Mechanisms Affecting the Sustainability and Scale-up of a System-wide Numeracy Reform

Janette Bobis

University of Sydney

With deliberate system-level reform now being acted upon around the world, both successful and unsuccessful cases provide a rich source of knowledge from which we can learn to improve large-scale reform. Research surrounding the effectiveness of a theory-based system-wide numeracy reform operating in primary schools across Australia is examined to extract the elements most influential for the sustainability and scaling-up of this reform. The evidence suggests that key mechanisms or tools such as a framework in number and diagnostic interview, alongside a focus on deep understanding in teachers of children's mathematical thinking strategies and the strategic placement of support structures at various levels of program implementation are important elements for the sustainability and scale-up of this program. This article introduces the professional development program in question, presents an overview of three research studies, and then links their findings to Coburn's (2003) multi-dimensional conceptualisation of scale.

The ability to sustain and scale up programs of professional development remain key issues in educational reform and improvement. Indeed, sustainability is a central challenge when bringing reforms to scale. In her reconceptualisation of scale, Coburn (2003) emphasises the importance of 'depth' in sustaining school reform. She asserts that teachers with deep understanding of underlying reform principles are more able to sustain reform practices over time and when challenges arise. Furthermore, she claims that teachers and schools are supported in this process when tools or mechanisms designed to assist sustainability of reforms are established at various levels of the school system. This paper draws upon the findings of three investigations involving district-level consultants, schools and classroom teachers engaged in a theory-based professional development program to discuss elements affecting sustainability of system-wide educational reform. This reform is the New South Wales [NSW] Count Me In Too [CMIT] numeracy program (NSW Department of Education and Training [DET], 2007). The sections that follow contextualise research about system-wide educational reform and present a conceptual framework for examining the effectiveness and readiness of reforms for scale-up. This is followed by a description of the program, focusing on its core principles and key elements, and an overview of the three studies forming the springboard for further discussions of sustainability and scale-up of reforms. The impact of the Count Me in Too program on student achievement has been reported elsewhere (see Bobis, 2001; Bobis et al., 2005; White & Mitchelmore, 2002); here the focus is on the effectiveness of specific aspects of the system-wide program across the three studies and spanning five years of research. The benefit of looking at data about the same program across studies that focus on different components (and

sometimes the same components) over an extended timeframe and in different contexts is the ability to identify core elements affecting the sustainability of such large-scale reforms.

Large-scale Reform

With considerable agreement among researchers and reform providers since the mid-1990s about the core elements of professional development that make them effective (Clarke & Clarke, 1998; Garet, Porter, Desimone, Birman, & Yeon, 2001; Loucks-Horsley, Hewson, Love, & Stiles, 1998; Whitcomb, Borko, & Liston, 2009), attention shifted from small-scale designs to examining ways to successfully scale up those designs that showed most promise in terms of teacher learning and student outcomes. By the late 1990s Stringfield and Datnow (1998) were of the opinion that "research on scaling up has just begun" (p. 275), with few studies systematically monitoring and reporting the effects of such programs in different contexts. They considered scaling up to be similar to the replication and expansion of an existing successful small-scale design to a greater number of settings. However, accompanying the deliberate spread came changes to a program's infrastructure necessary to support the reform as it grew. Datnow and Stringfield (2000) also discuss problems surrounding what they refer to as the 'longevity' of reforms, considering that full-implementation of a whole-school reform may take many years before it is truly ingrained in the school culture to the degree that it is 'institutionalised'.

By the turn of the century, scholars were explicitly linking the challenge of reform scale-up to sustainability. McLaughlin and Mitra (2001) drew upon five years of research to exemplify three main factors they thought designers should focus on when initiating reforms for successful scale-up: site selection, a proactive stance toward district contexts, and planned transfer of authority. They explain how in the process of expanding and sustaining reforms, implementation and infrastructure are often revisited and/or invented as the reform goes "deeper and broader" (p. 301) and conclude that issues of sustainability and scale occur simultaneously.

In his review of system-wide educational reforms, Fullan (2009) identifies the period 2003-2009 as the time in which large-scale reform came of age. He considers a "system to include a government and all its schools" and refers to "tri-level reform: schools and their communities, districts or regions, and state" (p. 102). In the period 2003-2009, he argues, large-scale reforms were flourishing—though not all were successful. Nevertheless, the prevalence of examples provided the opportunity to assess specific components so that we are now at a point of some agreement about the necessity of at least three key elements for the effectiveness of whole system reforms: high quality teachers, leaders and teachers working together, and a focus on student learning and achievement. However, as Fullan notes, there is great variation of strategies and much debate about how these elements are actually achieved.

In discussing various strategies used by educational systems that

consistently perform well on international comparative tests for literacy, numeracy and science, Fullan reflects on findings from the McKinsey report (Barber & Mourshed, 2007, cited in Fullan, 2009). This report examined the best performing educational systems in the world, including Finland, Singapore and Canada, to provide examples of successful policies and strategies that were deemed to make a positive difference to student outcomes. One such strategy, which was also highlighted by Timperley, Wilson, Barrar, and Fung (2007) in their report on teacher professional development to the New Zealand Ministry of Education, was the on-going use of data to monitor the progress of individual students, schools and clusters of schools with early intervention used to quickly address problems as they arise. Importantly, Fullan cautions against endorsing *"one factor at a time as key"*, since successful large-scale reform "usually means that a small number (up to half dozen) of powerful factors are interacting to produce a substantial impact. It is the interaction effect that accounts for the results" (p. 108). With deliberate system-level reform now being acted upon around the world, both successful and unsuccessful cases provide a rich source of knowledge from which we can learn to improve system-level reform.

