Desimone, L., & *Stuckey, D. (2014). Sustaining professional development. In. L. Martin, S. Kragler, D. Quatroche, & K. Bauserman (Eds.), *Handbook of professional development in Education: Successful Models and Practices, Prek-12* (pp. 467-482). New York, NY: Guilford Publications.

[cn]**Chapter 24**

[ct]**Sustaining Teacher Professional Development**

[ca]**Laura M. Desimone**

[ca]**Daniel Stuckey**

- [cint]Effective and sustainable professional development programs share certain core features.

- Accountability related to the quality of professional development and classroom monitoring and support is crucial for driving improvement.

- Professional development should be targeted to teacher's individual learning needs.

- Professional development works best in an environment of coherent policy.

- School administrators should have realistic expectations about what professional development is likely to do and when.

[a]**Sustainability in Professional Development**

To identify sustainable practices in teachers' professional development, we first identify the characteristics of high-quality, effective professional development. Second, we discuss how and why implementation of professional development in the classroom may vary, as well as challenges to measuring and studying professional development's effectiveness. Last, we discuss how to sustain models of high-quality professional development over time, at both the district and school levels.

[b]Defining High-Quality Professional Development

A growing body of empirical research suggests that effective professional development programs share a core set of features (e.g., Desimone, 2009; Garet et al., 2010; Penuel, Gallagher, & Moorthy, 2011). These core features include the following: (1) *content focus*: activities focused on subject matter content and how students learn that content; (2) *active learning*: opportunities for teachers to observe, receive feedback, analyze student work, or make presentations, as opposed to passively listening to lectures; (3) *coherence*: content, goals, and activities that are consistent with the school curriculum, teacher knowledge, and beliefs, the needs of students, and school, district, and state policies; (4) *sustained duration*: professional development activities that are ongoing throughout the school year and that include 20 hours or more of contact time; and (5) *collective participation*: groups of teachers from the same grade, subject, or school participate in professional development activities together to build an interactive learning community.

Over the past 5–8 years, several rigorous, randomized controlled trials (RCTs) have attempted to build off of the correlational and observational studies of professional development, experimentally testing the importance of some or all of these features of high-quality professional development. We draw on these recent RCTs on professional development to forward thinking about what works and how to sustain what works.

First, we discuss the role of implementation fidelity. Second, we discuss how different types of professional development (e.g., activities targeted to knowledge, behavior, curricula, or a combination) present different challenges and expected benefits. Third, we highlight the importance of furthering our understanding of how teachers vary in participation and response to learning activities. Fourth, we propose the policy-attributes theory as a way to organize thinking

about what needs to be in place in the policy environment to sustain models of high-quality professional development. Finally, we consider the results of recent professional development interventions and make preliminary conclusions about the type of outcomes we expect and how we might improve them.

Although the RCTs we discuss here are more rigorous than previous research, they are not perfect. They do not measure outcomes as often as we might like (e.g., once a year instead of periodically throughout the year); the control group is not "pure," because teachers in the control group are usually engaged in alternative learning experiences; and the studies are usually underpowered (i.e., enrolling too few teachers to detect small effects). Nevertheless, the findings from this recent and developing rigorous knowledge base point to several key areas we think hold promise for improving our understanding of how to achieve quality and sustainability in teachers' professional development.

[b]Theory of Change and Instruction: The Role of Implementation in Effects and Sustainability

Sustainability relates to whether both the theory of instruction and the theory of change work. That is, when implemented well, (1) does the new content or pedagogy learned in professional development improve student learning (theory of instruction), and (2) how well do the professional development activities elicit improvements in teacher knowledge and instruction (theory of change; Wayne, Yoon, Zhu, Cronen, & Garet, 2008)? If the intervention causes teachers to implement better instructional practices, then the intervention should be sustained. Positive effects on teaching and learning are probably the most powerful mechanism for institutionalizing an intervention because positive effects foster teacher buy-in and motivation (Desimone, 2002).

