived institutional support for data use (Chatterji et al., 2009), the ease of data collection and evaluation (Förster & Souvignier, 2015), and the importance of multiple data sources for decision-making (Rogers, 2015).

High-quality training features

Content focus. With regard to the data source examined in training sessions, formal data (k = 20) were more often examined than informal data (k = 13). Content covered for formal data included the theory behind progress monitoring (Albritton & Truscott, 2014; Castillo et al., 2016), scoring procedures (Newman-Thomas et al., 2012; Riccomini & Stecker, 2005), the data-based decision-making process (Albritton & Truscott, 2014; L. S. Fuchs et al., 1989, 1990, 1991, 1992, 1994; L. S. Fuchs & Fuchs, 1993; Jiminez et al., 2016; Kennedy et al., 2016; Kippers et al., 2018; Lembke et al., 2018; Reeves & Chiang, 2017, 2019; Reeves & Honig, 2015; van der Scheer & Visscher, 2016; Wesson, 1991), the response to intervention process (Castillo et al., 2016), error analysis (L. S. Fuchs et al., 1990), and using data to adjust instruction (Förster & Souvignier, 2015; L. S. Fuchs et al., 1992, 1994). Content covered for informal data included the theory behind formative assessment (Beesley et al., 2018; Cole, 2010; Fayadh, 2017), the properties of assessments (Fan et al., 2011), designing assessments (Fan et al., 2011; L. S. Fuchs et al., 1999; Martin et al., 2015; Randel et al., 2016; Schneider & Meyer, 2012; Wylie & Lyon, 2015), conducting error analysis (Chatterji et al., 2009; Polly et al., 2018; Vendlinksi & Phelan, 2011), providing feedback to students (Cole, 2010; Martin et al., 2015; Wylie & Lyon, 2015), tracking progress (Castillo et al., 2016; Martin et al., 2015; Randel et al., 2016; Rogers, 2015), and using data to adjust instruction (Castillo et al., 2016; L. S. Fuchs et al., 1999; Martin et al., 2015; Polly et al., 2018; Rogers, 2015; Schneider & Meyer, 2012; Vendlinski & Phelan, 2011). Nine studies covered a single component of data literacy, whereas twenty-three covered multiple components and one did not report.

Active learning and collective participation. A majority of studies included collaborative training components ($k = 24$). Ten of these delivered part of the intervention in collaborative groups and part in lecture or online learning formats (Albritton & Truscott, 2014; Beesley et al., 2018; Chatterji et al., 2009; Martin et al., 2015; Polly et al., 2018; Randel et al., 2016; Reeves & Chiang, 2019; Schneider & Meyer, 2012; Vendlinksi & Phelan, 2011; Wesson, 1991). The remaining 14 studies implemented the entire training in collaborative groups.

Duration. Twelve studies had an intensive duration (i.e., more than 14 hr), 19 studies were conducted over less than 14 hr, and two studies did not report this information (Fayadh, 2017; Randel et al., 2016).

Coaching. Seventeen studies included supplemental coaching. Coaches supported teachers' abilities to design assessment (Chatterji et al., 2009; Schneider & Meyer, 2012; Wylie & Lyon, 2015), administer assessments (Albritton & Truscott, 2014; Kippers et al., 2018; Lembke et al., 2018), conduct error analysis (Chatterji et al., 2009; L. S. Fuchs et al., 1990, 1994; Vendlinski & Phelan, 2011), follow the response to intervention model (Castillo et al., 2016), inspect progress monitoring graphs (L. S. Fuchs et al., 1989, 1990, 1991, 1992, 1994; L. S. Fuchs & Fuchs, 1993; Kippers et al., 2018; Lembke et al., 2018; van der Scheer & Visscher, 2016; Wesson, 1991), target instruction (Albritton et al., 2014; L. S. Fuchs et al., 1989, 1990, 1991, 1992, 1994; Fuchs & Fuchs, 1993; Kippers et al., 2018; Lembke et al., 2018; van der Scheer & Visscher, 2016; Wesson, 1991), and provide feedback (Beesley et al., 2018).

