# A Randomized Trial of Lesson Study with Mathematical Resource Kits: Analysis of Impact on Teachers' Beliefs and Learning Community

**Catherine C. Lewis and Rebecca Reed Perry**

Both theory and empirical findings suggest that improvement of teaching is not a one-shot activity: It requires ongoing effort by teachers, who must integrate improvements into the complex juggling act of classroom practice (e.g., Clarke & Hollingsworth, 2002; Lampert, 2001). To change teaching successfully often requires repeated cycles of classroom trial, reflection, feedback, and revised trial. For example, (Schorr & Koellner-Clark, 2003) chronicle the experience of a teacher who changed his classroom teaching with the intention of having students contribute to the construction of mathematical ideas in the classroom. He was initially pleased with his effort, but when he showed video of his classroom practice to colleagues, they saw the class discussion as lacking in mathematical rigor. Their reactions led him to reevaluate the quality of his class discussions and engage in further work to better establish mathematical focus while building student contributions.

Situations like those described by Schorr and Koellner-Clark (2003), in which teachers must engage in repeated cycles of experimentation and reflection to improve practice, are probably more the rule than the exception, even though models of professional learning impact often show a unidirectional arrow from professional development program to change in instruction (see review by Clarke & Hollingsworth, 2002). If repeated cycles of experimentation are typically needed to improve one's teaching, then what are the implications for the design of professional learning programs? For example, how might we design professional learning programs to catalyze development of the beliefs and dispositions needed to continue such challenging, ongoing work?

This chapter reports a randomized, controlled trial of lesson study supported by mathematical resources. The experimental treatment, described in more detail below, significantly increased the mathematical knowledge of both participating

---

C.C. Lewis (✉) • R.R. Perry
Mills College School of Education, 5000 MacArthur
Boulevard, Oakland, CA 94613, USA
e-mail: clewis@mills.edu

J.A. Middleton et al. (eds.), *Large-Scale Studies in Mathematics Education*,
Research in Mathematics Education, DOI 10.1007/978-3-319-07716-1_7

teachers and students, as reported elsewhere (Gersten, Taylor, Keys, Rolfhus, & Newman-Gonchar, 2014). This chapter focuses, however, on two "Intermediate Outcomes" shown in Fig. 1: teachers' beliefs and teacher learning community. We examine (1) the impact of the intervention (lesson study with mathematical resources) on teachers' beliefs and teacher learning community and (2) the role of teachers' beliefs and teacher learning community as mediators of teachers' and students' increases in mathematical knowledge.

## Background on Lesson Study

As shown on the left side of Fig. 1, lesson study is collaborative, practice-based professional learning in which teachers study the academic content of the curriculum and plan, enact, observe, and analyze a live classroom lesson (Fernandez & Yoshida, 2004; Lewis & Hurd, 2011; Lewis & Tsuchida, 1997, 1998; Perry & Lewis, 2010; Stigler & Hiebert, 1999; Wang-Iverson & Yoshida, 2005). As shown in the center of Fig. 1, the practice-based cycles that comprise lesson study are hypothesized to improve instruction by simultaneously improving five basic inputs to instruction: teachers' knowledge; teachers' beliefs and dispositions; teacher learning community; learning resources and tools; and system features.

Although the term "lesson study" often evokes images of lesson planning, in fact lesson planning is just a small portion of lesson study. It may be useful to think of lesson study and other familiar professional learning approaches as overlapping circles in a Venn diagram to highlight the characteristics lesson study shares with other familiar professional learning approaches. For example, lesson study overlaps with many other professional learning approaches in the shared element of *teachers' study of content knowledge*, which previous research has identified as a feature of effective professional learning (Garet, Porter, Desimone, Birman, & Yoon, 2001) and which in lesson study occurs most heavily during the first parts of the lesson study cycle, when teachers engage in *kyouzai kenkyuu*, the study of content and curriculum materials (Takahashi, Watanabe, Yoshida, & Wang-Iverson, 2005). In Japan, the teacher's manual provides a key resource for *kyouzai kenkyuu*, since it includes discussion of both the curriculum content and of common student thinking and misconceptions (Lewis, Perry, & Friedkin, 2011). In the current study, mathematical resource kits were designed to substitute for the materials available to Japanese teachers as they conduct *kyouzai kenkyuu*.

A second element of lesson study is *observation of live practice*, and this element is shared, for example, with many professional learning programs that include coaching or mentoring (e.g., Campbell & Malkus, 2011), although lesson study focuses on observation of students rather than on critique of teaching. A third element of lesson study is *analysis of student thinking and student work*; again, many well-known professional learning approaches strongly emphasize this element (Carpenter, Fennema, Franke, Levi, & Empson, 1999). In lesson study, teachers observe students *as* they think and work, as well as analyzing student work products.

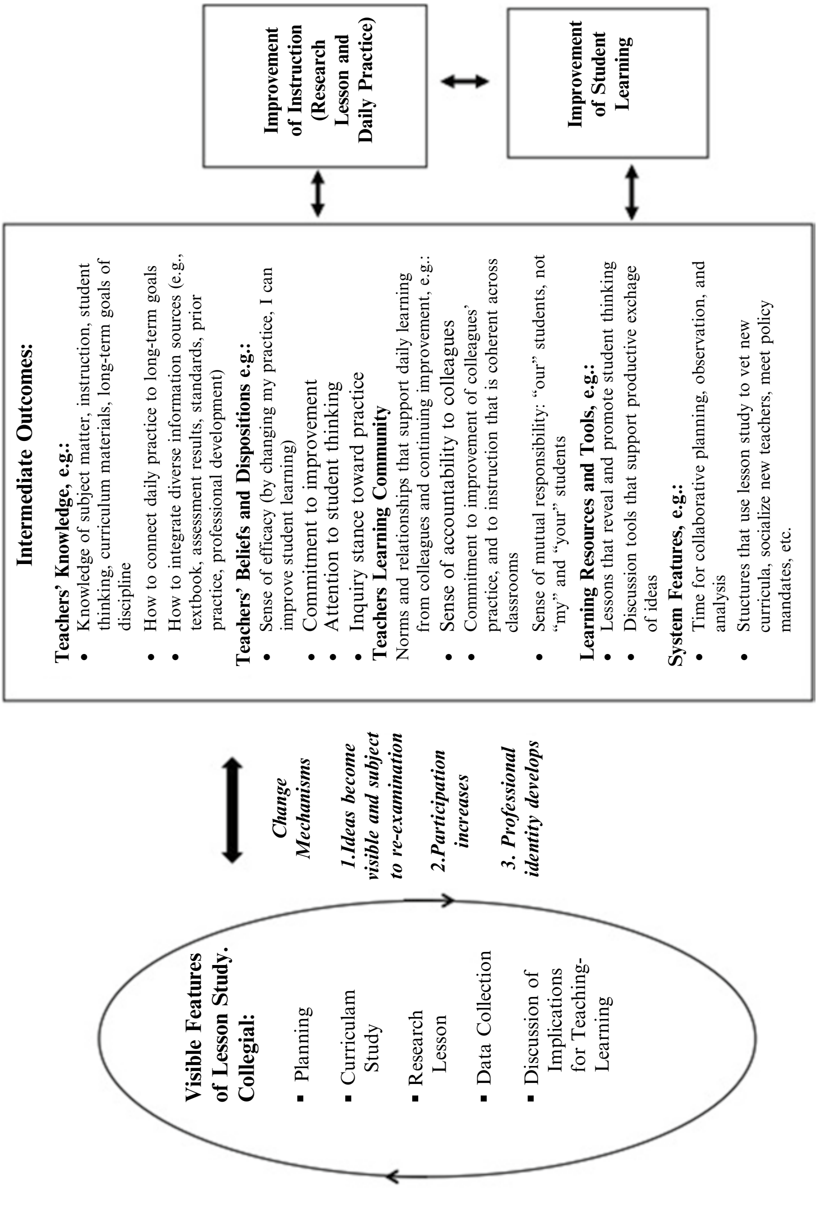

**Fig. 1** Lesson study cycle

A fourth element of lesson study is *teacher-led inquiry*. Lesson study asks teachers to pose and investigate research questions about practice and to answer them through the study of live practice; this element is shared with approaches to professional learning such as action research, self-study, and inquiry (e.g., Noffke & Somekh, 2009). A fifth element of lesson study is *collaboration with colleagues* (e.g., McLaughlin & Talbert, 2006), which is intended to build ongoing instructional collaboration and to reshape school routines to better focus on improvement of classroom practice (Sherer & Spillane, 2011).