Conceptual Framework for Scaling-up Reforms

Fullan (2009) and Whitcomb et al. (2009) note that large-scale reforms nearly always have their origins as small-scale initiatives that have shown potential for enhancing teacher practices and ultimately student outcomes. However, not all promising small-scale reforms have successfully scaled up and sustained their original intent with integrity (Stringfield & Datnow, 1998; Whitcomb et al., 2009). Scale-up remains a major challenge for reform stakeholders.

Coburn (2003) argues that part of this challenge is created by reform-designers' definitions of scale. Definitions of 'scaling up' not only impact on the strategies and principles reformers build into their programs, but also the way they are studied for their effectiveness. A traditional view of scale refers only to an increase in the number of sites or the replication of a program of reform and does not consider the depth to which a reform's core principles are integrated into practice or the "degree to which it is sustained" (p. 3). While expansion is essential for reforms intended for system-wide implementation, it is insufficient when effective long-term implementation is the aim.

In her multi-dimensional conceptualisation of scale, Coburn (2003) discusses four interrelated conditions necessary for effective scale-up of reforms: depth; sustainability; spread of norms, principles and beliefs; and a shift in reform ownership. She argues that it is only when a scaled-up reform is assessed in terms of each of these dimensions, does one truly get an indication of its effectiveness. Coburn states that for reforms to "effect deep and consequential change in classroom practice" (p. 4) they must go beyond surface-level changes to impact on teachers' core beliefs, knowledge and pedagogical principles underlying their teaching practice. Such depth of change may only be measured via qualitative means such as in-depth interviews and classroom observations.

Fundamental to this multi-dimensional conceptualisation of scale is sustainability. Unless a reform can be sustained, it cannot become large-scale. Sustainability is more than the maintenance of a reform at its current level of implementation; it involves 'depth' of change as beliefs, understandings and practices continue to develop more meaningfully in line with a reform's core principles even after initial external support for the reform is removed (McLaughlin & Mitra, 2001). According to Coburn (2003), sustaining change after external support has ceased is more likely if mechanisms are established at multiple levels of the system. Grossman, Smargorinsky, and Valencia (1999) suggest such mechanisms or tools can be 'conceptual' such as a well-defined set of reform principles or implementation plan. It could also include a theoretical framework for teaching and learning that teachers may use to guide their instruction and inform their professional discussions. Other mechanisms could be 'practical tools' such as protocols for assessing students' understanding or resources that inform classroom practice.

While 'spread' retains its meaning of expansion to other locations, Coburn's conceptualisation clarifies this definition to ensure that the expansion refers to not only surface level features of a reform, but also the underlying core principles, beliefs and norms. She further qualifies the term to refer to expansion *within* an existing reform site to include its deepening impact on, for example, school- or district-level policies and procedures. In this way, spread also refers to the gradual institutionalisation of the reform principles and is therefore inherently linked to sustainability.

Finally, a shift in ownership from the initial reformer to individual schools and teachers ensures that a "reform becomes self-generative" (Coburn, 2003, p. 7). Hence, an important strategy for reform designers is to build in mechanisms and processes allowing the transition of ownership to take place. Part of this capacity building process must focus on developing a supportive community of practice at multiple levels of the system, including active leaders – both principals and influential teaching staff (Coburn, 2005; Fullan, 2009; McLaughlin & Mitra, 2001). Along with the shift in ownership, issues surrounding the on-going development of teachers and problems associated with funding come to the fore.

As mentioned earlier, Coburn's conceptualisation of scale provides a useful framework for examining specific aspects of large-scale reforms, particularly when researchers wish to focus upon those components attributed with the reform's successful implementation. The next section provides a background to the system-wide professional development numeracy program, Count Me In Too, outlining the program's core principles and key elements so as to provide a context for the overview of the studies, the discussion surrounding specific support mechanisms or tools, and their role in sustaining the program.

Program Design and Context

The Count Me In Too [CMIT] program began in 1996 involving only 13 NSW Department of Education and Training [DET] primary schools, gradually scaling

up to involve nearly 1700 primary schools across the state in its first 8 years of implementation. CMIT (NSW DET, 2007) aims to improve the educational outcomes in mathematics for all students through the professional development of K-6 teachers. The theory-based design of the program means that its core principles and key features are informed by a synthesis of research surrounding how children learn mathematics and the effective development of teachers. This synergy of theory and practice, combined with the focus on building a community of learners via links between researchers, teachers, consultants and students, is represented in the program's model of professional development (see Figure 1).
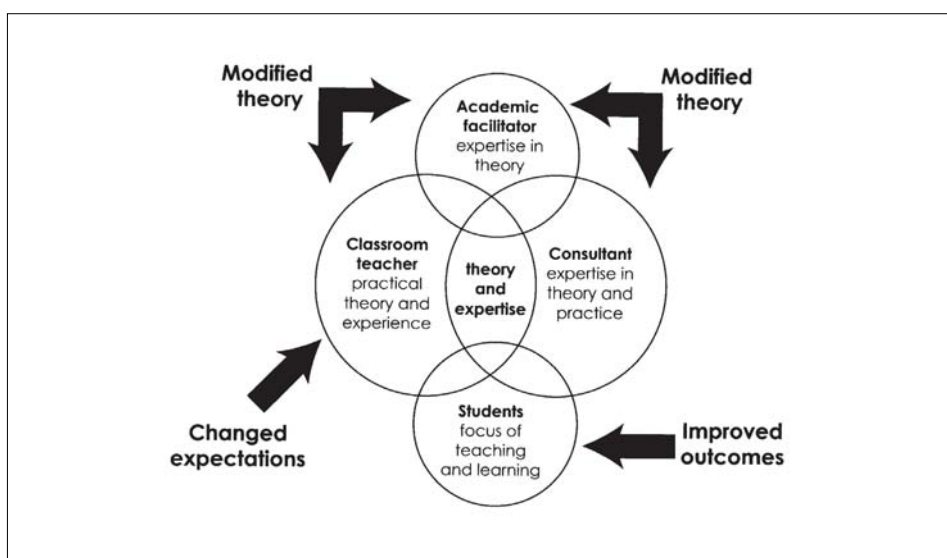

Figure 1. CMIT professional development model
(www.curriculumsupport.education.nsw.gov.au/countmein/
teachers_maintaining_momentum.html).

