

Chapter 61

Metaphor and Theory for Scale-up Research: Eagles in the Anacostia and Activity Systems

Sharon J. Lynch

As a researcher studying the scale-up of science curriculum units in middle schools, I was startled awake early one morning (c. 2004) when I heard the word ‘scale-up’ issuing from the National Public Radio news on the clock radio. In an interview about a worrisome impending winter flu epidemic and vaccine shortages, a reporter and a pharmaceutical researcher discussed the need to *scale up* vaccine production. Not only was much *more* of the stuff needed, but the industry also had to find *better ways* to produce it in quantities that could meet the rising demand from a concerned public. Another problem was *how to distribute* the vaccine to those who needed it most. Fully alert, I pondered the application of flu vaccine scale-up metaphor to issues facing my research team in studying the scale-up of science units with a partner school district. Our studies were designed to determine the interventions’ effectiveness at small scale, and explore if they could be taken to large scale without diluting their impact on student learning. As with the flu virus, both ‘production scale’ and ‘distribution’ of the curriculum units were mettlesome problems.

The field struggles to define, describe and understand the scale-up of interventions in education. The ultimate goal is to improve education by stimulating large-scale adoptions of interventions having strong evidence of effectiveness. My research programme was stimulated by funding from the Interagency Educational Research Initiative (IERI) whose goal is to ‘increase the knowledge of *scaling up* by supporting research that investigates the effectiveness of educational interventions ...[and]...requires ...understanding of the learning outcomes related to specific educational interventions with a rigorous analysis of the logistical, organizational, political, and economic factors that facilitate or impede [scale-up]’ (National Science Foundation 2002, p. i). IERI funded 101 educational research projects, about evenly distributed among reading, science and mathematics.

S.J. Lynch (✉)
Graduate School of Education and Human Development,
The George Washington University, Washington, DC, USA
e-mail: slynch@gwu.edu

The study of scale-up in education has been relatively untrammelled territory and is admittedly under-theorised (Coburn 2003; McDonald et al. 2006). In the fall of 2003, a group of IERI researchers and theorists met to discuss scale-up of educational interventions. Experts from fields outside education described scale-up from perspectives such as economics, technology, computer science, sociology, engineering, statistics, psychology and organisational behaviour. By exploring analogues in other disciplines, educational researchers perhaps would develop theories to explain the scale-up of interventions in school systems (cf. Schneider and McDonald 2007a, b).

Theorising Scale-up/Scale-up Research

Normative Dimensions for Outcomes of Scale-up

At the same time when the IERI conference was taking place, Cynthia Coburn (2003) framed her views on scale-up by developing criteria to guide the study of scale-up of whole-school reform interventions. She suggested that traditional definitions of scale-up (the deliberate expansion to many settings of an externally developed school restructuring design that has previously been used successfully in one or a small number of school settings) are too limiting and would not capture the normative aspects of scale-up in education settings. Coburn reconceptualised scale-up outcomes to include four interrelated dimensions:

- *Depth* (the reform must affect a deep and lasting change in classroom practice)
- *Sustainability* (it must last within the school or school district or continue to scale, even after start-up funding has run out)
- *Spread* (the intervention must include not only the spread of activity structures, materials and classroom organisation, but also spread of underlying beliefs, norms and principles to additional classrooms and schools)
- *Shift in reform ownership* (the reform is no longer external to the school and controlled by the reformer, but internal with the shift of authority and knowledge to teachers, schools and districts)

Coburn focused on scale-up from the standpoint of programmatic or normative outcomes of scale-up for schools or larger education entities. Her policy perspective applied to school systems, but was stimulated by her in-depth case study of a single elementary school that had participated in the Child Development Project (CDP), a whole-school reform programme for elementary schools. Coburn's study occurred the year *after* outside funding had ended and focused on CDP's ability to stick in this school after CDP researchers had withdrawn. She suggested that that scale-up theory ought to be about more than just numbers; it should also attend to the four dimensions listed above. Given the substantial human effort and financial

costs of whole-school reform efforts, questions of ‘worth’ arise. Coburn’s four normative dimensions for scale-up might be interpreted as criteria for determining the value of a scaled-up intervention over the long term, raising non-trivial questions about interventions intended for scale-up and their costs, consequences and long-term worth.

Scaling-up Exemplary Interventions

Sarah Kay McDonald and other members of the Data Research and Development Center (DRDC) at the University of Chicago provide a different view of scale-up research and theory based upon their unique position as a knowledge-building group charged with the management and dissemination of results from the IERI scale-up research portfolio (McDonald et al. 2006). McDonald et al. theorise about scale-up from the frame of university researchers focused on the knowledge that emerged from the IERI research portfolio, rather than that of researchers who worked directly with schools participating in such projects. McDonald et al.’s view is that scale-up research is primarily about numbers – valid and reliable data from studies constructed to be generalisable to increasingly large and varied contexts. They define scale-up as the practice of introducing proven interventions to new settings with the goal of producing similarly positive effects in larger, more diverse populations. Scale-up research examines factors that influence the effectiveness of interventions as they are brought to scale across settings. McDonald et al. partition scale-up and, correspondingly, scale-up research, into three stages. The goal of the first stage is to demonstrate that an intervention is *effective* and leads to improvements for students in a given set of circumstances. In the second stage, the goal is to determine if the intervention is *scalable*, spreading to more sites with varied contexts while maintaining its success. The third stage involves the ongoing evaluation of implementations (i.e. the intervention’s *sustainability and efficacy* across sites and over time). Scale-up research focuses on contextual factors necessary for success as an intervention scales. In contrast to Coburn, these authors believe that scale-up is inherently about size, numbers and doing more and about ‘extending the reach of an exemplary intervention to produce similarly positive effects in different settings to reach a greater number of students, teachers and setting’ (McDonald et al. 2006, p. 16). Research trials and comparisons allow generalisations about how and when to use the intervention in different contexts.