In summary, lesson study, which originated in Japan and has been practiced there for over a century, combines familiar elements including study of content and curriculum, study of student thinking and work, and observation of live instruction, with collaborative planning and analysis of instruction. These elements are brought together in cycles of teacher-led inquiry.

## Method

### *The Study Design and Conditions*

The tension between teacher "ownership" of an innovation and its faithful implementation is a quintessential dilemma of educational reform. Our intervention approached this dilemma by joining a teacher-led form of professional learning–lesson study–to research-based mathematics resources. We designed the mathematics resource kits because, as noted above, some US teacher's manuals do not provide sufficient information on mathematical content and student thinking to support the *kyouzai kenkyuu* during the first phase of lesson study (Lewis et al., 2011).

Using an electronic mailing list (lsnetwork@mailman.depaul.edu) and personal contacts, we recruited groups of US educators to engage in locally led lesson study, using resource kits centrally designed and distributed by our group. Since local educators took responsibility for recruiting lesson study teams and organizing and managing their own learning with the resource kits, this model departs substantially from centrally planned and "delivered" professional development in which educators are expected to faithfully implement a centrally designed set of instructional changes (Lewis, Perry, & Murata, 2006). The design of the experimental condition was intended to support local flexibility for users, by allowing each lesson study group to collaboratively make its own locally appropriate decisions about participants, use of the fractions resources we provided, and time allocation for lesson study and for specific components of the work.

More than 100 groups requested an opportunity to participate in the study. Four criteria were considered in selecting the sample of 39 sites: permission from local authorizing agencies and administrators; willingness to be randomly assigned to a study condition; site demographic characteristics (we sought diversity in region of the USA, urbanicity and student socioeconomic status); and the ability to participate within our study time frame. In the interest of supporting naturally occurring collaborative groups, we did not specify group membership, except to require that at least one group member be a classroom teacher within grades 2–5. Since some lesson study groups find it beneficial to collaborate across grades, groups could include educators at other levels or from non-classroom positions (e.g., mathematics coach). Educators who responded to our call for participation recruited local groups of 4–9 educators.

Triads of demographically matched sites (matched where possible on district and SES of students served) were created and one site from each triad was assigned by random draw to each of the three study conditions. Random assignment was not performed until after groups had completed teacher and student pre-assessments. *Condition 1* (C1) is the primary experimental treatment and consists of lesson study supported by a mathematical resource kit focused on fractions (described further in Lewis & Perry, 2014, and summarized in Fig. 2). Because we worried that the mathematical resource kit might be experienced by teachers as prescriptive, thereby undermining the sense of inquiry that should be integral to lesson study, we designed *Condition 2* (C2) as a control treatment in which teachers conducted lesson study *without* the fractions resource kit, on a topic of the group's *own* choosing other than fractions. We asked C2 groups to avoid fractions both for practical reasons (to avoid cross-condition "contamination," since many districts had groups in more than one condition) and for ethical reasons (we did not want to put local mathematics educators who had recruited more than one group in the position of withholding resources from some groups). C2 groups did receive sections 4 and 5 of the resource kit (see Fig. 2), the generic materials to support lesson study. These included tools and protocols to help groups set norms, anticipate student thinking, and plan, observe, and discuss a research lesson. Groups in *Condition 3* (C3) received no materials from our group, but participated in all study procedures (such as pre- and post-assessment) and received the study stipend upon documentation of expenses for professional development. C3 thus served as a control for selection factors (such as willingness to participate in lesson study) and for study procedures (such as study assessments and stipend). Groups in all three conditions were offered a $4,000 stipend upon documentation of expenses related to professional learning (e.g., substitutes, stipends for after-school work, course fees). The average length of participation (calendar days from student pretest to posttest) was roughly comparable across conditions: 91 days for Condition 1 groups; 80 days for Condition 2 groups; and 84 days for Condition 3 groups. No groups dropped out of the study, and only one teacher failed to complete the study.

Figure 2 summarizes the materials in the resource kit. Overall, the resource kit was designed to maintain the qualities of lesson study that are appealing to educators such as active investigation of a problem of practice, study of student thinking, and application of newly learned ideas in the classroom, while also providing ready access to mathematical resources. The basic flow of the toolkit, summarized in Fig. 2, begins with teachers solving mathematics tasks individually, sharing solutions, predicting how students might solve the tasks, and then examining actual student responses. Teachers then examine curriculum materials and research on

**Introduction**

**Section 1: Mathematics Tasks to Solve and Discuss**

Groups do and discuss problems such as Problem 2: "Find two fractions between $\frac{1}{2}$ and 1 and write them here," then study associated sample student work.

**Section 2: Curriculum Inquiry: Different Models of Fractions**

Groups examine different fraction models, the Japanese textbook and fractions curriculum trajectory, and classroom video of fractions instruction. Groups also solve a hands-on fractions task mirroring one shown on the video ("Mystery Strip").

**Section 3: Choosing a Focus for Your Lesson Study Work**

Groups choose either Path A or Path B. Path A centers around an introduction to fractions using the linear measurement context. Path A groups study materials based on the Japanese curriculum introduction to fractions (e.g., lesson plans; Japanese elementary Course of Study; Teaching Manuals). Path B groups study and read about another aspect of fractions, such as: 1) understanding that fractions are accumulations of unit fractions; or 2) understanding fractions on the number line, etc.

**Section 4: Planning, Conducting, and Discussing the Research Lesson**

Groups follow lesson study protocols, guidelines, and suggestions on reflection included in this section.

**Section 5: Lesson Study Refresher: Overview and Suggestions for Getting Started**

Groups new to lesson study may refer to this section for background information on how to conduct lesson study (e.g., setting norms in the group or choosing a research topic).

**Fig. 2** Fractions resource kit: summary and examples of contents fractions, with a particular focus on the linear measurement representation and how it may help students understand fractions as rational numbers. Many of the resources in this section are drawn from the Japanese curriculum, including a textbook, teacher's edition, lesson plan, and video of a Japanese educator introducing fractions to a US class, using the linear measurement representation of fractions. Finally, lesson study groups plan, conduct, observe, and discuss at least one research lesson on fractions, followed by reflection on what they learned during the lesson study cycle.

## *Measurement of Teachers' Beliefs and Teacher Learning Community*

The central portion of Fig. 1 posits that change in teachers' beliefs and dispositions is one route by which lesson study produces changes in instruction. Changes in beliefs and dispositions may directly influence instruction; for example, change in beliefs about the value of student struggle may lead a teacher to give students more time to struggle with a challenging problem. Changes may also operate indirectly; for example, a strengthened sense of efficacy may allow a teacher to engage in the repeated cycles of experimentation needed to successfully implement a new teaching strategy that initially proves difficult.

A large body of research documents the ways that teachers' beliefs, dispositions, and identity both *influence* and *are influenced* by professional learning experiences (Clarke & Hollingsworth, 2002; Goldsmith, Doerr, & Lewis, 2014; Zech, Gause-Vega, Bray, Secules, & Goldman, 2000). (In the current work, we do not try to distinguish among beliefs, dispositions, and identity, but refer to them all using the term "beliefs.") Teachers' expectations for student achievement have been the focus of a number of studies that have shown, for example, that opportunities to closely observe students can increase teachers' expectations (Borko, Davinroy, Bliem, & Cumbo, 2000; Chazan, Ben-Chaim, & Gormas, 1998; Kazemi & Franke, 2004; Lin, 2001; Puchner & Taylor, 2006), as can collegial learning that focuses on student thinking (Lin, 2002; Tobin & Espinet, 1990). Since working with colleagues to closely observe students is a core feature of lesson study, we included survey items designed to measure teachers' expectations for student achievement.