Effects of Data Literacy Training

Thirty-three studies met criteria for inclusion in this review of the literature. Four studies were quasi-experimental, nine were repeated measures design, and the remaining studies were experimental. Twenty-seven studies included assessments of teacher knowledge and skills, with a total of 91 effect sizes resulting in a weighted mean effect size of $g = .67$ (95% CI = [0.40, 0.93], $\tau^2 = .35$, $I^2 = 86.51\%$). Eleven studies included outcomes pertaining to teacher beliefs, with 72 effect sizes producing a weighted mean effect size of $g = .48$ (95% CI = [0.17, 0.79], $\tau^2 = .22$, $I^2 = 88.86\%$). Five studies included measures of both knowledge and skills, and beliefs; thus, these studies had relevant effect sizes included in both analyses.

Moderator Analysis

A moderator analysis for the knowledge and skills outcome was run in one meta-regression model that included coaching, delivery format, multiple data literacy skills, and data interpretation skills ($\tau^2 = .31$, $I^2 = 80.96\%$). Research design was included as a covariate; nine studies were of repeated measures design and this did not moderate effects ($\beta = .05$, 95% CI = [–0.41, 0.51). Fourteen studies included coaching, and it was determined that supplemental coaching did not influence the effects of training ($\beta = -.004$, 95% CI = [–0.46, 0.46]). Twenty-two studies included a collaborative format, and this had a significant positive impact on teacher knowledge and skills outcomes ($\beta = 1.31$, 95% CI = [0.41, 2.21]). Twenty-one studies covered multiple topics in training and this was found not to significantly influence the effects of training ($\beta = -.48$, 95% CI = [–1.39, 0.44]). Eighteen studies included interpretation of data and this also did not moderate effects ($\beta = .01$, 95% CI = [–0.80, 0.82]).

Discussion

The purpose of this meta-analysis was to evaluate the effects of teacher training in data literacy on outcomes related to teacher knowledge, skills, and beliefs pertaining to data use. Three research questions were asked, as follows: (a) What are the features of data literacy training for K–12 teachers? (b) What are the effects of data literacy training on K–12 teacher outcomes? and (c) Do training characteristics moderate the effects of data literacy training? Results indicated that training had a significant positive effect on teacher outcomes, and trainings with a collaborative component were associated with higher effects on knowledge and skills.

What Are the Effects of Data Literacy Training?

Overall, findings suggest that teacher training in data literacy has significant positive effects on both teacher knowledge and skills, and teacher beliefs. The average effect was large for knowledge and skills (g = .67), and moderate for teacher beliefs (g = .48; Cohen, 1988), with both more than large enough to be of interest to policy makers in the field of education (Hedges & Hedberg, 2007). As the underlying framework for data use emphasizes the importance of teachers in making meaning out of data to make instructional changes and improve student outcomes, this finding lends support to training teachers on this critical topic.

Despite the potential of training, research consistently indicates that teachers are not using formal data sources to inform their practice although they have increasing access to these data (Datnow & Hubbard, 2016; Gallagher et al., 2008; Mandinach & Gummer, 2016; Means et al., 2011). A wealth of studies has demonstrated the importance of teacher beliefs on their actual practice in the classroom (Boardman et al., 2005; Datnow & Hubbard, 2016; Hoogland et al., 2016; Jimerson & Wayman, 2012). As there was a slightly smaller average effect associated with teacher beliefs, as a result of training, it is important to explore potential reasons for this finding to promote more effective trainings for beliefs and to promote data use in practice.

Belief outcomes were separated into self-efficacy and the value of data use. There was a wide range of effects for measures of self-efficacy (g = –0.74 to g = 3.68) and value (g = –0.43 to g = 1.25). As there were more negative effects associated with beliefs about the value of data use, it seems that teachers' beliefs surrounding the value of data in general may be less malleable than their beliefs in their own ability to use data. This is an important distinction as buy-in is essential when it comes to long-term improvements in practice (Hoogland et al., 2016; Jimerson & Wayman, 2012). However, only three of the 11 studies that measured beliefs explicitly targeted the rationale behind data use (Albritton et al., 2014; Castillo et al., 2016; Cole, 2010). Rather than beliefs about the value of data use potentially being less malleable, it is possible that explicit training on the rationale for data use is needed to improve outcomes and practice.