Although McDonald et al. and Coburn approach scale-up in different ways, both contribute to theory building. The intention for this chapter is provide a third perspective on scale-up theory which arises from participation in a 6-year research programme on the scaling-up of middle school science curriculum units in a large and diverse public school system. The name of this research programme is Scaling-up Curriculum for Achievement Learning and Equity Project (scale-up).

Scale-up in Context: Science Curriculum Units

An Overview of Scale-up

Sharon Lynch, Joel Kuipers, Curtis Pyke and Michael Szesze (2005) designed scale-up to study the systematic scale-up of three reform-based science curriculum units in middle school classrooms in Montgomery County Public Schools (MCPS), Maryland. MCPS is the 14th largest school district in the USA and one of the most diverse. Scale-up was completed in 2007 after reaching 6th and 8th graders in about 35 middle schools and 7th graders in 10 schools, including about 250,000 students and over 120 science teachers. Scale-up involved three different middle school science curriculum units of limited duration (3–10 weeks) that were created by three different research-oriented institutions. Each of the units has well-defined instructional characteristics that are thought to be important for student learning according to criteria developed by Sophia Kesidou and Jo Ellen Roseman and their working group at Project 2061 (2002). Although each unit had been field-tested prior to scale-up, none had been studied using a rigorous quasi-experimental methodology, combined with an extensive ethnographic component. Consequently, the effectiveness of each unit was an open question, as was how it functioned in classrooms (Lynch et al. 2007a).

Each unit focused on a different challenging science target idea (conservation of matter, reasons for the seasons, and motion and force) that research shows as challenging for children (and adults) to understand (cf AAAS 1993). The dependent variables for each unit studied were student outcomes scores on curriculum-independent assessments. Classrooms of students from five pairs of carefully matched schools were randomly assigned to a treatment or comparison condition for each science curriculum unit. The resulting samples mirrored the middle school population. Each unit was studied in this way for at least 2 consecutive years (Lynch 2008). If an intervention curriculum unit was effective both overall and when data were disaggregated (by ethnicity, or eligibility for Free and Reduced Meal Status (FARMS), English for Speakers of Other Languages (ESOL), or special education services), then it would be considered for scale-up to 35 middle schools in the district. Scale-up would also study how the unit functioned in a classroom from an ethnographic perspective (Kuipers et al. in press) and explore additional factors of ‘school experience’, ‘outcomes at large versus small scale’ (Watson et al. 2007) and ‘fidelity of implementation’ (Lynch 2008; O’Donnell et al. 2007).

Pragmatics of Scale-up Research

Equity issues were paramount to scale-up’s curriculum effectiveness studies. The intervention curriculum units had certain instructional strategies, congruent with Project 2061’s Curriculum Analysis (Kesidou and Roseman 2002). These units seemed more likely to be more effective than the business-as-usual curriculum

materials in use in the district, such as traditional science textbooks, other reform-based curriculum materials, Internet and video resources, and district-constructed curriculum guides. Scale-up would test the effectiveness of each intervention units both overall and when student outcome data were disaggregated by ethnicity or eligibility for FARMS, ESOL or special education services (Lynch 2000). If some subgroups of students were disadvantaged by an intervention unit, then certainly the decision to scale it up would be problematic. However, if a unit was more effective overall and passed the equity litmus test, then the goal was to study its movement to scale in the school district, exploring its potential for closing achievement gaps in the long term.

Scale-up was dependent upon four conditions for research collaboration that emerged prior to, or very early on in, the endeavour: (1) close partnership between researchers and school district educators; (2) recognition that the success of any intervention is determined by the pervasive policy climate of the powerful school system; (3) quality of assessment feedback and other information that both permit and drive scale-up decisions; and (4) well-organised research agenda to systematically introduce new curriculum units to teachers. The scale-up of the units was unlikely to occur, as past experiences in the partner school district had already shown (Lynch et al. 2007a), unless each condition was addressed as discussed in detail below.

Close Partnership Between Collaborators

At the outset, the scale-up university researchers and school district science coordinators and evaluators had to establish common goals for the study. Scale-up was viewed initially by school district leaders as a long-term, intensive and thorough curriculum evaluation in which the district itself could and should engage, although it might not always have the means at its disposal. Scale-up funding provided those means. The role of the university side of the collaboration was to: develop the research design, guided by organisational patterns that existed within the school system; analyse data collected in classrooms; and report results. The role of the school district was to: direct the professional development required by the intervention units; coordinate the study across schools; and collect data. Interpretation of results and decision-making as the study progressed involved both sides of the collaboration.