Another type of belief that may be important to instructional improvement is interest in student thinking and in eliciting it during instruction because student thinking can provide crucial instructional feedback to teachers (e.g., Grandau, 2005; Kazemi & Franke, 2004; Remillard & Bryans, 2004; Seymour & Lehrer, 2006; Steinberg, Empson, & Carpenter, 2004). We included four survey items designed to measure teachers' sense of efficacy in eliciting and using students' mathematical thinking.

Inquiry stance toward practice and identity as a learner and teacher of mathematics is a third focus of the survey items we selected for inclusion. Prior research has shown that professional learning can impact teachers' inquiry stance, leading to increased use of analysis and knowledge-based reasoning and allowing a shift away from an evaluative stance (Sherin & Han, 2004; Sherin & van Es, 2009; Ticha & Hospesova, 2006; van Es & Sherin, 2008). Items in this area focused on teachers' interest in learning about mathematics and its teaching–learning. Prior research shows that teachers' identities as learners and teachers of mathematics can shape, for example, their learning from mathematics curriculum materials (Remillard & Bryans, 2004; Spillane, 2000). Because the resource kits we provided to the teachers in this study included curriculum and research materials, four survey items tapped teachers' perception that research and curricular materials (including those from other countries) are useful to teachers.

Another hypothesized route of lesson study influence on instruction (see Fig. 1) is through changes in the teacher learning community. Prior research indicates that professional development can, for example, engender collegial encouragement and support that enables teachers to try new types of teaching (Britt, Irwin, & Ritchie, 2001; Chazan et al., 1998; Manouchehri, 2001) and to see colleagues as a source of useful feedback and knowledge (Fisler & Firestone, 2006; Taylor, Anderson, Meyer, Wagner, & West, 2005; Thijs & van den Berg, 2002; Zech et al., 2000). The duration of the current study was somewhat brief to expect development of the teacher learning community, as it included just one cycle of lesson study. However, we included in the survey a number of items focused on workplace collaboration, including a number of items from established measures of school-site professional community (e.g., Michigan State University, 2003; CRC, 1994) and some new items designed to tap the perceived efficacy of working with colleagues to improve mathematics teaching ("Collegial Learning Effectiveness").

The pre- and post-teacher survey included the items shown in Appendix, which were interspersed for administration along with some additional items such as self-ratings of fractions knowledge (not reported here). Factor analysis and item content review were used to construct the scales, which are shown in Appendix along with scale alphas and item sources.

### *Measurement of Teachers' and Students' Fractions Knowledge*

Since some of our analyses examine the impact of teachers' beliefs and teacher learning community on the development of fractions knowledge, we briefly describe the assessments of teachers' and students' fraction knowledge. The measure of teachers' fraction knowledge was a 33-item assessment, with 21 of the items drawn from Learning Mathematics for Teaching (2007) and the remaining items drawn from other sources including Diagnostic Teacher Assessments in Math and Science (Center for Research in Mathematics and Science Teacher Development, 2005); New Zealand Maths (Ward & Thomas, 2009); and mathematics education research and curriculum materials (Beckmann, 2005; Newton, 2008; Norton & McCloskey, 2008; Post, Harel, Behr, & Lesh, 1988; Schifter, 1998; Zhou, Peverly, & Xin, 2006).

Because the study included students in grades 2–5, three overlapping assessments of students' fraction knowledge were constructed, including between 17 items (for grade 2–3 students) and 41 items (for grade 5 students) drawn from sources including published mathematics education research (Hackenberg, Norton, Wilkins, & Steffe, 2009; Saxe, 2007; Van de Walle, 2007); the National Assessment of Educational Progress (Institute of Education Sciences/National Center for Education Statistics (IES/NCES), 2007); Japanese teachers' manuals and student texts (Hironaka & Sugiyama, 2006); and the California Standards Test (California Department of Education, 2005).

In addition to the survey and assessment data, we collected written reflections at the end of the lesson study cycle from teachers in both lesson study conditions in response to the following prompt:

> *Describe in some detail two or three things you learned from this lesson study cycle that you want to remember, and that you think will affect your future practice*. These might be things about fractions or mathematics, about teaching, about student learning, or about working with colleagues. (If you don't feel you learned anything from this cycle of lesson study, please note that and identify changes that might have made the lesson study work more productive for you.)

## *Data Collection*

Teacher and student pre-assessments were mailed out to the sites, along with guidelines for administration. Once the completed assessments had been mailed back and received in our office, the site was randomly assigned to a condition and the appropriate study materials (for example, the resource kit) were mailed out to the site. Post-assessments were mailed out at the end of the study period. Participants in Conditions 1 and 2 (the lesson study conditions) were also asked to video record their lesson study meetings and research lessons, to collect materials from the lesson study cycle (such as student work and lesson plans), and to complete reflection forms at the end of each meeting and at the end of the lesson study cycle. Sites periodically mailed these data to our office. Due to budgetary constraints, we did not observe or attempt to measure changes in teachers' regular classroom instruction.

The 39 groups of educators included groups in 11 US states and the District of Columbia and in 27 school districts, totaling 213 teachers across the three study conditions. Table 1 provides demographic information on the teachers by study condition. The treatment and control conditions are generally comparable in teachers' years of experience and grade-level assignment. However, teachers in the treatment group were more likely to have a math degree or credential than control teachers ($\chi^2(2, N=213)=10.39, p=.006$) and also had slightly more lesson study experience ($t(122)=2.756, p=.007$), although the means for lesson study experience of all three conditions were in the range of 1–2 years. To control for baseline differences, these two teacher characteristics were included as covariates in subsequent analyses.

**Table 1** Demographic data at study baseline

| Indicator (dichotomous variable-D; continuous variable-C) | Percentage if dichotomous; mean (SD) if continuous | | | | Tests of difference between condition 1 and other control groups combined | | |
|---|---|---|---|---|---|---|---|
| | All groups ($N=213$) | Cond 1 ($N=73$) | Cond 2 ($N=73$) | Cond 3 ($N=67$) | $X^2/t$ | *df* | *p* |
| Elementary grade teacher (D) | 87 % | 86 % | 92 % | 84 % | $X^2=2.23$ | 211 | .329 |
| Less than 5 years experience (D) | 28 % | 23 % | 25 % | 37 % | $X^2=4.07$ | 211 | .130 |
| More than 15 years experience (D) | 25 % | 27 % | 30 % | 18 % | $X^2=3.01$ | 211 | .223 |
| Math degree/ credential (D) | 11 % | 21 % | 4 % | 9 % | $X^2=10.39$ | 211 | .006 |
| Lesson study experience (C, scale 1–5) | 2.27 (1.32) | 2.63 (1.48) | 2.10 (1.29) | 2.06 (1.09) | $t=2.76$ | 122 | .007 |

Although we suggested a time allocation of about 12–14 group meetings (including at least one classroom research lesson) for completion of the study requirements, groups organized their own meeting logistics, determining the total time, number of meetings, and meeting length. As a result, group participation time varied. Excluding time for assessments, estimated participation time for Condition 1 groups ranged from 7 to 42 h and time for Condition 2 groups ranged from 1.5 to 29 h. Meeting time was calculated from video records and self-reported meeting schedules to the extent they were available. Video records may err on the side of underestimation, since groups sometimes started the video camera late or let it run out before the meeting ended. Because teachers in Condition 3 engaged in various professional development activities (some individually, some in groups) a comparable participation figure is difficult to calculate. For example, teachers in one Condition 3 group jointly attended a regional mathematics conference, while other groups requested stipend funds to support future lesson study efforts. Variability in time devoted to lesson study (within the two lesson study conditions) is probably due to a range of factors. For example, some groups asked members to review materials as "homework," so that some of their time did not get picked up in the video record. Likewise, time spent planning the lesson or talking with group members outside of formal group meetings did not get captured. Hence, the time estimates should be considered imprecise. One factor we could identify that impacted participation time is that groups that decided to teach the research lesson more than once tended to have longer participation times.