The program's core principles are aligned with a situated perspective (Putnam & Borko, 2000) of teacher learning and development, namely, the model of teacher learning at the centre of the program does not rely on a pre-determined set of instructional strategies proven successful for learning elsewhere and transferred to a new setting. Instead, the core principles and key elements of the program work in tandem to develop teachers' knowledge to enable adjustments to instruction according to individual students' needs. The professional development program is designed to assist teachers to understand more clearly how children learn arithmetic and is focused on instructional implications of its two key elements:

1.   A theory of number development based on research from around the world and described in a Learning Framework In Number [LFIN] (NSW DET, 2007); and

   2.   An individualised Schedule for Early Number Assessment [SENA]
        (Wright, Martland, & Stafford, 2006), used by teachers to place each
        child at a point within the Learning Framework.
Each of these elements is now briefly discussed.

Learning Framework In Number

The CMIT program focuses on teachers possessing a deep understanding of how
children learn mathematics by deepening their knowledge of the LFIN. Initially
developed by Wright (1994), the LFIN has since undergone further development
through the impact of a wide range of research in early number (e.g., Bobis, 1996;
Gravemeijer, 1994; McClain & Cobb, 2001; Mulligan & Mitchelmore, 1997).
Through its application in CMIT, the LFIN has been revised to reflect the gradual
inclusion of content drawn from higher grades, new insights from recent
research, and critical feedback from teachers, consultants and academic advisers.
In short, it provides a description of knowledge and skills characterising major
stages of development in each of five key components of number knowledge,
ranging from counting-by-one strategies for addition and subtraction through to
grouping strategies for determining place value, and multiplication and division.
Teachers use these stage descriptions to profile their students' knowledge in each
component and to provide instructional guidance as to what individual students
need to work towards. The LFIN is presented in Figure 2 and a more detailed
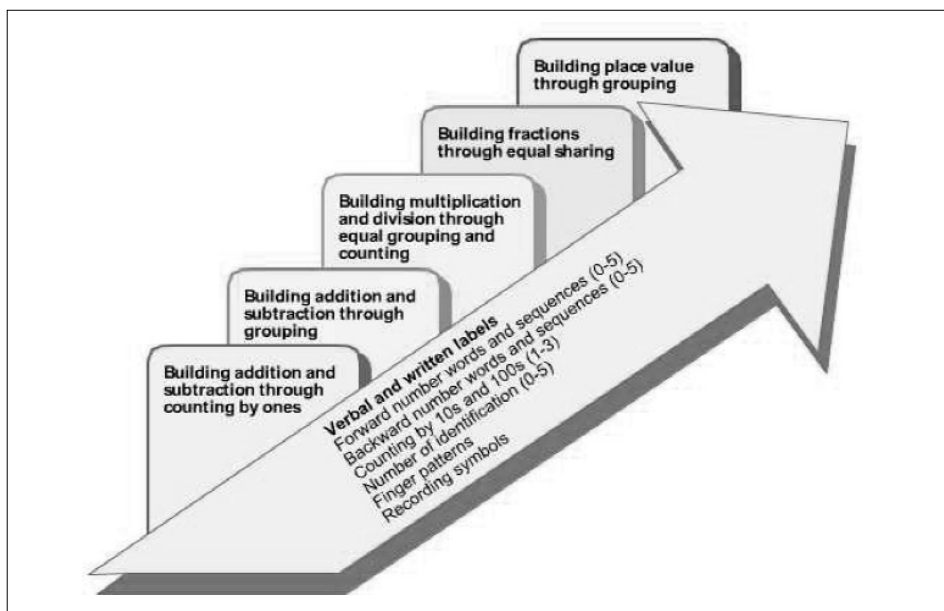description is provided in Appendix A.

Figure 2. Learning Framework In Number showing five major knowledge
components and important prerequisite number skills (NSW DET, 2007).

The Schedule for Early Number Assessment

The Schedule for Early Number Assessment (SENA) (Wright, Martland, & Stafford, 2006) is a task-based interview assessment instrument that is conducted individually with each child. Examples of tasks suitable for children in the early years of school include: asking the child to say the number before and after a given number that is less than 100; or given two covered collections of counters and asking the child how many counters altogether. More advanced tasks require children to count in multiples of tens and hundreds on and off the decade, to mentally compute answers to problems such as 43+21 and 50–27. During the interview, students are regularly asked to describe their solution strategies whether their answers are correct or not. It is the role of the interviewer to elicit a child's most sophisticated solution strategy and then to determine where each response might be categorised within the predetermined stages or levels of development described in the LFIN. The analysis of the SENA results and subsequent reflection of where each child is on the framework form the basis for teachers' instructional practices.

Context For The Studies

While CMIT started with a sole focus on number learning in the first three years of school (K-2), it has expanded to include K-6 and into other aspects of mathematics learning, such as measurement and geometry. In each case, the focus has been on using a synthesis of the research to provide clear direction for teaching and learning.

Prior to 2004, CMIT was introduced to teachers and maintained by one of 40 DET mathematics consultants. Consultants worked with a designated number of schools over a period of approximately 10 to 20 weeks to develop teachers' understanding of the LFIN and assessment schedule. This support was often at the individual classroom level. In 2004 system-level administration and structural changes necessitated the introduction of 62 facilitators located within schools where CMIT was already operating. Their roles were (and continue to be) specifically to assist schools sustain the program by providing school-based support. While the 40 mathematics consultant positions were retained, their main roles moved to providing regional-level support to each of the facilitators.