There is a substantial literature on instructional best practice, but identifying aspects of professional development that reliably change teacher behavior has been more elusive. Barriers to teacher implementation of instructional best practice include professional development that falls short on the core features (e.g., too short, not enough practice, not coherent with other initiatives, no opportunities for feedback, not integrated into the curriculum). We would expect that all the core features need to be in place for a reform to last in a school. For example, in an RCT of reading comprehension and vocabulary instruction in first grade, Gersten, Dimino, Jayanthi, Kim, and Santoro (2010) hypothesize that their professional development would have been more successful in changing practice if the professional development had been enacted for the duration that was originally planned; school officials allowed teachers to participate only during their planning time "due to scheduling constraints" (p. 731).

Even professional development with all the core features of quality sometimes does not have the strong effects on teachers and students that we might hope for. One way to further our understanding of why this might be is to examine effects across different teachers, which we do next.

[c]Variation in Implementation

In their simplest form, most theories of professional development boil down to this: Professional development alters teacher classroom behavior, which in turn alters student performance (see Yoon, Duncan, Lee, Scarloss, & Shapley, 2007, p. 4). Therefore, we expect that student outcomes will be heavily influenced by the degree to which teachers actually alter their behavior vis-à-vis a professional development intervention. In other words, fidelity of implementation plays a key role in determining whether professional development's effects on teachers and students will be realized.

Recent RCTs have shown that intervention models often include models of change that work for *some but not all teachers* and that student outcomes are higher for teachers who have higher fidelity of implementation, meaning that their execution of the desired instruction is closer to the ideal envisioned by the designers of the intervention. For example, Davidson, Fields, and Yang (2009), in a study of preschool phonological awareness, found no significant main effects for student performance but did find higher outcomes for the students of teachers who showed higher fidelity to the intervention. Thus the intervention failed to boost average student performance, but when teachers changed their behaviors to better align with the intervention, student performance did improve. When student outcomes are better for high-fidelity teachers, there is evidence that a professional development's theory of instruction is sound, even if its theory of change is not able to transform all teachers into "high implementers." This idea is consistent with past findings that a key driver of student improvement may be increasing teacher fidelity to interventions (e.g., Desimone, 2002). Professional development can be successful for subsets of teachers even when it does not seem to be successful on the whole. Understanding why professional development is successful with some teachers and understanding how to expand that subset is crucial for improving overall student achievement.

But how does this help us learn how to foster more widespread high implementation? The experimental literature on professional development tends to focus on average effects across teachers and students, and the experiments are designed and randomized to test these average effects. However, we suggest that professional development affects teachers differentially, depending on their previous knowledge, their level of experience, and other factors.

Of course, teachers differ from one another in many ways. Although this may seem obvious, Piasta and colleagues (2010) stress that teacher variability is a "major finding" and encourage

future research to explore "teacher × professional development interactions" (p. 369). We have not reviewed any study that incorporates teacher variation on key factors such as experience or a priori knowledge into the design of the experiment to see which teachers are most likely to take up the professional development. However, it is common to conduct exploratory or post hoc analyses to try to understand why some teachers became high implementers and others did not. These suggest that particular teacher characteristics do matter.

[d]Experience as a Key Source of Variation. Five recent RCT studies explored whether professional development affected novice and veteran teachers differently. Of these five, three found no differences—Garet and colleagues (2010), Gersten and colleagues (2010), and Sailors and Price (2010). Of the two other studies, one found that veteran teachers had higher fidelity, whereas the other found that novice teachers were the ones with higher fidelity. Davidson and colleagues (2009) found that veteran teachers were stronger implementers of a program in phonological awareness. On the other hand, Borman, Gamoran, and Bowdon (2008) found a significant interaction effect between years of experience and professional development around inquiry-oriented science teaching in the fourth grade. On average, the Borman and colleagues intervention had significantly negative main effects on students' elementary science skills. However, when the researchers looked only at students who were taught by teachers with fewer than 3 years of experience, the intervention had positive effects on student science achievement. Based on this finding, they conclude that "it appears that students of new teachers did benefit to a greater extent than students of more experienced teachers from their teachers' exposure to professional development in science inquiry" (p. 255).

One explanation of these mixed results may be that, in certain circumstances, veteran teachers have higher uptake of more "traditional" practices (e.g., phonological awareness), whereas

novice teachers might be more adept at reform-oriented professional development. It might also be that novice teachers are more open and flexible about new approaches, whereas veteran teachers are less likely to change easily because they have ingrained, established practices. Future studies could use block designs to explore the differential effects of professional development on novice versus veteran teachers, as well as other potentially influential characteristics.