The lack of persistence with data use is also an important factor when considering the ways in which outcomes are being measured. Across all of the studies on training, only one included a maintenance measure (Kennedy et al., 2016). Teachers in this study did maintain their knowledge and were able to apply it in a hypothetical scenario, but the authors did not examine whether teachers transferred this into practice (i.e., sustained data use in the classroom). Several studies used instructional plan sheets (L. S. Fuchs et al., 1989, 1994) and accuracy of implementation scales (e.g., Wesson, 1991) to determine frequency and accuracy of data use in the classroom setting during the training; however, the authors did not include a maintenance measure and, correspondingly, this too does not provide information about practice. The real-world applications of data literacy training remain largely unexplored in experimental research.

What Moderates the Effects of Training?

Moderator analyses of the knowledge and skills outcome revealed that the *content focus* of training (i.e., interpretation component, multiple data literacy components) did not influence the effects of training. These findings could be due to the measures used; specifically, most of the measures were directly aligned with the content that teachers received in training and were implemented immediately after training. It is therefore possible that teachers learned the skills targeted regardless of the content taught. This finding also contrasts with the study conducted by Kennedy (2016), which found that trainings that included a clear application to improve student learning were associated with larger effects. Prior studies of teacher data use have found that the act of collecting data improves student learning regardless of decision-making (Stecker et al., 2005). It is therefore possible that each of the components of data literacy can be connected to student learning, explaining the lack of differences between focus on each of the components.

Active learning and collective participation moderated effects, in that the effects of trainings with a collaborative format were significantly higher. This is no surprise as learning is theorized to be a social endeavor occurring through active participation in a community of practice (e.g., Vygotsky, 1978). Professional development literature in general espouses the importance of collaboration during training (e.g., Desimone, 2009; Yoon et al., 2007). It is possible that this component is especially important for data literacy training as collaboration among teachers is documented as a key feature of programs that enhance teachers' use of data for instruction (Marsh, 2012). Collaboration may also be important to promote the use of data in practice; for instance, teachers in small groups are observed to more accurately interpret data, clarify problems, ask follow-up questions, problem-solve, and correct errors (Means et al., 2011).

Finally, results of this review indicate that the presence of a *coach* did not moderate effects. This finding contradicts the emphasis on coaching seen across qualitative training literature (Datnow & Hubbard, 2016) and findings from a recent meta-analysis that found coaching to be effective across content areas (g = .49; Kraft et al., 2018). However, our findings align with a review of training in general, which found that coaching did not significantly moderate effects (Brock & Carter, 2017). It is possible that the collaborative component is more essential when it comes to data literacy training as compared with other trainings. This is illuminated when the results of studies that compared groups who received coaching with those who participated in a collaborative group are investigated. For instance, Wesson (1991) compared a group that received follow-up expert consultation alone with a group that received structured collaboration groups in addition to the follow-up expert consultation. The collaborative training group resulted in more positive effects than the group that only received expert coaching. L. S. Fuchs and colleagues (1992) compared collaborative training alone with collaborative training along with additional coaching and found both conditions to be comparable. Additional research that compares these methods is needed to determine whether and in what conditions coaching may support data literacy training.

How Can Data Literacy Training Features Help Contextualize Findings?

We observed important trends in the studies to contextualize the findings of the meta-analysis and subsequent moderator analysis. First, there was a notable amount of heterogeneity present for the knowledge and skills outcome ($\tau^2 = .35$, $I^2 = 86.51\%$), which was not able to be fully explained by theory-driven moderators ($\tau^2 = .31$, $I^2 = 80.96\%$). In other words, variability due to study differences decreased from 86.51% to 80.96%, indicating that at least some of the heterogeneity was explained by moderators. The remaining heterogeneity may be due in part to the variety of types of measurements (e.g., direct assessments, observation, and self-report) and topics covered by these measurements (e.g., data-based decision-making, accuracy of scoring procedures). It is also possible that there were differential effects for proximal researcher-developed tools compared with distal previously validated scales as larger effects have been noted for more proximal scales. We did detect a significant positive effect that indicates that the heterogeneity did not prevent a notable trend from being observed. Rather, this information provides an additional avenue for understanding the remaining heterogeneity.