The exact nature of each facilitator's role was dependent upon their school's needs and its implementation model. Diversity of schools' needs meant that there could be no one model or explanation of the facilitator's role. It could involve, for example, organising professional discussion and planning meetings for the team, distributing material, providing in-class support, organising demonstration lessons or ensuring assessment procedures are in place. An aim was to build a community of practice by working with a team of teachers to establish and achieve a specified action plan for the expansion of CMIT. Hence, the school-based facilitator is seen as a person not responsible for doing all the work but helping to drive the process.

By 2005 the model expanded to include 87 facilitators operating in primary schools across NSW. The facilitator model remains the main source of in-school support even today and is aimed at driving the process of sustaining and expanding CMIT from *within* schools by firmly embedding knowledge of the LFIN and the associated core principles into the culture of each school. By 2010, two more tiers of facilitation were operating, involving district-level facilitators and a system-level on-line support mechanism.

Information determining the effectiveness of CMIT as it expanded was collected during three separate studies that successively 'drilled down' to explore specific aspects of the program, including facilitators' and teachers' knowledge of the LFIN and its impact on their learning and instruction (Bobis, 2004, 2006, 2009). The findings of these studies were re-examined to reveal aspects perceived to support the sustainability and scale-up of the program. These aspects were then analysed in terms of Coburn's multi-dimensional conceptualisation of scale. An overview of these studies is now briefly presented, detailing the data sources from which the findings and analyses were derived.

Overview of the Studies

During 2004 data investigating the effectiveness of the school-based facilitator model were collected via surveys returned from 58 facilitators and 23 principals of schools where the model operated. Surveys were designed to collect information about facilitators' mathematical content and pedagogical content knowledge, specifically their knowledge and experience of the LFIN and assessment schedule. It also sought to evaluate various aspects of their role from their own perspective and that of the principal. Additionally, surveys returned from 35 consultants provided insight into the nature and level of support consultants gave to their associated facilitators and added another perspective of the perceived success of the facilitator model. While descriptive statistics were used to analyse the items on the surveys requiring quantitative responses, items requiring qualitative comments were analysed according to emergent themes or according to a rubric when it was necessary to distinguish degrees or levels inherent in respondents' comments. As part of supporting documentation, 55 facilitators submitted their action plans for the implementation and expansion of CMIT in their own schools. Plans were analysed for clarity of goals and the types of activities undertaken to achieve those goals.

A facilitator's success was gauged by the degree to which the program had expanded in the school and/or had met the goals inherent in the school's action plan. 'Expansion' was interpreted to refer to the program's spread to other grade levels in the school and its influence on parents, school policies and routines. Schools in which the model was considered not to be operating successfully were perceived to have no clear goals or a plan of how to implement the program, and received little support from the executive staff of the school. The quality of the facilitator, in terms of their knowledge of the LFIN and their ability to mentor colleagues towards a deeper understanding of the professional development

program, was seen to be a key factor to the success of the model. Other factors perceived by facilitators and principals as contributing to the effectiveness of the school-based facilitator model included the availability of time for the facilitator to work specifically on program support activities, the clarity of the action plan, support from the principal and/or other executives in the school, and a collaborative working relationship among team members. A large proportion of principals and facilitators specifically mentioned the benefit of having the facilitator based within their own school.

In 2005/6, case studies incorporating document analysis, interviews and classroom observations in four schools where the facilitator model was operating were conducted. The aim was to more closely explore the factors affecting sustainability emerging from findings of previous evaluations of CMIT. Schools were purposively selected based on the clarity of their action plans detailing expected outcomes and processes for sustaining the reforms in their school. Qualitative analysis of the data revealed a number of common factors that were considered most influential to the sustainability of CMIT across all four case study schools. Once again, the facilitator emerged as the "driving-force" of the program. A follow-up study conducted in 2008 also utilised survey data and case studies of a further three schools to investigate the extent to which the program's key elements (LFIN and SENA) were embedded in practices at school, class and community levels. The case study schools were purposively selected based on their involvement in CMIT being greater than four years in duration and the opinions of district-level facilitators that the program was "well integrated with the school cultures".

The survey utilised 'descriptors' of various aspects of CMIT and required teachers to rate the level they perceived best described the implementation of CMIT at their school, their community and in their own classrooms. These descriptors were adopted from an existing planning matrix that is used by schools to self-monitor the depth of program implementation. It describes four levels of CMIT implementation for various aspects of the program as perceived by NSW DET and derived from research literature surrounding effective reform development. For instance, for whole school management aspects, a school where CMIT is perceived to be operating at a 'deep' level (Level 1) is described in the matrix as having school targets and long-term goals for CMIT that are funded and embedded within the school three-year action plan. Alternatively, a school considered to be implementing the program at a very superficial level (Level 4) is described as making no mention of CMIT in the school action plans despite the fact that certain elements of the program might be operating in some classrooms.

The final section of the survey required an open-ended response to a scenario involving a description of a student's reaction to a mathematical task. Teachers were asked to use the available evidence to approximate the child's performance as described by the LFIN and to make suggestions about the types of learning experiences they considered would most suit the child's level of understanding. Follow-up interviews further probed individual teachers' depth

of understanding of the LFIN as they were asked to elaborate on the mathematical development of particular children from their classes and to explain how they adjusted their instruction as a consequence.

The self-report survey ratings were compared to scenario responses, interview information and data from school documents such as management plans and teaching programs. The data were used to determine the amount of consistency between teachers' perceptions of the degree to which CMIT was operating at their school and the degree to which this was actually evident in the teacher responses and school documents.

Survey respondent self-ratings were reported to be consistent with other data sources. For example, ratings by respondents from one school indicated that there was a fairly robust implementation of CMIT for most aspects at their school as determined by their selection of Level 1 descriptors on the survey. This was further supported by evidence from interviewees who described practices and provided documented evidence that such practices were indeed occurring at their school and at times firmly integrated into the school culture.