When results associated with a particular curriculum intervention were ambiguous (as they sometimes were), the entire process slowed and the research design was revisited. For instance, one of the treatment units unexpectedly proved to be less effective than the comparison condition for 2 consecutive years. As a consequence, the unit was not scaled-up (Lynch 2008). When another curriculum unit yielded ambiguous results both overall and when data were disaggregated for 2 years, Scale-up replicated the study with different schools and employed a much tighter research design in the third trial (Watson et al. 2007). Although this delayed the research agenda, eventually a collaborative decision was made to move to scale-up

the unit because the third trial yielded positive results under the more stringent design conditions.

McDonald et al. (2006) take a matter-of-fact approach to decision-making in scale-up: 'Each [scale-up study] produces an essentially dichotomous answer – either the intervention does or does not lead to an improvement in a given set of circumstances' (p. 16). However, scale-up's decision-making process was neither clearly dichotomous nor dispassionate. Unexpected results and the need to thoughtfully revisit aspects of research design resulted in extensive, careful analyses and decisions that focused on the best courses of action for students, the school district partner, and the research study. The decision to scale-up a curriculum unit or eliminate it was 'co-owned' by scale-up's school district and the university researchers.

In the era of No Child Left Behind (NCLB), school administrators might actively seek partnerships to test an intervention to remedy a specific problem if the intervention is believed to hold promise for meeting well-defined needs (Dahlkemper 2003; Lewis 2003; Daniel G. Bugler, personal communication, September, 2006). Although there might be instances when researchers approach a school district out of the blue to try an innovation, scale-up's experience suggests that school districts are unlikely to expend extensive resources without some assurance that improved student outcomes are likely to ensue, especially if a goal is to scale-up the intervention.

In research collaborations that involve scale-up, there is a group of goal-oriented early innovators including both researchers and school district educators. Other teachers are gradually introduced to the intervention and go along as it scales up. Still others remain estranged from the intervention for a variety of reasons, no matter when they encounter it. In this view, researchers and educators both are interveners in the scale-up environment. Over time, the intervention either takes root and is institutionalised, or it dies out. McDonald et al. might refer to this as a context-specific test of the efficacy of intervention at the level of the school district. When middle school science curriculum units or the like are studied as scaled-up research, then counting spread across schools makes sense, yet is too limited.

Scale-up experienced ongoing negotiation and problem-solving, bolstered by goodwill and considerable efforts to interpret results in a way that would result in a sound course of action for scale-up. An alternative view of scale-up suggests a reconsideration of what is actually being spread. When a school district adopts an intervention or restructuring model, in the longer view, information about the innovation moves through the system. Information includes student outcome data, changes in teacher beliefs, norms and principles, and declarative and procedural knowledge about the innovation itself. Information about the research goals, methods and outcomes is also distributed, including contextualised knowledge about the innovation's progress and who is involved, resistant or simply going with the flow.

During scale-up, middle school science teachers became involved with aspects of information flow in unanticipated ways. For instance, science teachers were asked to implement the intervention units with fidelity, while holding modifications to a minimum. Teachers asked, quite reasonably, about what exactly constituted fidelity of implementation and its parameters. Fidelity guidelines that were subse-

quently developed were as much a product of the science teachers' input as the researchers' (Lynch and O'Donnell 2005). For example, fidelity guidelines indicated that that teacher-developed homework and assessments not specified by the curriculum units were fine, but videos and Internet resources related to the units' concepts were not. Another example showed that, when one intervention unit proved to be less effective than the district's business-as-usual curriculum on the same topic, the teachers were sought out by the intervention unit's developers. Subsequent revisions of this unit were based in part on the teachers' input. Although this unit did not scale up, scale-up's partner-teachers' ideas will affect future versions of the unit in other school districts.

School Policy Climate

Principles, norms and beliefs are greatly affected by the school district's policy climate, which directs resources, professional development and incentives/disincentives to teachers. This sends a message about how the school district 'higher-ups' value a study. Scale-up's school district partner is large, influential and highly regarded. It is also becoming more diverse socio-economically, ethnically, culturally and linguistically. Predictably, because there are achievement gaps in science (and other subjects) between various demographic student subgroups, the school system must work hard to find new ways of reaching and teaching its increasingly diverse population in order to maintain its reputation. This was a priority of the strong and highly visible superintendent who has initiated several successful and highly publicised programmes to reduce achievement gaps (Weast 2000). This policy climate existed throughout scale-up and corresponded with the superintendent's mandate to improve student achievement for African American and Hispanic students. scale-up's goals aligned with school district policy goals, making it an opportune time for research on reducing science achievement gaps (Lynch et al. 2007b).