Limitations and Future Research

There are several limitations to this review, the first of which is the potential for publication bias. We took efforts to ensure that this was not likely within this article. First, we searched for and located unpublished literature, including dissertations and reports. Second, we used Egger et al.'s (1997) publication bias statistics and visual inspection of funnel plots. However, we note that it is possible that studies were still missed despite these precautions. The second limitation is the inclusion of studies of less rigorous design. As our intention was to provide an extensive look at the research on data literacy training, we chose to include these studies. We included steps to ensure that the inclusion of these studies did not impact results; however, we recommend readers interpret findings within this context. The third limitation is in the types of measures that were used. There are no standardized measures of teacher knowledge and skills; therefore, many studies included proximal measures and/or subjective rating scales. A fourth related limitation is the notable heterogeneity observed in the data set, which is commonly observed in meta-analyses (Higgins et al., 2003). We were also unable to conduct some planned moderator analyses (e.g., duration, participant characteristics) because there were few studies that included the variable of interest or reported the information needed. We included a descriptive analysis of training features to provide potential reasons for the heterogeneity and to evaluate each variable of interest to counteract these limitations. Future research may report on each of these variables of interest. Another limitation is related to the lack of maintenance measures of classroom application used in the studies; thus, there is still not an indication of how training translates into practice. Future research that evaluates how training impacts long-term effects on practice is necessary. In addition, coaching was rarely described with replicable precision. Given the challenges that teachers face with data use in practice, future research is needed that evaluates various coaching models that impact teachers' data literacy.

Implications

This review yields several important implications for the design of data literacy training and promotion of data use in the classroom. Most importantly, training in data literacy has a direct positive impact on teachers' data literacy outcomes. Training is therefore an important first step; however, we require additional knowledge about the mechanisms to maintain these outcomes and transfer learning into practice. Another important finding is that a collaborative component increases the effects of training. There is evidence to suggest that collaborative groups that discuss data improve data use (e.g., Schildkamp & Poortman, 2015). As such, these teams may be important not only in training but also as an ongoing structure to promote data use in practice. As expectations for teacher data use continue to rise, it is vital to support teachers with training targeted to meet their needs and, ultimately, to improve outcomes for the students they serve.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported in part by Grant H325D150056 from the Office of Special Education Programs, U.S. Department of Education. These contents do not necessarily represent the policy of this agency and one should not assume endorsement by the federal government.

Supplemental Material

Supplementary material for this article is available on the *Remedial and Special Education* website with the online version of this article.

References

*Studies preceded by an asterisk were included in the meta-analysis

*Albritton, K., & Truscott, S. (2014). Professional development to increase problem-solving skills in a response to intervention framework. *Contemporary School Psychology*, *18*, 44–58.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

*Beesley, A. D., Clark, T. F., Dempsey, K., & Tweed, A. (2018). Enhancing formative assessment practice and encouraging middle school mathematics engagement and persistence. *School Science and Mathematics*, *118*, 4–16.

Boardman, A. G., Argüelles, M. E., Vaughn, S., Hughes, M. T., & Klingner, J. (2005). Special education teachers' views of research-based practices. *The Journal of Special Education*, *39*, 168–180.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572.

Breiter, A., & Light, D. (2006). Data for school improvement: Factors for designing effective information systems to support decision-making in schools. *Journal of Educational Technology & Society*, *9*(3), 206–217.

Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education, 38*(3), 131–144.

*Castillo, J. M., March, A. L., Tan, S. Y., Stockslager, K. M., & Brundage, A. (2016). Relationships between ongoing professional development and educators' beliefs relative to response to intervention. *Journal of Applied School Psychology*, *32*, 287–312.