Drawing together the findings of the three studies, a number of recurring aspects emerge as affecting the sustainability of CMIT at the school and system levels. The most influential ones are listed below and have been categorised according to Coburn's (2003) four necessary conditions for effective reform scale-up:

Depth

- Facilitators who were considered to have deep knowledge of the LFIN and of how children learn mathematics;
- Teachers who possessed deep understanding of the LFIN and the SENA; and
- Instruction/programming that was guided by students' developmental levels as revealed by ongoing assessments and the LFIN.

Sustainability

- Facilitators based in schools;
- Facilitators who were perceived by teachers, consultants and principals to possess highly effective leadership and organisational skills;
- School-level action plans that contained clear goals and strategies for the implementation and expansion of the program;
- Action plans that were aligned with existing school and regional management plans;
- The routine occurrence of whole-staff and grade-level meetings with a focus on mathematics;
- A high degree of internal support from senior executive staff; and
- Regional-level support from consultants was regular, collaborative, practical and easily accessible.

Spread of norms, principles and beliefs

- A focus on integrating elements of the CMIT program with the mandated curriculum rather than treating them as separate; and
- The school kept parents informed about reform practices via information sessions and letters home.

Shift in reform ownership

- Staff who shared the responsibility for implementing the action plans;
- A high degree of collegiality amongst staff. For instance, grade-level planning and programming for mathematics occurred where teachers shared resources and teaching ideas; and
- Parents were encouraged to provide support in classrooms during mathematics lessons, particularly at the lower grades.

Foremost among all these aspects was the fact that facilitators were based in schools and perceived to possess deep knowledge of the LFIN. Importantly, Coburn refers to the four conditions as interrelated. Hence, despite each aspect being allocated to just one category, in reality, they may overlap into a number of conditions depending on the context in which they occur.

Discussion

This section revisits each of the conditions contained in Coburn's (2003) conceptual framework for reform scale-up and discusses them in terms of the findings from this series of studies.

Depth: According to Coburn's conceptualisation of scale, reforms can be implemented at varying degrees. For consequential change to occur, that is, change that is self-generative and provides improved learning outcomes for students, it must go beyond surface level features such as simply using new resources in classrooms. Deep change requires teachers and groups of teachers to modify their beliefs, norms, knowledge and pedagogical principles as enacted in the classroom. The depth to which CMIT reform principles impacted on classroom practice or were embedded into existing school procedures and policies was explored in each of the three studies examined. Twenty-six percent of principals surveyed in 2004 about the effectiveness of the facilitator considered reform implementation to be more effective when accompanied by changes in organisational structures and routines that were more closely aligned to the program. For instance, principals reported changes to the way teachers prepared their programs with the institutionalisation of regular grade-level planning meetings. Assessment practices changed to include "ongoing assessment of students' performances rather than episodes of assessment" (Bobis, 2004, p. 20) based purely on end of topic tests. Programs and teaching plans were more focused on student learning needs as highlighted by SENA results and other forms of assessment. As stated by a Year 2 teacher in a study exploring the impact of the LFIN:

> The LFIN is the reason for the choice of activities. We use the results of the SENA testing to determine how we group the children initially and what activities go into the program—that's why I have the results of the SENA and my observations at the front of my program.

While the accessibility of resources emerged from studies as an important element for the successful implementation of CMIT, it was also found that when there was an over-emphasis on resource production or the provision of activities without an equivalent focus on understanding how specific activities and resources were linked to the LFIN and the overall mathematical development of children, CMIT was less likely to be sustained and facilitators were perceived to be less effective in their roles. Consistent with this finding, Coburn emphasises the importance of going beyond "the presence or absence of specific materials or tasks" (p. 5) as an indication of how deeply a reform is being implemented. Instead, she urges researchers to focus on the underlying pedagogical principles of teachers and how they require students to engage with various resources and activities. Case studies of schools where CMIT was determined to have a deeper implementation found that significant changes were made to the structure of teaching programs and to the processes by which they were developed. Such changes included closer alignment with the LFIN and a collaborative team approach to their construction and to the planning of individual lessons. Teachers reported changes to their classroom practices, especially how they deliberately planned and successfully differentiated their instruction for the varying abilities of students. One teacher explained how her instruction changed as a result of her deeper understanding of the LFIN: "We explicitly teach concepts. I tell them why we are doing the lesson and … what I expect by the end of the lesson. I use the same activity but I make adaptations to suit their different learning levels".

The increased alignment of teaching programs to the LFIN and the associated changes to classroom practices are closely linked to the depth of facilitators' and teachers' knowledge of the LFIN. Overwhelmingly, the success of facilitators was found to be heavily dependent on their own depth of understanding about the LFIN. This depth of knowledge increased their confidence and enabled them to provide greater support to the teachers in their schools. Principals considered that the increased focus on understanding children's mathematical strategy development "has allowed staff to gain a greater insight into student learning".

Study after study confirmed that the LFIN is at the core of the CMIT program and that deep knowledge of it is necessary for the program's sustainability. Surveys incorporating the CMIT planning matrix as a tool for determining the depth of reform implementation at the school and classroom levels found a high degree of consensus amongst staff at the same school when they were asked to rate their own and their colleagues' understandings of the LFIN. Follow-up scenarios and interviews requiring teachers to use their knowledge of the LFIN to interpret students' thinking strategies and discuss appropriate instruction for such students were consistent with teachers'

perceptions of their understanding of the LFIN and the degree to which it impacted on their teaching. Additionally, there was a close relationship between the overall depth of teacher knowledge of the LFIN and the likelihood of a school's ability to expand the program to a greater number of grade levels. Two important phases in a teacher's career were identified as times when exposure to CMIT reform practices could be a most powerful influence on their pedagogy: the first few years of teaching, and when a teacher transitions to a grade that they have previously not taught, particularly when it was a grade very different to the one they just left. In both instances, teachers reported actively seeking guidance for classroom instruction. In grade levels or schools where the LFIN was a strong influence and there was a culture of collaboration and support, the LFIN was more easily embedded into their programs and practices. A teacher in her third year of teaching at a case study school with a strong commitment to CMIT reported that the LFIN "is the only way I know how to teach…".