## *Data Analysis*

HLM analyses were conducted to assess the impact of the experimental condition on changes in teachers' beliefs and teacher learning community. For teacher outcomes, we used a two-level HLM model with teachers at Level 1 ($n=213$) and groups at Level 2 ($n=39$) to account for the nesting of teachers within lesson study groups. We chose three Level 1 covariates on the basis of baseline data and prior similar research: pretest value on the scale, lesson study experience, and possession of a mathematics degree or credential (Akiba, Chiu, Zhuang, & Mueller, 2008; Birman et al., 2009; Bloom, Richburg-Hayes, & Black, 2007; Desimone, Smith, & Ueno, 2006; Hill, 2010; Smith & Desimone, 2003). For each outcome measure, the Level 1 pretest value, the dummy indicator for possession of a math degree/credential, and lesson study experience (continuous variable) were included as grand-mean centered variables in the model (Raudenbush & Bryk, 2002). At Level 2, we included as an uncentered variable the group assignment to Condition 1 (lesson study with resource kit), assigned a value of 1, and a value of 0 otherwise. Our primary interest in this analysis was the estimate of the treatment effect on each of the six measures of belief and teacher learning community, captured by the Level 2 parameter $\gamma_{01}$ in the fully conditional model shown below.

*Level-1 Model*

$$\begin{aligned} Y_{ij} = \beta_{0j} + \beta_{1j}\left(\text{pretest\_value}\right)_{ij} + \beta_{2j}\left(\text{math\_degree / credential}\right)_{ij} \\ + \beta_{3j}\left(\text{lesson\_study\_experience}\right)_{ij} + r_{ij}. \end{aligned}$$

*Level-2 Model*

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}\left(\text{toolkit\_assignment}\right)_{j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \end{aligned} .$$

In addition to the HLM analyses to look at the impact of experimental assignment on the six outcome measures (beliefs and teacher learning community), we conducted additional HLM analyses designed to explore the impact of the six belief and teacher community measures on teachers' and students' development of fractions knowledge during the study period. Specifically, we investigated whether the six measures of teachers' beliefs and learning community predicted changes in teachers' and students' fractions knowledge.

## Results

Table 2 shows the pre- and post-intervention scores and change scores for teachers' beliefs and teacher learning community for all three study conditions, as well as *t*-tests for the comparison of change rates between the experimental treatment (lesson study with mathematical resource kit) and the combined control conditions (lesson study only and locally chosen professional development). HLM analyses are shown in Table 3; they indicate a positive and statistically significant impact of the experimental treatment on two of the six scales: Collegial Learning Effectiveness, and Expectations for Student Achievement. In addition, the intervention shows a marginally significant impact on the Using and Promoting Student Thinking scale.

To avoid inflating the experiment-wise significance level, we limit significance testing to comparison of Condition 1, which is the full experimental treatment, with the remaining conditions. Tables 4 and 5 show the results of HLM analyses that examine changes in the teacher belief and teacher learning community measures as mediators of teachers' and students' change in fractions knowledge. Table 4 indicates that increases in collegial learning effectiveness and in professional community both significantly predict teachers' gain in fractions knowledge during the study period for the overall study sample. Likewise, Table 5 indicates that increase in teachers' collegial learning effectiveness significantly predicts students' gain in fractions knowledge during the study period.

The end-of-cycle written reflections provide insights into the kinds of experiences that increased teachers' beliefs in the effectiveness of learning with colleagues:

> This has made me think of how essential it is to observe other teachers and take as many ideas as possible to integrate in my classroom.
>
> I think this was my 7th or 8th cycle of working with lesson study and every time I am amazed at the amount of growth and learning that happens professionally for me.... The biggest impact for me is having more ears around the room listening to the students' conversations and what they are actually thinking. For example, during one of the lessons, a pair of students had recorded the correct fraction and written it the correct way, however, I overheard one partner say to the other, "1/2 means we have 1 m and 2 more." During a typical lesson and without "extra" ears around the room, the classroom teacher would have thought that pair of students knew the answer and the misconception would not have been noted.
>
> I found it so helpful to come together as a team, look closely at work that we had recently observed in action, and not all agree at what the student demonstrated. This made it clear to me that my "research" can be flawed if I am not listening and watching closely as my students talk and solve problems.
>
> I feel the collaboration piece is one of the greatest benefits for each of us. As I look back at each of the reflection notes, it is amazing how many things were discussed and how many different perspectives came out as we discussed any research or topics as part of the discussion.... The collaboration piece is also important *during* the lesson.... With more eyes there is more information, which we have found helps us create better lessons with more student learning. Even after the second lesson, our post lesson discussion has us thinking about what we still could improve and where to go from here. I definitely feel that the lack of collaboration is a weakness in our American schools. (italics not in original)

Likewise, the reflections highlight experiences that increased teachers' expectations for student achievement:

> As I watched the lesson unfold I saw how, with good intentions, we teachers stop the thinking of our students by providing too much scaffolding.... I saw students working

**Table 2** Teachers' beliefs and teacher learning community, by time and condition

| | C1-LS with resource kit ($N=73$)[a] | | | C2-lesson study only ($N=73$) | | | C3-control ($N=67$) | | | *t*-Test (211 *df*) for change in C1 vs. C2 and C3 |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Change | Pre | Post | Change | Pre | Post | Change | |
| *Collegial learning effectiveness* | | | | | | | | | | |
| Mean | 3.95 | 4.11 | **0.17**[b] | 3.96 | 3.94 | −0.02 | 3.83 | 3.84 | 0.00 | 2.265[c]* |
| *SD* | *0.65* | *0.58* | *0.51* | *0.56* | *0.66* | *0.54* | *0.67* | *0.61* | *0.55* | |
| *Expectations for student learning* | | | | | | | | | | |
| Mean | 4.26 | 4.27 | 0.01 | 4.33 | 3.97 | **−0.36** | 4.31 | 4.09 | **−0.22** | 3.927*** |
| *SD* | *0.48* | *0.44* | *0.53* | *0.48* | *0.62* | *0.57* | *0.52* | *0.55* | *0.51* | |
| *Interest in mathematics and inquiry stance* | | | | | | | | | | |
| Mean | 4.19 | 4.2 | 0.01 | 4.23 | 4.02 | **−0.2** | 4.13 | 4.12 | −0.01 | 1.928$^{+}$ |
| *SD* | *0.5* | *0.5* | *0.38* | *0.45* | *0.66* | *0.54* | *0.49* | *0.57* | *0.42* | |
| *Professional community* | | | | | | | | | | |
| Mean | 3.71 | 3.54 | **−0.18** | 3.62 | 3.56 | −0.06 | 3.58 | 3.4 | **−0.18** | −0.689 |
| *SD* | *0.71* | *0.84* | *0.62* | *0.71* | *0.75* | *0.55* | *0.74* | *0.75* | *0.53* | |
| *Research relevance for practice* | | | | | | | | | | |
| Mean | 4.05 | 4.18 | **0.13** | 4.2 | 4.01 | **−0.18** | 4.12 | 4.15 | 0.03 | 2.382* |
| *SD* | *0.57* | *0.52* | *0.57* | *0.58* | *0.73* | *0.65* | *0.63* | *0.61* | *0.6* | |
| *Using and promoting st. thinking* | | | | | | | | | | |
| Mean | 3.69 | 3.9 | **0.21** | 3.77 | 3.72 | −0.05 | 3.87 | 3.84 | −0.03 | 3.138** |
| *SD* | *0.53* | *0.5* | *0.51* | *0.49* | *0.6* | *0.6* | *0.49* | *0.5* | *0.54* | |

[a] *N* for each study condition represents the number of educators in each condition and is consistent within condition across the 6 belief measures