[d]Content Knowledge as a Key Source of Variation. The relationship between teachers' content knowledge and their fidelity of implementation is mixed, but it seems like a potentially important variable to consider. From Davidson and colleagues (2009), we see evidence that teachers' content knowledge matters very little: "there was not a relation between teacher scores on the [teacher knowledge test] and child performance on the outcome measures. The anticipated moderating effect of teacher knowledge did not occur" (p. 195). On the other hand, in a study of sixth-grade teachers' ability to teach fractions, ratios, and equations, Santagata (2009) reported higher student outcomes when teachers' math content knowledge was higher. To explain this result, Santagata suggests that teachers have difficulty focusing on intervention strategies or curricular changes when they do not understand the content well. In our own forthcoming randomized study (Desimone, Covay, & Caines, 2013), we witnessed similar trends. Some teachers with low content knowledge made use of the materials of the intervention without giving deep attention to the scientific principles behind the intervention, whereas other teachers indicated that they thought the best aspect of professional development was an opportunity to improve their content knowledge, even though this particular intervention was not set up for that purpose. Teachers also may react differently to each other and to professional development providers based on content knowledge differences. For example, Yoon, Liu, and Goh (2010)

describe how members of a professional development team shunned another member seemingly *because* he had higher content knowledge.

Thus variation in implementation due to teacher characteristics is a critical area of attention for schools and districts interested in sustaining the effects of professional development for all of their teachers. One powerful way for districts and schools to increase the effectiveness and sustainability of their professional development efforts is to acknowledge the way that professional development affects teachers differentially and to adjust it accordingly. This may mean adding additional supports for certain types of teachers, focusing professional development efforts on those teachers that they are most likely to work for, or choosing professional development programs that are best suited to the faculty of a particular school or district.

[c]The Role of Teacher Buy-In

Although it was not an explicit focus of any of the RCTs we surveyed, a synthesis of findings and previous research suggests that strong teacher buy-in goes a long way to improving implementation and facilitating sustainability (Desimone, 2002). We hypothesize that voluntary interventions have higher initial buy-in than mandatory interventions. If we use voluntary as a proxy for high buy-in, we see that when interventions are voluntary, student outcomes are improved. Of the twelve recent experimental studies, three were mandated, and of those three, one (Borman et al., 2008) had negative effects on student outcomes and the other two (Garet et al., 2010; Santagata, 2009) had no effects on student outcomes. In contrast, of the three professional development opportunities that were clearly voluntary, one (Gersten et al., 2010) had no effect on student outcomes, but the other two (Penuel et al., 2011; Sailors & Price, 2010) had significantly positive effects on student outcomes. Districts will frequently want to provide

professional development to all of their staff, not just to those who are willing to volunteer, but the importance of fostering buy-in should not be ignored.

[c]Implementation Dip

Another issue related to implementation is the "implementation dip," noted in many circumstances in which new knowledge and behavior are asked of teachers (Fullan, Cuttress, & Kilcher, 2005). Teacher and student performance can get worse before it gets better. Because studies usually last only 1 year, we do not have a good understanding of the ebb and flow of teacher learning or the way that implementation of new instruction continues into year 2, year 3, and beyond as a result of high-quality professional development. Borman and colleagues (2008) hypothesized that an implementation dip was responsible for the negative results of their professional development experiment, but it would take further years of study to determine whether that was actually the case. Dips in implementation (and in standardized test scores as a result) can be catastrophic for schools and districts because such dips are often used as reasons to abandon the reform before it has had a chance to take hold. Thus it is helpful to take a longer-term view of a teacher learning intervention and to establish reasonable expectations about when effects might occur (e.g., in year 2 rather than year 1).

[c]Depth of Implementation Measures Vary

Another complication in studying implementation is how we measure it. Methods range from checklists (Davidson et al., 2009) to observations (Gersten et al., 2010; Piasta et al., 2010; Sailors & Price, 2010) to surveys (Santagata, 2009), with each method offering advantages and disadvantages. There is also the possibility that the instruments used to measure implementation are not sensitive to differences among treatment and control (e.g., Kim et al., 2011). To address these possible weaknesses, any professional development intervention should have a

comprehensive system of evaluation, which includes observation, self-report, opportunities for feedback, and practice. These mechanisms help ensure that teachers will have the continuing monitoring and feedback shown to be necessary for teacher learning to transfer to institutionalized classroom instruction change (Cohen & Ball, 1990).