Sustainability: The degree to which particular reform features are embedded into school and system structures is a firm indicator of its sustainability—depth is a major factor in determining sustainability of reforms. According to Coburn's conceptualisation of scale, deep understanding of reform principles is needed for reform practices to be sustained over time and particularly in times of difficulty. Unless reforms can be sustained, they cannot be scaled up. It is therefore imperative that designers build into reforms capacity building and support mechanisms or tools at various levels of implementation (Coburn, 2003; Grossman et al., 1999). Fullan (2009, p. 102) discusses the need for "tri-level" support for system-wide reforms—at the state (or possibly Federal level), district level, and the school and community level. As CMIT has been scaled up, a number of changes have been introduced to support the program's sustainability and spread, while maintaining the integrity of its core elements—namely, the LFIN. These changes have not always occurred in discrete stages, nor were they always anticipated at the start of the program. McLaughlin and Mitra (2001) suggest that theory-based reforms evolve through a process of invention and implementation that often co-occur while the reform itself is sustained and expanded. In the case of CMIT, changes have mostly occurred in incremental steps as a result of systematic reviews providing recommendations and the existence of an effective feedback loop between system-level stakeholders and those at the school and class levels. At other times, external forces have necessitated major changes to the way the program has been supported.

The evidence base upon which this paper drew, revealed a range of support mechanisms at various levels of CMIT, some of which only emerged towards the end of the period in question. Central among changes at the school/class level was the introduction of the school-based facilitator model. Findings from all three studies concur that effective facilitators are those who are perceived to have deep knowledge of the LFIN, practical ideas due to their extensive experience as classroom teachers and as teachers who have successfully implemented the reform in their own classrooms. They were also described as being flexible with their planning, well organised and possessing a good rapport with staff and

children. In short, effective facilitators were viewed as leaders in mathematics at the school level. As one principal stated: "the quality of the facilitator is the key to the success of the model".

Importantly, while facilitators were (and remain) central to the on-going success of CMIT in schools, they are not the only support mechanism at this level. Once teachers sufficiently develop their understanding of the LFIN, it too was shown to be a powerful conceptual tool for teachers – and one that has been found to influence instructional decision-making at all grade levels. In each of the studies examined, teachers in schools where CMIT was firmly embedded utilised the LFIN on a daily basis. For some teachers, the framework was considered to provide greater detail about specific strategies and a "greater sense of direction" for teaching. During interviews conducted for two separate studies, teachers who had implemented CMIT for a number of years expressed the desire to learn more about the "theory behind the LFIN" as they felt it would help deepen their understanding even further. Many teachers reported that they kept a copy of the LFIN at the front of their programs to allow easy reference when programming. At one school, a teacher commented:

> We all have the *Framework In Number* on a table or up on the wall in the classroom as a constant reminder as to where you're taking these kids next … you can see where they have been and where they have to go.

Additional school-level support mechanisms that emerged across the studies included the *Developing Efficient Numeracy Strategies 1 and 2* (NSW DET, 1999, 2003) teacher resource books and the assessment instruments or SENA. Both these mechanisms are akin to the "practical tools" described by Grossman et al. (1999, p. 14). The conceptual understanding of the LFIN guides teachers in their use of SENA results for individual and groups of students and the teacher resource books further support classroom implementation of the LFIN via practical activities.

Another practical tool that repeatedly emerged from the findings was the CMIT action plans. Action plans varied enormously as they were dependent on schools' needs. While no one plan could be highlighted as an exemplar, those submitted from facilitators from schools where CMIT was more firmly embedded, contained certain commonalities. Such commonalities included a clear focus or goal, a variety of actions that progressed through a cycle of training, planning for implementation, in-class support, evaluation (via some form of data collection and reflection) and forward planning. For example, the action plan of one school where CMIT had been operating in K-3 classes for three years began the year with a focus on re-educating parents and training for new staff. The focus then shifted to an examination of whole-school assessment practices, grade-level programming based on the assessment data, in-class support and then to forward planning for the expansion of CMIT to the higher grades in the school. Throughout each stage of the plan professional dialogue and evaluation via feedback from both students and staff was emphasised. Throughout each of these activities, the focus remained on one area—

assessment—but included elements that would not only help teachers implement new practices, but would also assist with their sustained implementation via continued professional dialogue and collaboratively devised whole-school procedures for assessment.

At the regional level, consultants have provided facilitators with support since the introduction of the facilitator model. Survey results in 2004 indicated that nearly 42% of facilitators at that time considered the level of support they received from consultants to be "better than adequate". Interview data collected in subsequent studies revealed that the most effective support from consultants was typically described as "regular, collaborative, practical and easily accessible".

System-level support mechanisms were rarely mentioned in data from any of the studies examined over the five-year period 2004 to 2008. This is partly due to the fact that it was not specifically targeted for exploration in these studies. However, the effectiveness of CMIT Online has been the subject of a separate evaluation (Cavanagh, 2006). While largely successful, the evaluation revealed most teachers did not take full advantage of the ability to share knowledge and experiences of CMIT with teachers from other schools via the discussion site. However, a positive outcome of the online program was the opportunity it provided for collegial support to develop between teachers within schools as they worked together in school teams.