*Chatterji, M., Koh, N., Choi, L., & Iyengar, R. (2009). Closing learning gaps proximally with teacher-mediated diagnostic classroom assessment. *Research in the Schools*, *16*, 59–75.

Civic Impulse. (2017). H.R. 1532—112th Congress: Race to the Top Act of 2011. https://www.govtrack.us/congress/bills/112/hr1532

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Cole, M. W. (2010). *Influence of assessment for learning professional development in rural Georgia public schools* [Unpublished dissertation]. Liberty University.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Council of Chief State School Officers. (2013). *Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards and learning progressions for teachers. 1.0*.

Cuijpers, P., Weitz, E., Cristea, I., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*(4), 364–368.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession*. National Staff Development Council.

Data Quality Campaign (2016, April). *Time to act: Making data work for students*. https://dataqualitycampaign.org/wp-content/uploads/2016/04/Time-to-Act.pd

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, *117*(4). https://www.researchgate.net/publication/274136957_Teachers'_Use_of_Assessment_Data_to_Inform_Instruction_Lessons_From_the_Past_and_Prospects_for_the_Future

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*, 7–28.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181–199

Didion, L., Toste, J. R., & Filderman, M. J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness, 13*(1), 29–66.

Dillon, P., Erkens, C., Sanna, D., & Savastano, L. F. (2015). Crowdlearning. *JSD the Learning Forward Journal*, *36*(3), 28–31.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.

Endrew F. v. Douglas County School District, 798 F. *3d* 1329 (10th Cir. 2015).

Espin, C. A., Wayman, M. M., Deno, S. L., McMaster, K. L., & de Rooij, M. (2017). Data-based decision-making: Developing a method for capturing teachers' understanding of CBM graphs. *Learning Disabilities Research & Practice, 32*(1), 8–21

Every Student Succeeds Act (2015). Pub. L. No. 114-95. In *114th Congress*.

Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children, 8*1(2), 227–246.

*Fan, Y. C., Wang, T. H., & Wang, K. H. (2011). A web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education, 57*(2), 1727–1740.

*Fayadh, A. (2017). Effectiveness of the proposed training formative assessment programme and its impact on teaching style improvements of Saudi science teachers in Saudi Arabia. *Journal of Turkish Science Education*, *14*, 35–56.

Filderman, M. J., Toste, J. R., & Cooc, N. (2020). Does training predict second-grade teachers' use of student data for decision-making in reading and mathematics? *Assessment for Effective Intervention*, *46*, 247–258.

Filderman, M. J., Toste, J. R., Didion, L. A., Peng, P., & Clemens, N. H. (2018). Data-based decision making in reading interventions: A synthesis and meta-analysis of the effects for struggling readers. *The Journal of Special Education*, *52*(3), 174–187.

*Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students' reading progress. *School Psychology Review*, *44*, 60–75.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*, 263–279.

*Fuchs, L. S., Allinder, R. M., Hamlett, C. L., & Fuchs, D. (1990). An analysis of spelling curricula and teachers' skills in identifying error types. *Remedial and Special Education*, *11*, 42–52.

*Fuchs, L. S., & Fuchs, D. (1993). Effects of systematic observation and feedback on teachers' implementation of curriculum-based measurement. *Teacher Education and Special Education*, *16*, 178–187.

*Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading maze task. *Exceptional Children*, *58*, 436–450.

*Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children*, *60*, 518–537.

*Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, *28*, 617–641.

*Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal*, *36*, 609–646.

*Fuchs, L. S., Fuchs, D., & Stecker, P. M. (1989). Effects of curriculum-based measurement on teachers' instructional planning. *Journal of Learning Disabilities*, *22*, 51–59.

Gallagher, L., Means, B., & Padilla, C. (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007*. U.S. Department of Education.

Graney, S. B. (2008). General education teacher judgments of their low-performing students' short-term reading progress. *Psychology in the Schools*, *45*(6), 537–549.

Griffin, P., Murray, L., Care, E., Thomas, A., & Perri, P. (2010). Developmental assessment: Lifting literacy through professional learning teams. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 383–397.