[b]The Interaction of Dosage and Substance: Are We Aiming to Change Knowledge, Behavior, or Decision Making?

Beyond teacher characteristics and the ebb and flow of change represented by the implementation dip, the extent to which a professional development intervention is successful depends on both its duration and its focus. Unless teachers receive a certain "dosage" of hours or sessions, it seems unlikely that professional development will be effective. In the experiments reviewed by Yoon and colleagues (2007), for example, teachers experienced 49 hours, on average, and students performed significantly better whenever their teachers experienced at least 14 hours of professional development.

 In the more recent experiments, teachers have received as many as 114 hours (Garet et al., 2010) without significantly influencing student outcomes. On the other hand, professional development was sometimes effective even at a low dosage. In particular, the preschool literacy intervention studied by Piasta and colleagues (2010) is good news for those hoping to implement sustainable professional development programs. The researchers deliberately examined a program that fit within the resources that schools and districts usually have available to them. The professional development took only 11 hours, the teachers missed little teaching, and the curricular materials were relatively inexpensive. Monitoring, accountability, and feedback were accomplished through a pattern of submitting videos and receiving letters, thereby obviating the need for expensive coaching.

Why would an 11-hour professional development produce student achievement gains, whereas a 114-hour professional development failed to? Clearly, factors besides dosage contribute to an intervention's effectiveness. One possible explanation is that the 114-hour intervention focused on content, whereas the 11-hour intervention focused on changing specific instructional behaviors.

As we reviewed 12 recent RCTs, we noticed that interventions tended to fall into one or more of these three categorizations: (1) examining academic content, (2) changing specific teacher behaviors, or (3) improving teacher decision making. The 114-hour mathematics intervention studied by Garet and colleagues (2010) focused primarily on content. First, it aimed to improve teachers' content knowledge of rational numbers, and second, it aimed to help teachers deliver more "precise definitions…explain rational number concepts and procedures, identify and address persistent student misconceptions…and use representations of rational number concepts in teaching" (p. xvi). Through deep engagement with content, developers hoped to improve teachers' understanding of rational numbers and their ability to explain the content to students. However, teachers who received professional development scored no better than control teachers on tests that measured their "common knowledge (CK) of mathematics and specialized knowledge (SK) of mathematics for teaching" (p. 14).

[c]Content Knowledge

This pattern was common across the studies that examined teacher content knowledge. In four out of five studies (Davidson et al., 2009; Garet et al., 2010; Neuman & Cunningham, 2009; Santagata, 2009), teachers experiencing professional development showed no significant improvement in content knowledge. In the fifth study (Gersten et al., 2010), teachers showed significant improvement in knowledge for teaching vocabulary, but not in knowledge for

teaching comprehension. Though common sense tells us that improving teachers' content knowledge is a good idea for improving student outcomes, the studies demonstrate how hard it is to improve teachers' content knowledge.

[c]Pedagogy

A second approach to professional development focuses less on content and more on curricular, pedagogical, or behavioral changes. This type of intervention provides teachers with additional materials or suggests specific instructional strategies that teachers should use. The Piasta and colleagues (2010) study is emblematic. Teachers in the treatment group were asked to read a specific book to preschool students four times a week; the books included inserts alerting teachers to how they might explicitly make "print references when reading" (p. 353). Similarly, Sailors and Price (2010) studied teachers who were asked to teach a discrete set of cognitive reading strategies to students in grades 2–8. Coaches helped these teachers to modify the curriculum and sometimes engaged in team teaching of the cognitive strategies, leading to significant positive effects.