[b] Bolded means in the "change" column for each of the study assignments indicate a statistically significant difference between pretest and posttest as follows. C1—Collegial learning effectiveness ($t(72)=2.751$, $p<.01$); Professional community ($t(72)=2.414$, $p<.05$); Research relevance for practice ($t(72)=1.945$, $p<.10$); Using and promoting student thinking ($t(72)=3.499$, $p<.01$). C2—Expectations for student achievement ($t(72)=5.421$, $p<.001$); Interest in mathematics and inquiry stance ($t(72)=3.253$, $p<.01$); Research relevance for practice ($t(72)=2.381$, $p<.05$). C3—Expectations for student achievement ($t(66)=3.461$, $p<.01$); Professional community ($t(66)=2.802$, $p<.01$)

[c] $^{+}p<.10$, $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.0$

**Table 3** HLM results: impact of lesson study with mathematical resource kits on teachers' beliefs and learning community

| | Collegial learning effectiveness | Expectations for student achievement | Interest in mathematics and inquiry stance | Professional community | Research relevance for practice | Using and promoting student thinking |
|---|---|---|---|---|---|---|
| Fixed effects | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient |
| Intercept | 3.907 (.039)*** | 4.027 (.046)*** | 4.086 (.051)*** | 3.502 (.054)*** | 4.084 (.057)*** | 3.773 (.044)*** |
| *Teacher predictors* | | | | | | |
| Pretest value | .604 (.058)*** | .528 (.089)*** | .750 (.067)*** | .737 (.037)*** | .496 (.072)*** | .435 (.062)*** |
| Math degree or credential | .085 (.098) | .111 (.114) | .168 (.069)* | −.082 (.141) | .133 (.092) | .202 (.077)* |
| Lesson study experience | .018 (.026) | .011 (.023) | .017 (.018) | .050 (.033) | −.023 (.031) | .019 (.023) |
| *Group predictor*—assignment to toolkit | .173 (.075)* | .251 (.069)*** | .089 (.063) | −.066 (.131) | .082 (.086) | .146 (.075)$^{+}$ |
| Random effects | Variance component | Variance component | Variance component | Variance component | Variance component | Variance component |
| *Unconditional model* | | | | | | |
| Variance between groups | 0.033 | 0.022 | 0.029 | 0.187 | 0.005 | 0.033 |
| Variance within groups | 0.359 | 0.284 | 0.311 | 0.434 | 0.390 | 0.257 |
| *Full model* | | | | | | |
| Variance between groups | 0.000 | 0.006 | 0.018 | 0.073 | 0.016 | 0.013 |
| Variance within groups | 0.232 | 0.229 | 0.180 | 0.235 | 0.310 | 0.219 |
| *Effect size—resource kit* | **0.03** | **0.07** | **0.02** | −0.01 | **0.02** | **0.03** |

L1 sample = 213 teachers; L2 sample = 39 groups

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{+}p<.10$

**Table 4** HLM results: changes in teachers' beliefs and professional community as predictors of teachers' knowledge gain

| Fixed effects | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) |
|---|---|---|---|---|---|---|---|---|
| Intercept | −.062 (.056) | −.059 (.055) | −.049 (.055) | −.049 (.057) | −.052 (.055) | −.063 (.055) | −.065 (.055) | −.058 (.055) |
| *Teacher predictors* | | | | | | | | |
| Pretest value | .802 (.046)*** | .806 (.046)*** | .804 (.045)*** | .802 (.047)*** | .802 (.045)*** | .800 (.046)*** | .791 (.043)*** | .802 (.046)*** |
| Math degree or credential | .084 (.101) | .123 (.110) | .102 (.099) | .079 (.100) | .067 (.097) | .084 (.101) | .104 (.099) | .076 (.100) |
| Lesson study experience | −.032 (.030) | −.025 (.029) | −.029 (.028) | −.032 (.029) | −.032 (.030) | −.033 (.030) | −.039 (.030) | −.032 (.029) |
| Collegial learning effectiveness (change score) | – | – | .181 (.083)* | – | – | – | – | – |
| Expectations for student achievement (change score) | – | – | – | .124 (.084) | – | – | – | – |
| Using and promoting student thinking (change score) | – | – | – | – | .115 (.076) | – | – | – |
| Research relevance for practice (change score) | – | – | – | – | – | −.210 (.063) | – | – |
| Professional community (change score) | – | – | – | – | – | – | .186 (.061)** | – |
| Interest in math/inquiry stance (change score) | – | – | – | – | – | – | – | .093 (.116) |
| *Group predictors* | | | | | | | | |
| Assignment to toolkit | .178 (.077)* | .173 (.075)* | .142 (.078)$^{+}$ | .141 (.082)$^{+}$ | .151 (.078)$^{+}$ | .183 (.077)* | .190 (.077)* | .168 (.076)* |
| Random effects | | | | | | | | |

(continued)

**Table 4** (continued)

| Fixed effects | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) | Fractions knowledge (Z score) |
|---|---|---|---|---|---|---|---|---|
| Unconditional model | | | | | | | | |
| Intercept (variance between groups) | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 |
| Level 1 (variance within groups) | 0.970 | 0.970 | 0.970 | 0.970 | 0.970 | 0.970 | 0.970 | 0.97 |
| Full model | | | | | | | | |
| Intercept (variance between groups) | 0.004 | 0.002 | 0.002 | 0.007 | 0.003 | 0.004 | 0.005 | 0.004 |
| Level 1 (variance within groups) | 0.359 | 0.361 | 0.353 | 0.354 | 0.358 | 0.361 | 0.349 | 0.360 |
| Effect size of assignment to resource kit[a] | **0.18** | **0.17** | **0.14** | **0.14** | **0.15** | **0.18** | **0.19** | **0.17** |
| Effect size of assignment to resource kit[b] | **0.02** | **0.02** | **0.02** | **0.02** | **0.02** | **0.02** | **0.02** | **0.02** |

L1 sample = 213 teachers; L2 sample = 39 groups

[a] Effect sizes are standardized coefficients

[b] Computed effect sizes using formula from Song & Herman, 2010

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .10$

**Table 5** HLM results: changes in teachers' beliefs and professional community as predictors of students' knowledge gain

| | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) |
|---|---|---|---|---|---|---|---|
| Fixed effects | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient |
| Intercept | −.186 (.050)*** | −.160 (.051)*** | −.185 (.049)*** | −.182 (.048)** | −.188 (.051)** | −.182 (.049)** | −.186 (.051)** |
| *Student predictors* | | | | | | | |
| Pretest score | .682 (.029)*** | .681 (.028)*** | .682 (.029)*** | .683 (.027)*** | .682 (.029)*** | .683 (.029)*** | .682 (.029)*** |
| Grade 23 test | −.056 (.099) | −.102 (.091) | −.058 (.095) | −.058 (.095) | −.037 (.103) | −.1058 (.096) | −.049 (.099) |
| Grade 5 test | −.085 (.127) | −.148 (.136) | −.089 (.139) | −.105 (.121) | −.064 (.135) | −.091 (.119) | −.075 (.131) |
| *Teacher predictors* | | | | | | | |
| Math degree or credential | −.193 (.069)** | −.133 (.081)** | −.197 (.062)** | −.287 (.082)** | −.186 (.063)** | −.195 (.060)** | −.196 (.069)** |
| Lesson study experience | .016 (.033) | .023 (.031) | .017 (.033) | .015 (.032) | .012 (.032) | .021 (.033) | .013 (.032) |
| Collegial learning effectiveness (change score) | – | .219 (.073)** | – | – | – | – | – |
| Expectations for student achievement (change score) | – | – | .011 (.072) | – | – | – | – |
| Using and promoting student thinking (change score) | – | – | – | .101 (.076) | – | – | – |
| Research relevance for practice (change score) | – | – | – | – | −.066 (.064) | – | – |
| Professional community (change score) | – | – | – | – | – | −.070 (.059) | – |
| Interest in math/inquiry stance (change score) | – | – | – | – | – | – | −.061 (.066) |

(continued)