Gummer, E., & Mandinach, E. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, *117*(4), 1–22.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*, 495–500.

Hedberg, E. C. (2011). *Robumeta: Stata module to perform robust variance estimation in meta-regression with dependent effect size estimates* [Stata ado file]. https://www.researchgate.net/publication/254395315_ROBUMETA_Stata_module_to_perform_robust_variance_estimation_in_meta-regression_with_dependent_effect_size_estimates

Hedges, L. V., & Hedberg, E. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*, 60–87.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560.

Hoogland, I., Schildkamp, K., Van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, *60*, 377–386.

*Jimenez, B. A., Mims, P. J., & Baker, J. (2016). The effects of an online data-based decisions professional development for in-service teachers of students with significant disability. *Rural Special Education Quarterly*, *35*(3), 30–40.

Jimerson, J. B., & Wayman, J. C. (2012, April). *Branding educational data use through professional learning: Findings from a study in three school districts*. Annual Meeting of the American Educational Research Association Vancouver, BC, Canada.

Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice*, *33*, 144–155.

*Kennedy, M. J., Wagner, D., Stegall, J., Lembke, E., Miciak, J., Alves, K. D., Brown, T., Driver, M. K., & Hirsch, S. E. (2016). Using content acquisition podcasts to improve teacher candidate knowledge of curriculum-based measurement. *Exceptional Children*, *82*, 303–320.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980.

*Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in Educational Evaluation*, *56*, 21–31.

Klinger, D. A., McDivitt, P. R., Howard, B. B., Munoz, M. A., Rogers, W. T., & Wylie, E.C. (2015). *The classroom assessment standards for preK-12 teachers: Joint committee on standards for educational evaluation*. Kindle Direct Press.

Koesters, M. (2017). Every effect size has its place: A commentary on the avoidance of pre–post effect sizes. *Epidemiology and Psychiatric Sciences*, *26*(4), 369–370.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588.

*Lembke, E. S., McMaster, K. L., Smith, R. A., Allen, A., Brandes, D., & Wagner, K. (2018). Professional development for data-based instruction in early writing: Tools, learning, and collaborative support. *Teacher Education and Special Education, 41*(2), 106–120.

Lifter, K., Kruger, L., Okun, B., Tabol, C., Poklop, L., & Shishmanian, E. (2005). Transformation to a web-based preservice training program: A case study. *Topics in Early Childhood Special Education*, *25*(1), 15–24.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE.

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education, 118*(2), 143–166.

Mandinach, E. B. (2012) A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*(2), 71–85.

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376.

Mandinach, E. B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. In E.B. Mandinach, M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 13–31). Teachers College Press.

Marsden, E., & Torgerson, C. J. (2012). Single group, pre-and post-test research designs: Some methodological concerns. *Oxford Review of Education*, *38*(5), 583–616.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1–48.

*Martin, C. S., Polly, D., Wang, C., Lambert, R. G., & Pugalee, D. K. (2016). Perspectives and practices of elementary teachers using an internet-based formative assessment tool: The case of assessing mathematics concepts. *International Journal for Technology in Mathematics Education, 23*(1), 3–11.

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. U.S. Department of Education.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools: Teacher access, supports and use*. U.S. Department of Education.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125.

National Assessment of Educational Progress. (2019). *The nation's report card: 2019 mathematics and reading assessments*. National Center for Educational Statistics.

National Center on Intensive Intervention. (2013). *Data-based individualization: A framework for intensive intervention*. Office of Special Education, U.S. Department of Education.

National Council for the Accreditation of Teacher Education. (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers: A national strategy to prepare effective teachers*.

*Newman-Thomas, C., Smith, C. A., Zhao, X., Kethley, C. I., Rieth, H. J., Swanson, E. A., & Heo, Y. (2012). Technology-based practice to teach preservice teachers to assess oral reading fluency. *Journal of Special Education Technology*, *27*, 15–32.

Piro, J. S., Dunlap, K., & Shutt, T. (2014). A collaborative Data Chat: Teaching summative assessment data use in pre-service teacher education. *Cogent Education, 1*(1), 1–24.