On the other hand, Borman and colleagues (2008) found significant *negative* effects for a professional development intervention meant to transition teachers into a new elementary science curriculum based on inquiry-oriented practices. These negative effects may be due to the relative complexity of inquiry-oriented practices compared with the relative simplicity of reading strategies such as *ask questions* and *make predictions*. The negative results may also be due to the inadequacy of a train-the-trainer model; in the Borman study, only one teacher from each school received the training, and that teacher was then responsible for training the other teachers. Together, the Piasta and colleagues (2010), Sailors and Price (2010), and Borman and colleagues (2008) studies suggest that simple behavioral changes can be effected by professional

development and can have positive effects on student achievement but that more complicated behavioral changes are less tractable. Of course, professional development is likely to focus on content *and* instructional behavior. Neuman and Cunningham (2009) write, "professional development that contains both content and pedagogical knowledge may best support the ability of teachers to apply literacy knowledge in practice" (p. 534). Santagata (2009) suggests that the superficial changes brought about by these pedagogical interventions are necessary prerequisites for establishing more transformative change.

[c]Decision Making

A third approach focuses on teacher decision making, which has received less empirical attention. If some researchers hypothesize that the key to improving professional development outcomes is increasing fidelity to a set program, Penuel and colleagues (2011) hypothesize that teachers will always adapt curricula and professional development to suit their own settings and needs. Therefore, Penuel and colleagues attempt to build an adaptive professional development program that helps teachers make better decisions about content and pacing, helping teachers to see how to use both "expert- and teacher-designed activities" to teach earth science to students in grades 6–8 (p. 1018). The students of teachers receiving such development—focused on decision making and adapting materials—had significantly higher science achievement than the students of teachers who received only a new curriculum. Ultimately, the key to sustainable professional development might not be putting stable practices in place but rather helping teachers become adaptive planners capable of making good decisions over time.

[b]The Policy-Attributes Theory

Now we take a step back to consider the broader organizational and policy context that facilitates implementation and sustainability. Recent studies of professional development (and previous

research on reform) have identified several organizational variables that are strongly linked with the initial success and sustainability of reform focused on teacher learning. These organizational variables can be described by a policy-attributes framework, which identifies authority, power, coherence, and specificity as key factors in determining the effects and longevity of a reform (Desimone, 2002). The recent RCT studies of professional development that we review here shed light on which aspects of these attributes may be most important for the sustainability of high-quality professional development.

[c]Authority

As with most reform efforts, professional development interventions rely to a great extent on the leadership support at the school and district levels. Although some studies are not clear on how leadership influenced the implementation and success of their professional development interventions (e.g., Borman et al., 2008; Davidson et al., 2009), others emphasize the crucial role that principals can play in structuring time for teachers to participate in professional development (Gersten et al., 2010). Santagata (2009) provides an example of how tension between teachers and the district can cause teachers to view a professional development program unfavorably if teachers associate the professional development with the district. The influential role that support and backing from authority figures can play in improving initial implementation and increasing the longevity of an intervention is well documented in the broader school reform literature (e.g., Desimone, 2002).

[c]Power (Accountability Pressure)

Previous studies have shown that power exerted through the pressure of rewards or sanctions can alter teacher behavior, but such changes are usually not as long-lasting as behavior changes that result from self-motivation or buy-in (e.g., Desimone, 2002). One critical question is, To what

extent should we encourage districts and schools to use accountability pressures to motivate teachers to adopt practices encouraged in professional development?

It seems possible that teacher change is highly mediated by power, or accountability pressures. For example, Santagata (2009) found that when observed, teachers "treated" with professional development spent significantly more time working on assessment problems than did control teachers, but that treated teachers did not increase the cognitive demand on their students. This may suggest that teachers felt pressure to comply but did so in a superficial way, rather than deeply changing their underlying behaviors in ways proposed by professional developers. Garet and colleagues (2010) hypothesize that one reason their professional development on teaching rational numbers to middle schoolers may not have had strong effects was that there was no pressure from the district or school for teachers to change what they were doing.

Further support for the possible effects of accountability comes from Neuman and Cunningham (2009), who found that when preschool child-care workers received a course on preliteracy skills, the teachers did not change their practice, but that when the course was paired with coaching at the child-care center, workers *did* significantly change their practice. In this case, we might construe coaching at least in part as an accountability mechanism—when teachers know they will be observed by experts looking for particular teaching behaviors, they may be more likely to adopt those new behaviors. Videotaping and feedback may work similarly (Piasta et al., 2010). In this way, accountability mechanisms can be considered a type of "implementation driver" (Hulleman & Cordray, 2009).