Even in a favourable policy climate, competing mandates created tricky crosswinds for both school district science educators and university researchers during scale-up. The study's initial goal was to compare the effectiveness of three different curriculum units with the business-as-usual curriculum materials. However, 2 years into the study, the district's central administration called for the development of new science curriculum guides aimed precisely at a new state curriculum framework. The district science office complied and formed teams of teachers to write the new middle school science guides. Thus, completely unanticipated, somewhat different comparison conditions were born during scale-up, a research programme that relied upon a series of quasi-experiments. Middle school science teachers in this district might be involved in implementing scale-up's curriculum units, or the development and implementation of the new locally constructed guides, or both. This messy situation shows how the school policy climate, while favourable for scale-up research on equity issues, could blow in new directives that are potentially confounding to the research design, which potentially could place science teachers and researchers in conflict. Although scale-up rode out these competing mandates, by

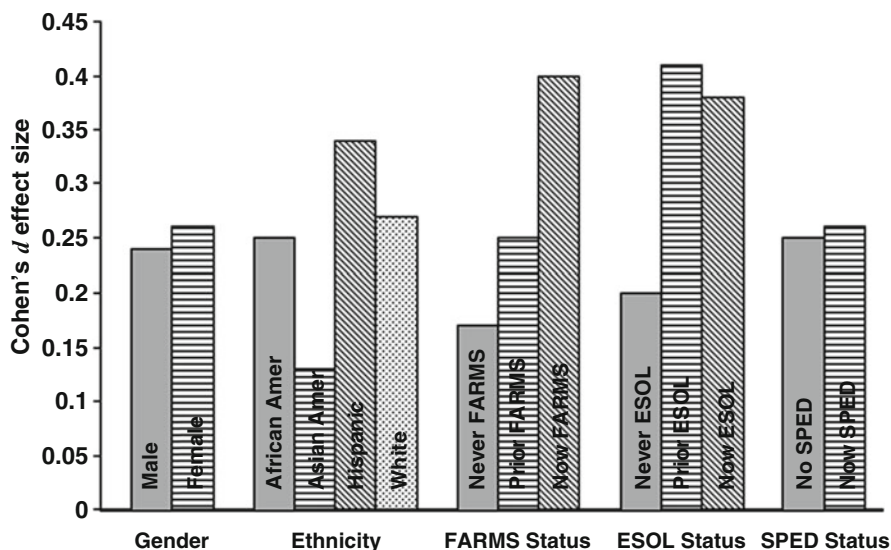


Fig. 61.1 Cohen's *d* effect sizes for *Chemistry That Applies* (2002–2003) according to gender, ethnic group or eligibility for Free and Reduced Meals Status (FARMS), English for Speakers of Other Languages (ESOL) or Special Education Services (SPED)

the end of the study, the comparison condition had been somewhat changed. It included the new locally produced guides, as well as the older menu of curricular options. Teachers' attachment to the new guides developed by local teams could affect the sustainability of scale-up's interventions after funding ended, despite the fact that two intervention units produced evidence indicating overall effectiveness and potential to close achievement gaps.

Assessment Feedback and Scale-up

Scale-up's initial effectiveness studies for the first curriculum unit, *Chemistry That Applies* (CTA) (State of Michigan 1993), showed significant mean differences on a curriculum-independent assessment of the target idea (conservation of matter) for 2 consecutive years. Disaggregated data showed that CTA was more successful than the business-as-usual curriculum materials for virtually all demographic subgroups of students (Lynch et al. 2005). This included student subgroups under-served in science education, such as students eligible for FARMS, ESOL or special education services, or African American or Hispanic students (Lynch et al. 2007b). CTA's overall effect size was 0.25 and ranged from 0.25 to 0.41 for under-served subgroups (see Fig. 61.1). Figure 61.1's representation of results made a compelling case to district science educators for CTA's scale-up. Measures of student engagement and goal orientation also pointed in a positive direction for the unit. Video data from classroom

observations created insights about how the curriculum unit was actually functioning in the classroom to help students to learn (Kuipers et al. in press).

Information about the CTA unit spread, emanating not only from teachers, administrators and researchers, but also from students who had received the intervention and their parents. If student responses were positive, then teachers also were more likely to respond positively. Crucial to the successful scale-up of an intervention, however, is capturing and reporting solid student outcome data early, rather than relying solely on impressions of the unit. Some teachers liked CTA, some found it repetitious and others thought that they could teach the target concept better without the unit. But 2 years of data indicated that, overall, CTA was more effective than the standard fare, especially for under-served subgroups. Thus, the reciprocal relationship between the intervention's spread and student responses to it seems obvious *if* researchers can make the results publically accessible in a timely fashion. This increased CTA's chance of going to scale.

Scale-up was deliberately designed not to be an accountability system that could link student outcome data to teachers or schools. Rather, the goal was to generalise to the entire school system, with disaggregated student data providing an evidence-based voice from student subgroups that might not ordinarily be heard when making curriculum decisions. Although two of the three units scaled up over the duration of the study, and the public nature of the scale-up's data dissemination made the intervention research hard to ignore early on, it remains a question whether the study of the effectiveness of the units and their scale-up would have had more impact if there had been a direct linkage to classroom performance within schools.