Spread: Coburn's (2003) conceptualisation of scaling-up reforms retains the traditional meaning of spread, referring to increasing numbers of sites where the reform is implemented. However, her insistence that spread must go beyond superficial aspects of a reform to also include beliefs, norms and principles means that spread should also be considered in terms of a reforms influence on "policies, procedures and professional development" at the school and regional levels (p. 7). In this sense, a reform is considered to have spread even *within* a school where it has been implemented for a number of years, because it is more deeply embedded (or institutionalised) in the school documents and daily routines. A key role of the school-based facilitator was to embed aspects of CMIT into the school culture. Interview data reveal that at schools where CMIT was most influential, teachers shared a professional dialogue to describe children's cognitive development of mathematics based on the language of the LFIN. This was achieved more often in schools where regular grade-level planning meetings were held and whole-school staff meetings focused on student achievement in mathematics. A number of schools had instituted classroom observational visits between teachers of different grades followed by professional conversations to 'de-brief'. Teachers in these schools often commented that such opportunities allowed for deeper conversations about pedagogy and reinforced their collaborative planning efforts. Generally, it was found that such practices worked best in the lower grades, where CMIT had been operating in schools longest. However, in case study schools with principals who were also highly informed and supportive of the program, there were plans to spread the practice to more grades.

Coburn (2003, p. 7) laments that studies taking more traditional views of reform spread have not considered "spread within" schools as a dimension to be explored. School management plans (as distinct from the CMIT action plans) were another element linked to a school's capacity for broadening and deepening the program's influence within the school that were examined in two separate studies reviewed for this paper. School management plans that incorporated budgetary allowances for CMIT resources and staff training provided explicit evidence that the sustainability and expansion of the program was a high priority for school leadership teams in some schools. Teachers at schools where CMIT was perceived as a priority in management plans and well supported by principals and other executive were adamant that "the school is really committed to the program, and you know that it's here to stay".

Another dimension to spread that has not figured greatly in previous studies of reform expansion, is the involvement of the school community. Many of the CMIT action plans referred to "re-educating" parents. One principal commented that the intention of communicating and educating the parents about the principles behind and benefits of the program was so that they too would learn to "value and support the program". It was felt that with community support, the program would become more deeply embedded into the school's culture. The extent to which the sustainability and scale-up of CMIT can be impacted upon by greater community support and involvement is an area for future investigations of the program by its developers.

Shift in reform ownership: There is significant agreement among researchers that for a reform to be considered 'at scale' there must be a shift in reform ownership to the teachers, schools and regions enacting the reform (Coburn, 2003; Fullan, 2009; McLaughlin & Mitra, 2001). It is only through this shift in ownership that a reform becomes truly self-generative (McLaughlin & Mitra, 2001). Foremost among such capacity-building strategies in the CMIT program has been the shift of authority and knowledge to school-based facilitators and the school management team in terms of the best way to implement and deepen the program's influence. Importantly, schools where the facilitator model worked most effectively to expand the program, responsibility for enacting the CMIT action plan was shared amongst staff, with the facilitator viewed as the "driving-force". As one consultant remarked when reflecting on the benefits of the facilitator model:

> I think the idea of having a school-based facilitator is great! It enables schools to take greater control of their own professional learning and gives some teachers a chance to try some leadership.

While the school-based facilitator model repeatedly emerged in studies as a determining factor for the sustained success of the program, Fullan's (2009, p. 108) warning against endorsing "*one factor at a time as key*" is worth repeating. Even with an effective facilitator possessing all the attributes considered necessary for success, it is evident from our previous discussions that other conditions play major influential roles. These conditions include the presence of

structures for on-going teacher learning about the reform, such as whole-school professional development, grade-level planning meetings, and classroom observation visits. Another condition necessary for reform ownership to shift to teachers is the existence of school-level strategies that provide continued funding for CMIT activities. To this end, CMIT was more likely to be sustained in schools where it was a priority in management plans and routinely factored into school funding arrangements.

Conclusion

The findings of this linked sequence of studies and their subsequent re-examination in the light of Coburn's conceptualisation of scale have implications for sustaining change in schools that have moved beyond the initial implementation phase of professional development programs, particularly when intensive external support ceases. Namely, sustainability is dependent on the interaction of a variety of elements – no one element can be highlighted as singularly influential. These elements are not normally definitive instructional practices, but include a deep knowledge-base of appropriate instructional practices derived from children's developmental needs, a coherent school plan that clearly articulates shared goals and establishes firm structures for on-going teacher learning, and the provision of time and institutionalised structures for these shared goals to be realised. As outlined by Coburn (2003, p. 6), these elements are more effective at sustaining reform when "there are mechanisms in place at multiple levels of the system". In the case of CMIT, the school-based facilitator works as a mechanism to nurture deep knowledge of children's developmental thinking strategies and the consequential changes to instruction required by increasing teachers' understanding of the LFIN. Importantly, the facilitator model is only one of a number of mechanisms acting to sustain and scale up the professional development reform principles over time.

References

Bobis, J., (1996). Visualisation and the development of number sense with kindergarten children. In J. Mulligan, & M. Mitchelmore (Eds.), *Children's Number Learning* (pp. 17-33). Adelaide: Australian Association of Mathematics Teachers.

Bobis, J. (2001). *The effect of Count Me In Too on Year 3 Basic Skills Test results*. Sydney: New South Wales Department of Education and Training.

Bobis, J. (2004). *Count Me In Too: An evaluation of the facilitator model*. Sydney: New South Wales Department of Education and Training.

Bobis, J. (2006). *Factors affecting the sustainability and expansion of Count Me In Too*. Sydney: New South Wales Department of Education and Training.

Bobis, J. (2009). *Report of the CMIT Project 2008*. Sydney: New South Wales Department of Education and Training.

Bobis, J., Clarke, B., Clarke, D., Thomas, G., Wright, R., Young-Loveridge, J., & Gould, P. (2005). Supporting teachers in the development of young children's mathematical thinking: Three large scale cases. *Mathematics Education Research Journal, 16*(3), 27-57.