[c]Coherence

The extent to which the professional development is compatible with the curriculum and standards the teacher is using can play a substantial role in influencing a teacher's willingness

and ability to adopt new content or practices. Penuel and colleagues' (2011) elegant study was set up for success. The participating district was engaging in large-scale change, and the professional development intervention was intentionally integrated with the more holistic change that was occurring across the district. In contrast, when districts communicate competing priorities, creating an incoherent environment for teachers, we would expect implementation to be weak (e.g., Santagata, 2009).

Another dimension of coherence may be autonomy; in Neuman and Cunningham's (2009) study, in which teachers had autonomy to change, there was little chance the professional development was incoherent with demands on them because teachers did not have many rules and regulations to abide by. Perhaps the most troubling circumstance occurs when we are asking teachers to change inside a system that is not itself changing, or at least not openly supportive of change. Thus it is important for districts and schools to align their teacher learning initiatives with other reforms that teachers are being asked to follow.

[c]Specificity

Previous research has demonstrated that when a policy or intervention clearly states the activities and behaviors that are the target, teachers are more likely to show higher implementation. We read the recent professional development literature as coalescing on a similar finding: The more concrete the behaviors asked for in the professional development, the more likely teachers are to be high implementers. For example, teachers are more likely to be able to implement a concrete, specific task—such as more use of phonological exercises (Davidson et al., 2009)—than to engage in a more nuanced, complicated task, such as inquiry-oriented science practices (Borman et al., 2008). This is likely because more conceptual teaching requires a complex knowledge base and sustained practice (Cohen & Ball, 2002), whereas more process- or pedagogically oriented

behaviors are likely much easier to implement. Similarly, interventions are more likely to find effects on more proximal outcomes specific to the intervention rather than on more broad measures such as standardized tests.

[a]Discussion: What Recent Intervention Research Tells Us about the Sustainability of Professional Development

Given contemporary reform pressures, professional development programs are likely to be sustained only if they can show results. Ultimately, "showing results" may require improving student achievement on the state's standardized tests, but districts seeking sustainability should first look for more gradual or proximal signs of success. When developing and evaluating professional development programs, districts should remember several lessons about how professional development affects student outcomes: (1) Professional development is likely to improve some teachers' instruction more than other teachers' instruction; (2) teacher behaviors are likely to change before student achievement changes; (3) it may take a while for student improvement to occur; and (4) student improvement may be too narrow to register on state tests.

[b]Teachers First

Improving student achievement is difficult to do, and even when student achievement is accomplished, it may be difficult to measure or confirm. However, the recent batch of RCTs demonstrates that professional development is good at eliciting some kinds of behavioral changes in teachers. Professional development interventions can seek to affect teacher behavior, teacher knowledge, or both. Nine of the recent studies measured whether teacher behaviors or practice changed as a result of the professional development, and in seven of them, researchers found significantly positive behavioral changes. In these seven cases, researchers thought that

teachers who experienced professional development acted in ways more in line with best practice.

Of course, best practice is always theoretical unless empirical support shows that such behaviors lead to improved student outcomes—but the important lesson is that professional development gets teachers to change. If the professional development is based on an appropriate theory of instruction, it should get student outcomes to improve if given enough time.

On the other hand, increasing teacher content knowledge seems more difficult. We identified four studies (Garet et al., 2010; Gersten et al., 2010; Neuman & Cunningham, 2009; Santagata, 2009) in which the researchers attempted to improve teacher content knowledge and also provided a posttest of content knowledge in order to test significant differences between teachers in the control and treatment groups. Three of these studies showed no significant effects, and the fourth (Gersten et al., 2010) showed increased teacher knowledge of best practices in vocabulary instruction but not best practices in reading comprehension instruction. These studies are the most recent evidence that building teacher content knowledge is complicated and requires sustained, ongoing, high-quality supports throughout a teacher's career (Cohen & Ball, 1990).

[b]Teacher Variability

As discussed earlier in this chapter, professional development programs have variable effects on teachers and their students. Although professional development seems to change teachers' behaviors *on average*, it does not change all teachers equally. At the inception of professional development programs, districts may want to focus on the teachers who stand to benefit the most from development, either because they are the teachers in the most need of improvement or because their characteristics imply that they are most likely to change their behavior. As professional development continues, districts may want to add additional supports or

accountability pressures in order to further the development of teachers who have shown the least behavioral change during the early stages of the professional development efforts. Another idea is to structure and plan professional development targeted to teachers with different levels of classroom experience or content knowledge.