An Organised Agenda for Scale-up

In school districts, interventions are constantly introduced, but quickly disappear: an administrator buys some software for schools to use; a professional development effort pushes a particular approach for instruction; or the state assessment system changes and so must teachers' everyday assessments. Such interventions can be fleeting because they were never really evaluated in the district and later subjected to decisions based on beliefs or fluctuating funding levels. Scale-up research is based on accumulating evidence of an intervention's effectiveness in different contexts; typically, school districts cannot do that sort of painstaking research (cf McDonald et al. 2006). Scale-up's studies could not have occurred without substantial research funding. Many decision-makers claim to want such evidence for better decision-making.

The ability to demonstrate an intervention's impact seems absolutely necessary but, oddly enough, probably not sufficient for spread and sustainability (cf Borman and Hewes 2002; Desimone 2002). Because most interventions require teachers to do things differently and often demand additional work and skills, the justification for the inevitable extra effort would be to add value for students or teachers. Even when outcomes are positive, some interventions fade because they are too labour-intensive, require too much change or have prohibitive maintenance costs. Teachers

are faced with competing mandates. The effort associated with any one might not seem worthwhile, given an onslaught of innovations with differing priorities. If an intervention actually reduces teachers' work/effort while increasing student outcomes, it is likely to go to scale. Examples of scalable, sustainable innovations are scarce, particularly if they require changing beliefs, norms and principles.

Metaphor and Theory for Scale-up

Eagles in the Anacostia and Scale-up

Coburn (2003) and McDonald et al. (2006) contributed to the under-theorised study of scale-up in education. Coburn's definition includes often-neglected normative *outcomes* of scale-up at the school district level: depth, spread, sustainability and transfer of ownership. In contrast, McDonald et al.'s view is primarily *methodological*, given their vantage as managers of the knowledge diffusion from IERI scale-up portfolio. This chapter provides a practitioner/researcher view of a scale-up research study, emphasising conditions necessary for an intervention to go to scale; the flow of information is crucial, including its accumulation, interpretation, representation and presentation to stakeholders, and dissemination to a wider audience.

An analogy could illuminate the importance of information flow in a scale-up research system; it would require a situation in which one thing is obviously scaling-up, while something less obvious, but fundamental to growth and change, actually creates conditions for healthy proliferation. Ecological metaphors for scale-up have been used before (cf Cohen et al. 2001; McLaughlin and March 1978) and could be helpful here.

The Anacostia watershed of the Potomac River in Washington, DC, runs through a socio-economically stressed, ruderal area, where natural beauty competes with human neglect. Nonetheless, bald eagles were reintroduced to their ecological niche, and their population has been steadily increasing (Planet Maryland, March 21, 2001). The eagles, analogous to an educational intervention, are scaling up. Eagles are easy to spot; they are symbols of environmental health and wildness; and they have a patriotic connotation that allows them more public support for protection than other species. Because the watershed is constantly cleaned up, it can support a bald eagle population, as the media happily report. This environment has improved in several ways, but probably the most important is the healthy *flow of biomass* (fixed carbon) throughout the ecosystem that allows the eagles to find the fish that ate the plankton that fixed the sun's energy in carbon-hydrogen bonds. Thus, although the eagles are a visible symbol of scale-up, what has actually improved is the health of the system through better natural biomass cycling. In the Anacostia watershed, too much human trash or storm sewage runoff could affect the healthy cycle and lead to fewer eagles. Similarly, if eagles faced stiff competition for their food source from other introduced species, they could die off or find a better place to live. Biomass flow is analogous to the movement of information in a school

district successfully involved in scale-up efforts. Poor communication or mixed mandates would impede the flow of information about the progress of an intervention going to scale, eventually resulting in its extinction.

In order to scale-up, the intervention should fit the school district's needs (like the ecological niche occupied by the Anacostia's eagles). The district probably is best at determining its needs, and a close, long-term collaboration with researchers is a good way of finding or developing a likely intervention. The district's policy environment further determines the intervention's success by creating incentives or disincentives for it to go to scale. For instance, if middle school science students are newly required to take high-stakes tests that assess their abilities to reason from evidence, and if the intervention can be shown to encourage such reasoning, then the intervention is likely to have a greater chance of success. If the assessment system relies on factual minutia covering a lot of ground, but does not require reasoning from evidence, the intervention might be doomed despite its success in helping students to reason deeply.

It is unlikely that any large educational entity would expend resources in moving an intervention to scale unless the policy climate demands or supports it. This could include initial buy-in from administrators and teachers. But eventually, convincing, positive results must flow from the students themselves, influencing decision-makers' and teachers' beliefs, values and attitudes. The ability to stream accurate information into the environment depends on the mechanisms already in place within the school district (email, accessible websites, professional development meetings, human networks, policy systems and administrative hierarchies) and how well the researchers can tap into them or create new ones. In severely stressed school districts, scale-up is difficult because positive information gets lost in the detritus of bad news or a swirl of new initiatives that roil through the schools.

Scale-up research can bring external funding for new resources and services for teachers and students. To sustain the intervention at scale, a commitment for continuing support is crucial as research funding comes to an end. If feedback on student learning was vital to the intervention, it must continue. If professional development meetings for teachers were the means to exchange of information on improved implementation, then such meetings must persist.