*Polly, D., Wang, C., Martin, C., Lambert, R., Pugalee, D., & Middleton, C. (2018). The influence of mathematics professional development, school-level, and teacher-level variables on primary students' mathematics achievement. *Early Childhood Education Journal*, *46*, 31–45.

*Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research*, *109*, 491–502.

*Reeves, T. D., & Chiang, J. L. (2017). Building pre-service teacher capacity to use external assessment data: An intervention study. *The Teacher Educator*, *52*, 155–172.

*Reeves, T. D., & Chiang, J. L. (2019). Effects of an asynchronous online data literacy intervention on pre-service and in-service educators' beliefs, self-efficacy, and practices. *Computers & Education, 136*, 13–33

*Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, *50*, 90–101.

Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, *27*, 472–482.

*Riccomini, P. J., & Stecker, P. M. (2005). Effects of technology-enhanced practice on scoring accuracy of oral reading fluency. *Journal of Special Education Technology*, *20*, 5–12.

*Rogers, M. A. (2015). *A developmental study examining the value, effectiveness, and quality of a data literacy intervention* [Unpublished dissertation]. University of Iowa.

Rule, S., Fodor-Davis, J., Morgan, R., Salzberg, C. L., & Chen, J. (1990). An inservice training model to encourage collaborative consultation. *Teacher Education and Special Education*, *13*(3–4), 225–227.

Schildkamp, K., & Poortman, C. (2015). Factors influencing the functioning of data teams. *Teachers College Record*, *117*(4), 1–42.

*Schneider, M. C., & Meyer, J. P. (2012). Investigating the efficacy of a professional development program in formative classroom assessment in middle school English language arts and mathematics. *Journal of Multi-disciplinary Evaluation*, *8*(17), 1–24.

*Schütze, B., Rakoczy, K., Klieme, E., Besser, M., & Leiss, D. (2017). Training effects on teachers' feedback practice: The mediating function of feedback knowledge and the moderating role of self-efficacy. *ZDM Mathematics Education*, *49*, 475–489.

Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 4*1(2), 160–175.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal, 343,* d4002.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795–819.

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal*, *110*, 228–245.

van den Bosch, R. M., Espin, C. A., Chung, S., & Saab, N. (2017). Data-based decision-making: Teachers' comprehension of curriculum-based measurement progress-monitoring graphs. *Learning Disabilities Research & Practice*, *32*, 46–60.

*van der Scheer, E. A., & Visscher, A. J. (2016). Effects of an intensive data-based decision making intervention on teacher efficacy. *Teaching and Teacher Education*, *60*, 34–43.

Vaughn, S., Elbaum, B., Wanzek, J., Scammacca, N., & Walker, M. (2014). *Code sheet and guide for education-related intervention study syntheses*. The Meadows Center for Preventing Educational Risk.

*Vendlinski, T. P., & Phelan, J. (2011). *Using key conceptual ideas to improve teacher use of formative assessment data* (CRESST Report 794). University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Vygotsky, L. S. (1978). *Mind in Society*. Harvard University Press.

Wagner, D. L., Hammerschmidt-Snidarich, S. M., Espin, C. A., Seifert, K., & McMaster, K. L. (2017). Pre-service teachers' interpretation of CBM progress monitoring data. *Learning Disabilities Research & Practice*, *32*, 22–31.

Wayman, J. C., Midgley, S., & Stringfield, S. (2017). Leadership for data-based decision making: Collaborative educator teams. In A. Danzig, K. Borman, B. Jones & W. Wright (Eds.), *Learner-centered leadership* (pp. 189–206). Routledge.

Wayman, J. C., & Jimerson, J. B. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation*, *42*, 25–34.

*Wesson, C. L. (1991). Curriculum-based measurement and two models of follow-up consultation. *Exceptional Children*, *57*, 246–256.

*Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, *22*, 140–160.

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement. Issues & answers. REL 2007-No. 033*. Regional Educational Laboratory Southwest, U.S. Department of Special Education.