**Table 5** (continued)

| | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) | Fractions knowledge (*Z* score) |
|---|---|---|---|---|---|---|---|
| Fixed effects | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient |
| *Group predictors* | | | | | | | |
| Assignment to toolkit | .496 (.136)** | .405 (.125)** | .494 (.133)*** | .477 (.130)** | .505 (.137)** | .489 (.136)** | .500 (.136)** |
| Random effects | Variance component | Variance component | Variance component | Variance component | Variance component | Variance component | Variance component |
| Unconditional model | | | | | | | |
| Intercept1/intercept2 (variance between groups) | 0.250 | 0.250 | 0.250 | 0.250 | 0.250 | 0.250 | 0.250 |
| Intercept (variance within groups between classes) | 0.124 | 0.124 | 0.124 | 0.124 | 0.124 | 0.124 | 0.124 |
| Level 1 (variance within classes) | 0.603 | 0.603 | 0.603 | 0.603 | 0.603 | 0.603 | 0.603 |
| Full model | | | | | | | |
| Intercept1/intercept2 (variance between groups) | 0.044 | 0.047 | 0.044 | 0.035 | 0.048 | 0.040 | 0.049 |
| Intercept (variance within groups between classes) | 0.050 | 0.036 | 0.050 | 0.053 | 0.046 | 0.051 | 0.047 |
| Level 1 (variance within classes) | 0.290 | 0.290 | 0.290 | 0.290 | 0.290 | 0.290 | 0.290 |
| Effect size of assignment to resource kit | **0.5** | **0.41** | **0.49** | **0.48** | **0.51** | **0.49** | **0.41** |

> themselves from an incorrect answer to recognizing the answer was wrong, puzzling over how to correct it only to have a teacher ask “yes–no” questions that stopped their problem solving and led them to the correct answer. I recognize this trait in myself and have committed myself to allowing the students time to struggle…
>
> My students discovered on their own that the more you divide the whole the smaller the fractional parts.… Because of this discovery many students began to make awesome connections.… Once when comparing 3/4 and 5/6 Daniella claimed that she can compare the fractions just by looking at them. Other students thought it would be too difficult because the size of the parts and number of parts were different. Daniella used her understanding of unit fractions to compare the numbers. She said that 3/4 is 1/4 away from equaling 1 whole but 5/6 is only 1/6 away from equaling 1 whole, 1/4 is larger than 1/6, so 3/4 is less than 5/6. This came completely unprompted and it led to a student explaining and demonstrating using fraction strips and the other students agreeing and taking part in a cool discovery. This never happened before because I never put much time as a 5th grade teacher into my student understanding of unit fractions.
>
> I love that one teacher did a 360 [complete turnaround] from her initial response to the math lesson, “My students cannot do this,” to “I would love to see my students do this.” That raising of the bar, while at the same time knowing the students well enough to plan for success, proved to be the best surprise of all.

## Discussion

The HLM analyses indicate that participation in lesson study with mathematical resources significantly increased two of the six outcome measures related to teachers’ beliefs and teacher learning community: Expectations for Student Achievement and Collegial Learning Effectiveness. The intervention also had a marginally significant effect on a third outcome: Using and Promoting Student Thinking. End-of-cycle reflections illuminate the specific experiences that enabled these changes in beliefs, such as hearing other teachers’ perspectives and seeing students respond to a challenging mathematical task.

Although we generally combined the two control groups for analysis to avoid inflating the experiment-wise significance levels, examining the data for all three conditions provides insights into how lesson study with the specially designed mathematical resources (Condition 1) differed from more typical lesson study (Condition 2) and from locally chosen professional development other than lesson study (Condition 3). Going down the “change” columns in Table 2 for each of the three conditions suggests that the two control conditions may be more similar to each other than to the experimental condition (lesson study with the mathematical resource kit) in terms of impact on teachers’ beliefs and professional learning community. Why would this be? The mathematical resources provided to Condition 1 teachers may have catalyzed more opportunities to change beliefs than the resources Condition 2 teachers located on their own. For example, the mathematical resources included fractions chapters from a Japanese teacher’s edition, and previous research has shown major differences between USA and Japanese teacher’s editions, such as more presentation of varied student thinking in the Japanese vs. the US teacher’s edition (28 % vs. 1 % of statements) and more discussion of the rationale for tasks and instructional design (10 % vs. 0 %) (Lewis et al., 2011). So the resource kits may have catalyzed a more substantial collegial discussion than the materials (such as US teacher's editions) located by Condition 2 groups, making colleagues more valuable in making sense of the materials. A number of participants mentioned how helpful it was to see how colleagues solved the math tasks; it is likely that Condition 2 lesson study groups located, solved, and discussed fewer tasks than did teachers using the resource kit, since it included a number of mathematical tasks and specific prompts to solve them individually and then discuss.

Similarly, the resource kit's emphasis on linear measure models and unit fractions seem to have been useful in revealing students' mathematical potential. The linear measure model made it easy for students to compare 3/4 and 5/6 (as noted by the teacher quoted above) and to use this in classroom discussion, which in turn allowed the teacher to see students' potential to reason mathematically. After describing students' "awesome discovery" the teacher wrote: "This never happened before because I never put much time as a 5th grade teacher into… unit fractions." Condition 2 teachers, who sought out lesson materials on their own, may have had a harder time finding materials that supported such changes during the brief period of the study.

One interesting feature of the findings is the difference in results for the two scales related to learning from colleagues: Collegial Learning Effectiveness and Professional Community. These scales differ in two major ways. First, several of the items on the Professional Community scale focus on all colleagues at the school site, for example: "There is a lot of discussion among teachers at this school about how to teach." In contrast, the Collegial Learning Effectiveness scale refers to colleagues self-identified by the respondent, for example, "I have learned a great deal about mathematics teaching from colleagues." Since the lesson study we report was conducted by small groups of teachers, not by all teachers in a school, differences between the two measures would be expected if teachers' attitudes toward their lesson study colleagues do not necessarily extend to the broader set of all colleagues at their school site. A second difference between the scales is that the Professional Community scale focuses on the *frequency* of learning with colleagues, whereas the Collegial Learning Effectiveness Scale focuses on its *usefulness and impact* (for example, whether respondents believe they have learned about student thinking from colleagues). Finally, the Collegial Learning Effectiveness Scale is more heavily focused on mathematics (4 of 5 items) than the Professional Community Scale (2 of 6 items). Hence, Collegial Learning Effectiveness is better designed to pick up changes in usefulness of collegial learning among educators with whom the respondent collaborates, as opposed to frequency of collegial interaction within the school as a whole.

Another interesting aspect of Table 2 is that scores on many of the scales declined in the 3–4 months period between the baseline administration (usually in September) and the final administration (usually in January), especially in the two control groups. The September administration may have captured the most hopeful moment of the school year.

One limitation of this study is that several of the scales have a relatively small number of items, marginal scale reliability, and little or no prior evidence of predictive validity. Given the length of the fractions assessment, there was not sufficient time to administer a large number of survey items related to beliefs or teacher learning community. Many existing scales (with evidence of predictive validity) did not seem adequately aligned to the intervention at hand, lesson study by a small group of educators at a site (or across sites) on a particular topic within mathematics (fractions). More thorough investigation of the middle box of Fig. 2, the changes that occur during lesson study in teachers' beliefs and collegial relationships, is certainly warranted.

## Conclusions

As far as we are aware, this is the first randomized, controlled trial of a lesson study intervention, and we believe that it contributes in several ways to the current research base on professional learning. First, it documents that in a brief period of about 3 months, self-organized groups of educators scattered across the USA, supported by mathematical resource kits, were able to conduct lesson study that significantly increased not only their own and students' knowledge of fractions but also their expectations for student achievement and the reported efficacy of working with colleagues–beliefs that may have enormous implications for *future* efforts to improve. Prior qualitative research has provided evidence of changes in teachers' beliefs and professional relationships during lesson study (e.g., Lewis, Perry, & Hurd, 2009; Murata, 2003), and the current study confirms these changes in a much larger sample using a randomized trial. The findings suggest the fruitfulness of taking a much closer look at the middle box of Fig. 1 to document the changes in beliefs and collegial work that may allow lesson study to produce changes in both teachers' and students' learning and to support teachers' continued learning from practice over time.