Scale-up researchers inquired about existing, sustained, scaled-up interventions in science in its partner school district. There were two examples (B. Hansen and M. Szesze, personal communication, February 5, 2004). In one instance, a middle school environmental education programme requiring overnight stays at an outdoor centre had been in place for decades. It is integral to the middle school science programme and is a rite of passage for students who might never have been 'away at camp'. It is one person's full-time job to manage this programme for the entire district. A second example is 'kit-based elementary school curriculum units', formerly funded through an NSF grant. This intervention is sustained by employing a full-time science equipment czar whose job is to procure equipment inexpensively for the kits, package it and send it off to elementary schools. His role expanded to include procurement for secondary school science programmes (including equipment for scale-up's interventions) and is fully integrated into the system. Just as the

Anacostia's eagle population is unlikely to be sustained without continuous human stewardship, it seems likely that relatively sophisticated education interventions also require ongoing stewardship, something that ought to be acknowledged and built into scale-up research if sustainability is a serious goal.

Activity Systems in Scale-up Intervention Research

The ecological metaphor for the scale-up of eagles in the Anacostia and the scale-up of science curriculum units in a large school district suggest overlapping, interrelated systems and layers of complexity, as well as the human actors crucial to scale-up and sustainability. Activity theory can help to explain scale-up's research programme, and more generally capture the complexity of relationships and meanings for the scale-up of education interventions in school systems. Activity theory's roots come from animal evolution and the natural environment, but it has been applied in human cultural evolution. What used to be ecological and natural becomes economic and historical (University of Helsinki Center for Activity Theory and Developmental Work Research 2006, p. 1). According to Linda Gilbert (1999), it is a development of socio-cultural theory, with much in common with current learning theories that marry notions of distributed cognition and situated cognition. There is a common focus on the interaction of the individual with the environment in gaining or using knowledge, with origins in the work of Vygotsky and his follower Leont'ev (1978). It neither is a theory in a strict interpretation of the term, nor is it predictive. Rather it is a powerful and clarifying descriptive tool and can be viewed as a general conceptual system. John Carroll (1997) describes activity theory:

The object of description in this approach is an "activity system," the ensemble of technological factors with social factors, and of individual attitudes, experiences, and actions with community practices, traditions, and values. Activity theory emphasizes that these ensembles are essentially contingent and changing, that human activities are mediated and transformed by human creations such as technologies, and that people make themselves through their use of tools...Activity theory shifts attention from characterizing static and individual competencies toward characterizing how people can negotiate with the social and technological environment to solve problems and learn, which subsumes many of the issues of distributed and situated cognition (p. 512).

According to Graham Nuttall (2000), although activity theory research sometimes focuses on the use of computers as technological tools, it is also used as the basis for the generic analysis of the patterning of classroom experiences. According to Wells (as cited in Nuttall 2000), an activity is a relatively self-contained, goal-oriented unit, such as carrying out an experiment or writing a story. It consists of a series of behaviours or tasks that follow an expected pattern to achieve a goal, held together by the mutual interrelated expectations of participants (although how it is carried out can vary in time and place). Activity theory allows researchers to take a socio-cultural perspective in understanding how diverse students learn. On the other hand, some researchers equate student activities with 'learning', without direct reference to what might be occurring in students' minds; this idealises participation in

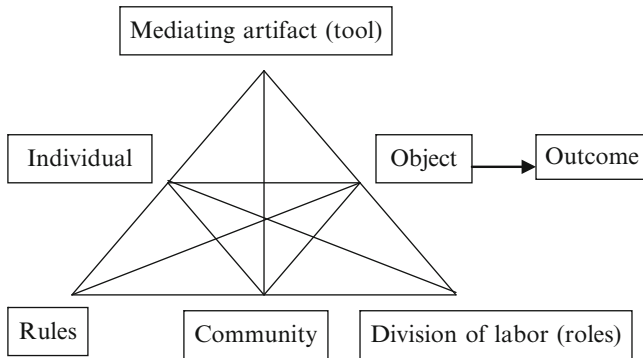


Fig. 61.2 Model of activity system (Adapted from Engeström 1992)

classroom activities as both the process and an end of learning. Nuttall (2000) believes that there are serious problems with this assumption, including the difficulty in interpreting what the activities of students mean, because of students' varied cultural backgrounds. Students might display interest when they in fact are not interested. Moreover, they could expend the least possible effort in carrying out tasks that are likely to be noticed or evaluated. Consequently, Nuttall urges that researchers attempt to capture what is in students' minds (concepts), as well as attitudes and beliefs, but uses activity theory as the basis for research. Yrjö Engeström (1992) provides a framework to describe a mediated activity system which consists of individuals, colleagues in the workplace community, conceptual and practical tools, and the shared objects (similar to objectives) as a unified and dynamic whole, depicted in Fig. 61.2.