[b]Time

Ideally, professional development programs would have immediate impact on student achievement, but because improving student achievement is difficult, professional development is unlikely to have immediate effects and may even have initially negative effects. As mentioned, Fullan and colleagues (2005) calls this the implementation dip, referring to the ways teachers stumble as they first experiment with new practices.

The recent RCTs fail to show the longitudinal effects of professional development programs. None of the 12 recent interventions engaged teachers for more than 2 years, and none of the interventions measured students for longer than 1 year. Thus it remains unknown how teacher or student outcomes persist over time. Although several studies reported teachers changing their behaviors during an intervention, it is unknown whether teachers maintain new behaviors once the intervention is over. (In light of concerns that teachers are really responding to accountability pressures—not professional development—we might be especially worried that teacher behavior would not persist absent ongoing development or accountability pressure.) In a similar vein, we do not know how student outcomes are affected over time. Because none of the trials tracks student achievement in subsequent years, we cannot know whether student gains persist or fade away, nor whether there are any latent effects on student achievement. All of these questions are worthy of investigation.

[b]Narrow Success

Yoon and colleagues' (2007) meta-analysis of high-quality professional development studies found that across nine interventions, student outcomes improved on 18 out of 20 measures, and for the RCTs, the average effect size was 0.51. However, the RCTs since 2007 imply that professional development is considerably less successful at effecting student achievement gains. Of the nine studies that report student outcomes, one reports negative effects on fourth graders' performance on an elementary science benchmark (Borman et al., 2008). Four studies report no effects on tests that measure either elementary math or literacy (Davidson et al., 2009; Garet et al., 2010; Gersten et al., 2010; Santagata, 2009), and only three demonstrate significant positive results. Kim and colleagues (2011) find that Latino English language learners in grades 6–12 get better at writing text-based, analytical essays. Sailors and Price (2010) find that students in grades 2– 8 improve their reading comprehension. Penuel and colleagues (2011) find that students in grades 6–8 have higher achievement on a test of earth science knowledge. Although we often hope that professional development will improve student achievement across a content area, professional development is more likely to improve student achievement on some narrow skill. For example, Simmons and colleagues (2010) found that fourth-grade students of teachers receiving professional development did significantly better than their peers on an achievement test that closely matched the content of the intervention. However, gains were washed out on a broader test of reading comprehension and vocabulary. The researchers note that this finding is consistent with earlier research on professional development; it seems easier to effect change on targeted content than to effect the generalized improvement necessary to improve state test scores. If policymakers and school leaders are interested in using professional development to improve achievement, they should target their professional development to the skills students are most lacking or to the skills that are most generalizable.

Districts should keep these lessons in mind. Lack of results can cause a district to abandon professional development, but districts would be wise to look for proximal results in the short term. If teacher behaviors are changing, or if student scores are changing on more narrow measures, these proximal results can be the flywheel of continued improvement, and they can provide an argument for sustaining current professional development activities.

[a] Recommendations and Future Directions

From our reading of recent rigorous causal studies of professional development, we draw several conclusions and suggest several ideas that may be helpful in designing sustainable professional development.

First, for practices to become institutionalized, professional development must have the core features of quality—content focus, active learning, coherence, sufficient duration, and collective participation.

Second, we emphasize the importance of ongoing support and monitoring of implementation. This includes comprehensive assessment, which might include videotaping instruction, observation, feedback, discussion, practice, and modeling.

Third, districts seeking sustainable and effective professional development programs should keep in mind several lessons about teacher variation. Teachers vary in their ability and their responses to professional development. Even when professional development programs boost average student outcomes, the effects are likely to be uneven. In light of teacher differences, districts might consider differentiating professional development according to teacher characteristics. In order to maximize effects, districts might match teachers with development opportunities most likely to suit their knowledge, experience, and learning needs. Fourth, teachers are more likely to implement professional development interventions with fidelity when the policy/organizational

environment has the following attributes, integrated and in balance: accountability mechanisms in place to motivate implementation; authority in the form of support and buy-in from school officials and teachers; coherence in terms of integration with curriculum, standards, assessments, and other demands on teachers; and specificity, in terms of identifying specific activities and behaviors to be targeted by the professional development intervention.