Given the arguments made at the outset that improvement of instruction is likely to require teachers to engage in repeated cycles of trial and revision, it is essential to identify the beliefs and collegial learning structures that allow teachers to keep up this effortful work over time. Our findings indicate that the intervention significantly increased teachers' perceptions of the usefulness of collegial work and their expectations for student achievement and that these changes significantly predicted increases in teachers' and students' mathematical knowledge over the study period. This was true for the intervention group and for the study sample as a whole.

Finally, the results of this study should encourage us to think in new ways about scale-up of instructional improvement. The intervention was "low-touch," in that local, self-managed groups of educators worked independently at a distance from us, without any centralized supervision. These groups organized their learning in ways that made sense locally, rather than adhering to centrally prescribed rules designed to achieve implementation fidelity. In this way, the intervention supported educators' own agency and leadership, while also allowing them to build their mathematical knowledge.

These results suggest a promising solution to the conundrum of faithful implementation of high-quality materials versus teachers' "ownership" of professional learning. Through a lesson study process supported by mathematical resources, teachers can participate in a process that values their ideas and leadership, while at the same time increasing their expectations for student achievement and the effectiveness of their collegial work, as well as their own mathematical knowledge and that of their students.

## Appendix: Scales to Measure Teachers' Beliefs and Teacher Learning Community

Stem: "Please indicate how well each of the following statements describes your attitude" (Rated on a 5-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree.")).

*Expectations for student achievement* (7 items; Alpha = .63 on pretest; .64 on posttest)

No matter how hard I try, some students will not be able to learn aspects of mathematics (reverse coded) (McLaughlin & Talbert, 2001).
My expectations about how much students should learn are not as high as they used to be (reverse coded) (McLaughlin & Talbert, 2001).
Students who work hard and do well deserve more of my time than those who do not (reverse coded) (McLaughlin & Talbert, 2001).
The attitudes and habits students bring to my classes greatly reduce their chances for academic success (reverse coded) (McLaughlin & Talbert, 2001).
There is really very little I can do to ensure that most of my students achieve at a high level (reverse coded) (McLaughlin & Talbert, 2001).
Most of the students I teach are not capable of learning material I should be teaching them (reverse coded) (McLaughlin & Talbert, 2001).
By trying a different teaching method, I can significantly affect a student's achievement (CRC, 1994).

*Using and promoting student thinking*: (4 items; .63 at pretest and .68 at posttest)

I am able to figure out what students know about fractions (Project-developed).
I have some good strategies for making students' mathematical thinking visible (Project-developed).
I can help students "catch up" who come to me lacking in math skills (Adapted from CRC, 1994).
When students are confused about fractions, I am able to provide good examples and explanations (Project-developed).

*Interest in mathematics and inquiry stance* (*8 items*; Alpha = .74 on pretest; .84 on posttest)

I enjoy teaching mathematics (Horizon Research, 2000).
I like solving mathematics problems (Project-developed).
Student mathematical thinking is fascinating to me (Project-developed).
I think of myself as a researcher in the classroom (Project-developed).
I am always curious about student thinking (Adapted from MSU, 2003).
I actively look for opportunities to learn more mathematics (Project-developed).
I am interested in the mathematics taught at many grade levels (Project-developed).
I would like to learn more about fractions (Adapted from LMT, 2007).

*Research relevance for practice* (4 items; .64 at pretest and .66 at posttest)

Educational research often provides useful insights for teaching (Project-developed).

In general, curriculum materials from other countries are not useful (Project-developed).
Most research is not relevant to my needs as a teacher (Project-developed).
I find it interesting to read about a variety of educational programs and ideas (Project-developed).

*Collegial learning effectiveness* (5 items; .62 on pretest and .63 on posttest; based on items adapted from CRC, 1994 and Horizon Research Inc., 2000.)

I have learned a lot about student thinking by working with colleagues.
Working with colleagues on mathematical tasks is often unpleasant (reverse coded) (Project-developed).
I have good opportunities to learn about the mathematics taught at different grade levels (Adapted from CRC, 1994).
I have learned a great deal about mathematics teaching from colleagues.
I find it useful to solve mathematics problems with colleagues (Project-developed).

*Professional Community* (6 items; .80 at pretest and .82 at posttest)

My colleagues and I regularly share ideas and materials related to mathematics teaching.
Mathematics teachers in this school regularly observe each other teaching classes as part of sharing and improving instructional strategies.
I feel supported by other teachers to try out new ideas in teaching.
There is a lot of discussion among teachers at this school about how to teach (Adapted from CRC, 1994; MSU, 2003).
I plan and coordinate with other teachers (MSU, 2003).
I don't know how other teachers in this school teach (Adapted from CRC, 1994).

## References

Akiba, M., Chiu, Y.-F., Zhuang, Y.-L., & Mueller, H. E. (2008). Standards-based mathematics reforms and mathematics achievement of American Indian/Alaska native eighth graders. *Education Policy Analysis Archives, 16*(20), 1–31.
Beckmann, S. (2005). *Mathematics for elementary teachers*. Boston: Pearson.
Birman, B. F., Boyle, A., Le Floch, K., Elledge, A., Holtzman, D., Song, M., et al. (2009). *State and local implementation of the no child left behind act: Volume VIII–teacher quality under NCLB: Final report*. Washington, DC: U.S. Department of Education.
Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.
Borko, H., Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two third-grade teachers' experiences in a mathematics and literacy staff development project. *The Elementary School Journal, 100*(4), 273–306.
Britt, M. S., Irwin, K. C., & Ritchie, G. (2001). Professional conversations and professional growth. *Journal of Mathematics Teacher Education, 4*(1), 29–53.
California Department of Education. (2005). CST released test questions—standardized testing and reporting (STAR). From http://www.cde.ca.gov/ta/tg/sr/css05rtq.asp
Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 11*(3), 430–454.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Porthsmouth, NH: Hienemann.

Center for Research in Mathematics and Science Teacher Development. (2005). Diagnostic teacher assessment in mathematics and science, elementary mathematics, rational numbers assessment, version 1.2 and middle school mathematics, number and computation assessment, version 2. From http://louisville.edu/education/research/centers/crmstd/diag_math_assess_elem_teachers.html

Chazan, D., Ben-Chaim, D., & Gormas, J. (1998). Shared teaching assignments in the service of mathematics reform: Situated professional development. *Teaching and Teacher Education, 14*(7), 687–702.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967.

CRC (Center for Research on the Context of Teaching). (1994). *Survey of elementary mathematics education in California: Teacher questionnaire*. Stanford, CA: Stanford University School of Education.

Desimone, L. M., Smith, T. M., & Ueno, K. (2006). Are teachers who need sustained, content-focused professional development getting it? An administrator's dilemma. *Educational Administration Quarterly, 42*(2), 179–215.

Fernandez, C., & Yoshida, M. (2004). *Lesson study: A case of a Japanese approach to improving instruction through school-based teacher development*. Mahwah, NJ: Lawrence Erlbaum.

Fisler, J. L., & Firestone, W. A. (2006). Teacher learning in a school-university partnership: Exploring the role of social trust and teaching efficacy beliefs. *Teachers College Record, 108*(6), 1155–1185.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 304–313.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches.* (No. REL 2014-010). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2014). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education, 17*(1), 5–36.

Grandau, L. (2005). Learning from self-study: Gaining knowledge about how fourth graders move from relational description to algebraic generalization. *Harvard Educational Review, 75*(2), 202–221.

Hackenberg, A., Norton, A., Wilkins, J., & Steffe, L. (2009). *Testing hypotheses about students' operational development of fractions*. Paper presented at the NCTM Research Pre-Session.

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education, 41*(5), 513–545.