Student Activity System

In scale-up research on middle school science curriculum units, the first activity system to consider is that of the student (individual) who is learning with the new curriculum unit, the mediating artefact or tool (see Fig. 61.2). The enactment of the curriculum unit in the classroom requires the student to follow rules in a community that consists of other students and the teacher. The teacher defines the division of labour in the classroom, further shaped by the curriculum materials and other students' actions, particularly students who are organised into laboratory groups. The object for the student is to learn the concepts from the curriculum unit, scaffolded by the teacher, curriculum materials and student peers, as the unit is enacted. The outcomes consist of laboratory journal responses, performance on assessments and grades. Improved understanding of complex science ideas is the ultimate outcome. Although the student was not much aware of it, in scale-up research, the disaggregated student outcome data made the most compelling public case for the intervention's impact.

Teacher Activity System

Concurrently, the teacher activity system sets the teacher's object as the implementation of the intervention unit, aided by professional development experiences and direct interactions with the written materials. In scale-up, the tool is the curriculum unit for both teacher and student activity systems. Each unit also has a teacher manual that guides the teacher to further explain the object. The teacher should follow the rules of the school and district, as well as those of the professional development and science teaching communities engaged in the study. The rules for teachers engaged in a scale-up research study might differ somewhat from those for teachers who are using the tool in a more routine way. In an effectiveness research study, the teacher should use the tool as intended/implied by the curriculum materials in order to ensure that the research is valid. Teachers also determine student division of labour and interact with other professionals who assume roles such as peers/coaches, supervisors, evaluators and researchers. The teacher's immediate outcome is the perception that the unit's lessons are going well or failing, discerned through students' daily interactions and cumulative work. However, if the research also provides collective measures of positive student outcome data in other classes and in other schools, the teacher has another way of weighing the unit's effectiveness. Thus, even if individual teachers have doubts, there is a feedback mechanism that can reassure them.

Researcher Activity System

In scale-up, the research team consisted of university researchers, science education administrators, evaluators and teacher peer/coaches. The object for this activity system was to study the curriculum unit's (tool) impact and scale-up. Each member of the research team had a specialised role, while keeping the interests of classroom teachers and students in mind. Researchers operated in the system of rules set for the project, as well as rules of the school system and the larger educational research communities. Each member of the research team participated in different kinds of actions related to the object, including the formal and informal collection and analyses of evidence, fully aware that valid student outcome data are crucial to scale-up research.

In summary, there were three different activity systems in play (student, teacher, researcher) in scale-up. Each had much in common with the other, in membership, community and rules. All relied on the same tool (the curriculum unit) and object (to achieve positive student outcomes that are valid and reliable, to be distributed publically). Activity theory captures the complexity of scale-up research, while fixing common terms and ideas to explain what is occurring. The extent of agreement between the activities systems related to a common object probably can predict an intervention's degree of success. The four conditions identified in this chapter as crucial to the scale-up of new curriculum materials (close school district/research collaboration, a positive climate for the intervention, the collection of

student outcome data, and the system that makes data accessible to move the scale-up research agenda) help the three overlapping activity systems to adhere and remain congruent.

On the other hand, the student, teacher and researcher activity systems, no matter how compatible early on in scale-up, soon faced competition. Middle school students come to school with many competing goals (social and academic), some of which could conflict with object/intervention of a science curriculum unit. The teacher activity system focuses on the intervention science curriculum unit, but teachers have other objects as well. Teachers participating in scale-up voiced concerns about the amount of time that intervention units take and whether this leaves sufficient time to cover other topics than mandated by the new state curriculum and high-stakes assessment system. For the researcher activity system, school district science collaborators were subject to competing accountability structures. Their jobs demanded participation in other activity systems, such as designing new middle school science curriculum guides. They used the information gained from scale-up research to influence the design of the new guides and the associated professional development. Ultimately, it was the flow of information that was really scaling-up.

Summary

This chapter is intended to advance theory-building for scale-up research. Prior articles by Coburn (2003) and McDonald et al. (2006) offered different views of scale-up: one is normative and retrospective; and the other is methodological and general. In contrast, this chapter provides a highly contextualised perspective on scale-up from the ground level, as this study developed over 6 years. Metaphors such as the scale-up of eagles in the Anacostia illuminate the scale-up of middle school science curriculum units in a large public school district. This metaphor is important because it demonstrates that what might be obviously scaling up could be dependent on the underlying health of the information system that nurtures it. These complex interactions are further explored from the standpoint of activity theory. An understanding the congruence of overlapping activity systems provides a way in which to see the potential for an intervention to go to scale. Moreover, the continued alignment of important activity systems is likely to determine the sustainability of the intervention over the long run. Changes in human activity systems, no doubt, are inevitable and responsive to factors outside any one given school system. Seeing such changes as natural, but not necessarily inevitable, is helpful in understanding the ecology of educational reform and shaping its future.

Acknowledgement This work (REC-0228447) was supported by the National Science Foundation (NSF), the US Department of Education (USDOE), and the National Institute of Health (NIH). Any opinions, findings, conclusions or recommendations are those of the author and do not necessarily reflect the position, policy or endorsement of the funding agencies. The author thanks Okhee Lee, Joel Kuipers, Annie Knight, Bill Watson and Carol O'Donnell for their contributions to this chapter.