Last, this most recent set of RCTs allows us to see both strengths and weaknesses in the way we support teachers. They provide evidence that well-structured, relatively short professional development can change specific behaviors, but changing underlying content knowledge in meaningful ways is much more challenging. Links to student learning are often tenuous and are most easily demonstrated by tests that measure narrow content. The field needs more longitudinal and in-depth studies of how professional development affects individual teachers and also how professional development affects the trajectory of teacher practices and student learning over time. We can only learn to sustain high-quality professional development by sustaining our studies of it.

[Qhd]Questions for Discussion

[Q]1. What evidence should school administrators collect to help them evaluate whether professional development is successful? What counts as success?

2. What strategies can administrators use to increase the chances that professional development will be successful for their teachers?

3. How might administrators decide which type of professional development is most appropriate for their teachers—for example, professional development targeted to using a particular curriculum, adopting a specific type of pedagogy, building content knowledge, or training teachers to make better decisions about how to adapt materials to their students?

4. What is the boundary between teacher professional development and the rest of the work that teachers do? Is professional development best conceived of as a "program" or an "intervention," or does the goal of sustainability require some other characterization?

5. What should we make of the discrepancy between the large and consistent effects on student achievement found by Yoon and colleagues (2007) and the more muted and inconsistent effects of the RCTs completed since 2007?

[ref hd]References

[ref]Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. *Journal of Research on Educational Effectiveness*, *1*(4), 237–264.

Cohen, D., & Ball, D. L. (1990). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis, 12*(3), 249–256.

Davidson, M. R., Fields, M. K., & Yang, J. (2009). A randomized trial study of a preschool literacy curriculum: The importance of implementation. *Journal of Research on Educational Effectiveness*, *2*(3), 177–208.

Desimone, L. M. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*(3), 433–479.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Desimone, L. M., Covay, E., & Caines, J. (2013*). Implementation of a middle school science curriculum intervention: Teacher challenges* (Working paper). Philadelphia: University of Pennsylvania.

Fullan, M., Cuttress, C., & Kilcher, A. (2005). Forces for leaders of change. *Journal of Staff Development*, *26*(4), 54–59.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., et al. (2010). Middle school mathematics professional development impact study: Findings after the first year of implementation (NCEE Publication No. 2010-4009). Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance.

Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first-grade classrooms. *American Educational Research Journal*, *47*(3), 694–739.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88–110.

Kim, J. S., Olson, C. B., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., et al. (2011). A randomized experiment of a cognitive strategies approach to text-based analytical writing for mainstreamed Latino English language learners in grades 6 to 12. *Journal of Research on Educational Effectiveness*, *4*(3), 231–263.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, *46*(2), 532–566.

Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, *48*(4), 996–1025.

Piasta, S. B., Dynia, J. M., Justice, L. M., Pentimonti, J. M., Kaderavek, J. N., & Schatschneider, C. (2010). Impact of professional development on preschool teachers' print references during shared read alouds: A latent growth curve analysis. *Journal of Research on Educational Effectiveness*, *3*(4), 343–380.

Sailors, M., & Price, L. R. (2010). Professional development that supports the teaching of cognitive reading strategy instruction. *Elementary School Journal*, *110*(3), 301–322.

Santagata, R. (2009). Designing video-based professional development for mathematics teachers in low-performing schools. *Journal of Teacher Education*, *60*(1), 38–51.

Simmons, D., Hairrell, A., Edmonds, M., Vaughn, S., Larsen, R., Willson, V., et al. (2010). A comparison of multiple-strategy methods: Effects on fourth-grade students' general and content-specific reading comprehension and vocabulary development. *Journal of Research on Educational Effectiveness*, *3*(2), 121–156.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, *37*(8), 469–479.

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*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance.

Yoon, S., Liu, L., & Goh, S. (2010). Convergent adaptation in small groups: Understanding professional development activities through a complex systems lens. *Journal of Technology and Teacher Education*, *18*(2), 319–344.