Cavanagh, M. (2006). *Count Me In Too online professional development program*. Sydney: New South Wales Department of Education and Training.

Clarke, B., & Clarke, D. (1998). Purposes, principles and practices in the professional development of mathematics teachers. In N. Ellerton (Ed.), *Current issues in mathematics education* (pp.1-17). Perth: MASTEC.

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3-12.

Coburn, C. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy, 19*(3), 476-509.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk, 5*(1–2), 183–204.

Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change, 10*, 101-113.

Garet, M., Porter, D., Desimone, L., Birman, B., & Yeon, K. (2001). What makes professional development effective: Results from a national sample of teachers. *American Educational Research Journal, 38*(40), 915-945.

Gravemeijer, K. P. E. (1994). *Developing realistic mathematics education*. Utrecht, The Netherlands: CD-B Press.

Grossman, P., Smargorinsky, P., & Valencia, S. (1999). Appropriating tools for teaching English: A theoretical framework for research on learning to teach. *American Journal of Education, 108*, 1-28.

Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks: Corwin Press.

McClain, K., & Cobb, P. (2001). An analysis of development of sociomathematical norms in one first-grade classroom. *Journal for Research in Mathematics Education. 32*(3), 236-266.

McLaughlin, M., & Mitra, D. (2001). Theory-based change and change-based theory: Going deeper, going broader. *Journal of Educational Change, 2*, 301-323.

Mulligan, J. T., & Mitchelmore, M. C. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education, 28*, 309-330.

New South Wales Department of Education and Training. (1999). *Developing efficient numeracy strategies Early Stage 1 and Stage 1*. Sydney: Author.

New South Wales Department of Education and Training. (2003). *Developing efficient numeracy strategies Stage 2*. Sydney: Author.

New South Wales Department of Education and Training. (2007). *Count Me In Too. Learning framework in number*. Sydney: Author.

Putnam, R., & Borko, H., (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Stringfield, S., & Datnow, A. (1998). Introduction: Scaling up school restructuring designs in urban schools. *Education and Urban Society, 30*, 269-279.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Wellington: Ministry of Education.

Whitcomb, J., Borko, H., & Liston, D. (2009). Growing talent: Promising professional development models and practices. *Journal of Teacher Education, 60*, 207-212.

White, P., & Mitchelmore, M. (2002). *The impact of Count Me in Too on Basic Skills Test numeracy scores: 2001-2002 follow-up report*, Sydney: New South Wales Department of Education and Training.

Wright, R. J. (1994). Working with teachers to advance the arithmetical knowledge of low-attaining 6- and 7-year-olds: First year results. In J. Ponte, & J. Matos (Eds.), *Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education, Vol. 4*, (pp. 377-384). Lisbon, Portugal: PME.

Wright, R. J., Martland, J., & Stafford, A. (2006). *Early numeracy: Assessment for teaching and intervention*. London: Paul Chapman.

Authors

Janette Bobis, Faculty of Education and Social Work, The University of Sydney, Sydney, Australia. Email: <janette.bobis@sydney.edu.au>

Appendix A: The Learning Framework in Number

1. Building addition and subtraction through counting by ones

Stage	Name of Stage	Description of Strategy
0	Emergent	Cannot count visible items
1	Perceptual	Can count visible items only.
2	Figurative	Can count invisible items, but starts from one.
3	Counting-on	Can count invisible items, using a counting-on strategy to solve addition or missing addend tasks, and may use a counting-back strategy to solve missing subtrahend or removed items tasks.
4	Facile	Can use non-counting-by-one strategies, such as doubles, add through ten, compensation, etc.

2. Building addition and subtraction through grouping

Combining and Partitioning

Level	Name of Level	Description of Strategy
1	To 10	Knows number combinations to 10 and how many more to make 10.
2	To 20	Can provide standard and non-standard partitioning of a number to 20.

Subitising

0	Emergent	Unable to instantly recognise the total number of objects for small groups.
1	Perceptual	Can instantly recognise the total number of up to 6 objects arranged in common dice patterns.
2	Conceptual	Able to recognise combinations of domino patterns (e.g. 4 and 5 dots) as both two groups and as "a whole".

3. Building place value through grouping

Level	Name of Level	Description of Strategy
1	Ten as a count	Not able to see ten as a unit composed of ten ones. The child solves tens and ones tasks using a counting-on or counting-back strategy.

| 2 | Ten as a unit | Able to see ten as a unit composed of ten ones. The child can count on and back by tens on and off the decade. Child is dependent on representations of units of ten. |
| 3 | Tens and ones | Able to mentally solve 2 digit addition and subtraction tasks by adding and/or subtracting units of ten and ones. Does not need materials or representations. |

4. Building multiplication and division through grouping and counting

Level	Name of Level	Description of Strategy
1	Perceptual counting by ones	Does not see equal groups as composite units and thus counts each item by ones.
2	Perceptual counting in multiples	Uses groups or multiples in perceptual counting and sharing e.g. rhythmic or skip counting.
3	Repeated numerical composites	Equal grouping and counting without individual items visible. Uses group markers or organisers.
4	Repeated abstract composites	Constructs composites and coordinates the count. Uses repeated addition and subtraction.
5	Multiplication and division as operations	Uses known facts and inverse operations to derive answers.

5. Building fractions through equal sharing

Quantity Fractions

Level	Name of Level	Description of Strategy
1	Halving	Uses halving to create the 2-partition and the 4-partition. Interpreting halves and quarters formed.
2	Odd partitions	Creates measureable unit fractions without repeated halving. Can create thirds by subdividing a given whole.
3	Re-unitising	Carries out measurement division involving a unit fraction. Can iterate a fraction part beyond the whole and reform the unit whole.
4	Quantity fractions using the equal whole	Developed a sense of the universal equal whole and can use multiplicative structures to create equivalence.
5	Coordinating composition of partitioning	Can reverse the process of creating a fractional part.