Hironaka, H., & Sugiyama, Y. (2006). *Mathematics for elementary school, grades 1-6*. Tokyo: Tokyo Shoseki. English language version available at http://www.globaledresources.com.

Horizon Research, Inc. (2000). 2000 National Survey of Science and Mathematics Education Mathematics Teacher Questionnaire. http://www.horizon-research.com/horizonresearchwp/wp-content/uploads/2013/04/math_teacher.pdf.

Institute of Education Sciences/National Center for Education Statistics (IES/NCES). (2007). NAEP questions tool. Retrieved from http://nces.ed.gov/nationsreportcard/itmrlsx/landing.aspx

Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education, 7*(3), 203–235.

Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

Learning Mathematics for Teaching. (2007). *Survey of teachers of mathematics, winter 2007, form LMT PR-07*. Ann Arbor, MI: University of Michigan, School of Education.

Lewis, C., & Hurd, J. (2011). *Lesson study step by step: How teacher learning communities improve instruction*. Portsmouth, NH: Heinneman.

Lewis, C., & Perry, R. (2014). Lesson study with mathematical resources: A sustainable model for locally led teacher professional learning. *Mathematics Teacher Education and Development, 16*(1), 22–42.

Lewis, C., & Perry, R. (under review). *Scale-up of instructional improvement through lesson study.* Unpublished manuscript.

Lewis, C., Perry, R., & Friedkin, S. (2011). Using Japanese curriculum materials to support lesson study outside Japan: Toward coherent curriculum. *Educational Studies in Japan: International Yearbook, 6*, 5–19.

Lewis, C., Perry, R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *Journal of Mathematics Teacher Education, 12*(4), 285–304.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3–14.

Lewis, C., & Tsuchida, I. (1997). Planned educational change in Japan: The case of elementary science instruction. *Journal of Educational Policy, 12*(5), 313–331.

Lewis, C., & Tsuchida, I. (1998). A lesson is like a swiftly flowing river: Research lessons and the improvement of Japanese education. *American Educator*, Winter, 14–17 & 50–52.

Lin, X. (2001). Reflective adaptation of a technology artifact: A case study of classroom change. *Cognition and Instruction, 19*(4), 395–440.

Lin, P. (2002). On enhancing teachers' knowledge by constructing cases in the classrooms. *Journal of Mathematics Teacher Education, 5*(4), 317–349.

Manouchehri, A. (2001). Collegial interaction and reflective practice. *Action in Teacher Education, 22*(4), 86–97.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. University of Chicago Press.

McLaughlin, M. W., & Talbert, J. E. (2006). *Building school-based teacher learning communities*. New York: Teachers College Press.

Michigan State University. (2003). *Teacher learning through professional development, 2003 SVMI teacher survey*. East Lansing, MI: Michigan State University.

Murata, A. (2003). *Teacher learning and lesson study: Developing efficacy through experiencing student learning.* Paper presented at the annual meeting of the School Science and Mathematics Association, Columbus, OH.

Newton, K. J. (2008). An extensive analysis of preservice elementary teachers' knowledge of fractions. *American Educational Research Journal, 45*(4), 1080–1110.

Noffke, S., & Somekh, B. (Eds.). (2009). *The SAGE handbook of educational action research*. London: Sage.

Norton, A. H., & McCloskey, A. V. (2008). Modeling students' mathematics using Steffe's fraction schemes. *Teaching Children Mathematics, 15*(1), 48–54.

Perry, R., & Lewis, C. (2010). Building demand for research through lesson study. In C. E. Coburn & M. K. Stein (Eds.), *Research and practice in education: Building alliances, bridging the divide* (pp. 131–145). Lanham, MD: Rowman & Littlefield.

Post, T. R., Harel, G., Behr, M. J., & Lesh, R. A. (1988). Intermediate teachers' knowledge of rational number concepts. In E. Fennema (Ed.), *Papers from the First Wisconsin Symposium for Research on Teaching and Learning Mathematics* (pp. 194–219). Madison, WI: Wisconsin Center for Education Research.

Puchner, L. D., & Taylor, A. R. (2006). Lesson study, collaboration and teacher efficacy: Stories from two school-based math lesson study groups. *Teaching and Teacher Education, 22*(7), 922–934.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Remillard, J. T., & Bryans, M. B. (2004). Teachers' orientations toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education, 35*(5), 352–388.

Saxe, G. B. (2007). Learning about fractions as points on a number line. In W. G. Martin, M. E. Strutchens, & P. C. Elliott (Eds.), *The Learning of Mathematics* (Vol. 69th Yearbook, pp. 221–237). Reston, VA: The National Council of Teachers of Mathematics.

Schifter, D. (1998). Learning mathematics for teaching: From the teachers' seminar to the classroom. *Journal of Mathematics Teacher Education, 1*(1), 55–87.

Schorr, R. Y., & Koellner-Clark, K. (2003). Using a modeling approach to analyze the ways in which teachers consider new ways to teach mathematics. *Mathematical Thinking and Learning, 5*(2), 191–210.

Seymour, J. R., & Lehrer, R. (2006). Tracing the evolution of pedagogical content knowledge as the development of interanimated discourses. *The Journal of the Learning Sciences, 15*(4), 549–582.

Sherer, J., & Spillane, J. (2011). Constancy and change in work practice in schools: The role of organizational routines. *Teachers College Record, 113*(3), 611–657.

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education, 20*(2), 163–183.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*, 20–37.

Smith, T., & Desimone, L. M. (2003). Do changes in patterns of participation in teachers' professional development reflect the goals of standards-based reforms? *Educational Horizons, 81*(3), 119–129.

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education lessons learned from the what works clearinghouse (Phase I). *Educational Evaluation and Policy Analysis, 32*(3), 351–371.

Spillane, J. P. (2000). A fifth-grade teacher's reconstruction of mathematics and literacy teaching: Exploring interactions among identity, learning, and subject matter. *The Elementary School Journal, 100*(4), 307–330.

Steinberg, R. M., Empson, S. B., & Carpenter, T. P. (2004). Inquiry into children's mathematical thinking as a means to teacher change. *Journal of Mathematics Teacher Education, 7*(3), 237–267.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Summit Books.

Takahashi, A., Watanabe, T., Yoshida, M., & Wang-Iverson, P. (2005). Improving content and pedagogical knowledge through kyozaikenkyu. In P. Wang-Iverson & M. Yoshida (Eds.), *Building our understanding of lesson study* (pp. 77–84). Philadelphia: Research for Better Schools.

Taylor, A. R., Anderson, S., Meyer, K., Wagner, M. K., & West, C. (2005). Lesson study: A professional development model for mathematics reform. *Rural Educator, 26*(2), 17–22.

Thijs, A., & van den Berg, E. (2002). Peer coaching as part of a professional development program for science teachers in Botswana. *International Journal of Educational Development, 22*, 55–68.

Ticha, M., & Hospesova, A. (2006). Qualified pedagogical reflection as a way to improve mathematics education. *Journal of Mathematics Teacher Education, 9*(2), 129–156.

Tobin, K., & Espinet, M. (1990). Teachers as peer coaches in high school mathematics. *School Science and Mathematics, 90*(3), 232–244.

Van de Walle, J. A. (2007). *Elementary and middle school mathematics: Teaching developmentally* (6th ed.). Boston, MA: Pearson.

van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*(2), 244–276.

Wang-Iverson, P., & Yoshida, M. (2005). *Building our understanding of lesson study*. Philadelphia: Research for Better Schools.

Ward, J., & Thomas, G. (2009). From presentation on teaching scenarios. Retrieved June 1, 2009, from http://www.nzmaths.co.nz/fractions

Zech, L. K., Gause-Vega, C. L., Bray, M. H., Secules, T., & Goldman, S. R. (2000). Content-based collaborative inquiry: A professional development model for sustaining educational reform. *Educational Psychologist, 35*(3), 207–217.

Zhou, Z., Peverly, S. T., & Xin, T. (2006). Knowing and teaching fractions: A cross-cultural study of American and Chinese mathematics teachers. *Contemporary Educational Psychology, 31*(4), 438–457.