References

- American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost effectiveness of Success for All. *Educational Evaluation and Policy Analysis*, 24, 243–266.
- Caroll, J. M. (1997). Human-computer interaction: Psychology as the science of design. *International Journal of Human-Computer Studies*, 46, 501–522.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3–12.
- Cohen, D. K., Ball, D. L., & Barnes, C. A. (2001). *Scale in improving instruction* (Proposal submitted to NSF/IERI 2001 competition). Ann Arbor, MI: University of Michigan.
- Dahlkemper, L. (2003). What does scientifically based research mean for schools? *SEDL Newsletter*, XV(1), 3–6.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72, 433–479.
- Engeström, Y. (1992). *Interactive expertise: Studies in distributive working intelligence* (Research Bulletin No. 83). Helsinki, Finland: Helsinki University Department of Education. (ERIC Document Reproduction Service No. ED349956).
- Gilbert, L. S. (1999, February). *Where is my brain? Distributed cognition, activity theory, and cognitive tools*. Proceedings of selected research and development papers presented at the National Convention on the Association for Educational Communications and Technology, Houston, TX.
- Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching*, 39, 522–549.
- Kuipers, J. C., Viechnicki, G. B., Massoud, L., & Wright, L. J. (in press). Science, culture, and equity in curriculum: An ethnographic approach to the study of a highly-rated curriculum unit. In K. Richardson & K. Gomez (Eds.), *Talking science, writing science: The work of language in multicultural classrooms*. Mahwah, NJ: Lawrence Erlbaum and Associates.
- Leont'ev, A. N. (1978). *Activity, consciousness, and personality* (M. J. Hall, Trans.). Englewood Cliffs, NJ: Prentice-Hall.
- Lewis, A. (2003, Fall). Research goes to school—Part 3. *CRESST LINE Newsletter of the National Center for Research on Evaluation, Standards, and Testing*, 4–9.
- Lynch, S. (2000). *Equity and science education reform*. Mahwah, NJ: Lawrence Erlbaum and Associates.
- Lynch, S. J. (2008, March). *How do curriculum materials improve student outcomes at the class level? Study of construct validity, fidelity of implementation, and the comparison group*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, 42, 912–946.
- Lynch, S., & O'Donnell, C. L. (2005, April). *The evolving definition, measurement, and conceptualization of fidelity of implementation in scale-up of highly rated science curriculum units in diverse middle schools*. Paper presented at the annual meeting of the American Educational Researchers Association, Montreal, Canada.
- Lynch, S., Szesze, M., Pyke, C., & Kuipers, J. C. (2007a). Scaling-up highly rated middle science curriculum units for diverse student populations: Features that affect collaborative research, and vice versa. In B. Schneider & S. K. McDonald (Eds.), *Scale-up in education, Volume 2: Issues in practice*. Lanham, MD: Rowman & Littlefield.
- Lynch, S., Taymans, J., Watson, W. A., Ochsendorf, R., Pyke, C., & Szesze, M. (2007b). Scaling up highly rated curriculum units for students with disabilities in mainstream classrooms: Initial findings and implications for scale-up. *Exceptional Children*, 73, 202–223.

- McDonald, S. K., Keesler, V. A., Kaufman, N. J., & Schneider, B. (2006) Scaling-up exemplary interventions. *Educational Researcher*, 35(3), 15–24.
- McLaughlin, M. B., & March, D. D. (1978). Staff development and school change. *Teachers College Record*, 80, 71–94.
- National Science Foundation. (2002). *Interagency Education Research Initiative (FY2002) (IERI) program solicitation*, NSF-02-062. Arlington, VA: Author.
- Nuttall, G. (2000, November). *Understanding what students learn in school*. Paper presented at the annual meeting of the New Zealand Association for Research in Education, Hamilton, New Zealand (ERIC Document Reproduction Service No. ED455205).
- O'Donnell, C., Lynch, S., Lastica, J., & Merchlinsky, S. (2007, April). *Analyzing the relationship between Fidelity of Implementation (FOI) and student outcomes in a quasi-experiment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.
- Planet Maryland. (2001, March 21). *Bald eagle population soars to 23-year high along Chesapeake Bay* [Television Broadcast]. Maryland Public Television.
- Schneider, B., & McDonald, S. K. (Eds.). (2007a). *Scale-up in education, Volume 1: Ideas in principle*. Lanham, MD: Rowman & Littlefield.
- Schneider, B., & McDonald, S. K. (Eds.). (2007b). *Scale-up in education, Volume 2: Issues in practice*. Lanham, MD: Rowman & Littlefield.
- State of Michigan. (1993). *Chemistry that applies*. Lansing, MI: Author.
- University of Helsinki Center for Activity Theory and Developmental Work Research. (2006). *The activity system*. Retrieved October 25, 2006 from the University of Helsinki Center for Activity Theory and Developmental Work Research, Department of Education Web site: <http://www.edu.helsinki.fi/activity/pages/chatanddwr/activitysystem/>
- Watson, W., Pyke, C., Lynch, S., & Ochsendorf, R. (2007, April). *Understanding the effectiveness of curriculum materials through replication*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA.
- Weast, J. D. (2000). *Studies of mathematics instruction and curriculum: Implications for the future* (Memorandum to MCPS Board of Education). Rockville, MD: Montgomery County Public Schools.