The scaling of reading interventions: Building multilevel insight

FINBARR C. SLOANE

doi:10.1598/RRQ.40.3.4

The National Science Foundation, Arlington, Virginia, USA

Peter Freebody and Margery Hornibrook (later in this issue) raise the following question: What kind of research can guide educators and policymakers with a commitment to literacy and social justice? They address their question by drawing our attention to policy actions being taken in Australia and New Zealand regarding the use of advanced technologies to support reading instruction. In my mind, this question raises other important issues for the reading research community. These include, but are not limited to, clarity about the measurement and modeling of reading comprehension and the development of multilevel theory to support reading interventions at scale. My response to their question is to take on the more mundane task posed by the latter issue: the need for clear multilevel theory to support policy efforts. I specifically address this need here. In doing so, I draw on my experiences as program director responsible, along with my colleagues at the Interagency Education Research Initiative (IERI), for the oversight of recent scaling efforts in U.S. schools for research interventions in the content areas of reading, mathematics, and science.

Research in the development and testing of multilevel models has increased dramatically in the past 15 years (Raudenbush & Bryk, 2002), yet there have been relatively few efforts made that provide a broad multilevel theoretical framework for education research, let alone reading research. For the purposes of this article, I define multilevel reading research as "reading research that links teacher and student behavior, and honors both of these units and the linkages across these units theoretically and empirically." For the purposes of this presentation, I confine the discussion to two-level models. I do this solely for the purposes of clarity.

Multilevel theory building represents a substantial challenge to educational researchers trained to think "micro" (i.e., psychologists in particular and reading researchers in general), or to think "macro" (i.e., sociologists), but not to think "micro and macro" (e.g., see Barr & Dreeben, 1983, who try to apply both reading and organizational theory to learning to read). Given the training of those involved in reading research and the tradition of research in the community, cross-level theory building is a challenge. However, the problems of education and of reading interventions are invariably multilevel in nature. Students are nested in reading groups, and reading groups are in classrooms, where instruction is provided. Alternatively, instruction is provided by reading specialists in small groups distinct from the classroom. When instruction is provided to groups of individuals (in their classroom or remotely), then the problem under study can be configured so as to address this nested or multilevel structure.

In this article, I describe some of the possible components needed to develop multilevel theory that support the scaling of reading interventions. Scaling can be defined in a number of ways: depth, sustainability, spread, and shift (see Coburn, 2003). I treat "scaling" here as it has been traditionally defined—that is, simply to increase the number and types of sites in which the intervention is being studied. This traditional definition of scale, correctly considered by Coburn as theoretically noncomplex, still considerably increases the complexity of the phenomenon to be studied for intervention researchers. This is because it is highly unlikely that interventions will be delivered with the same degree of fidelity as is true in smaller studies of the same intervention.

In contrast to the assumption that the intervention can be estimated as a fixed effect, these problems are best dealt with theoretically and empirically. This can be done by explicitly understanding and accounting for the multilevel structure inherent in the data and developing clear and coherent multilevel theory. Briefly, the elements of such multilevel theory I examine here include the what, the how, the where, and

the when of such studies. The list is not exhaustive, but rather it points to the need for such theoretical development. Reading research is multilevel by nature because of the naturally occurring nesting in reading data. Few, if any, researchers explicitly leverage this multilevel structure to support theoretical developments in the study of reading interventions. Currently, the multilevel structure of reading data is either ignored or is utilized to reduce bias and better estimate our statistical models. However, the time is now ripe for the development of more sophisticated theory as we begin to implement reading interventions at scales never seen before.

Multilevel theory building in reading

Multilevel theory is neither always needed nor always better than single-level theory. Micro theorists may articulate theoretical models capturing individual-level processes that are invariant across contexts, or they may examine constructs and processes that have no meaningful parallels at higher levels. This is true for much of the interventionist experimental work in reading, where the implementation researchers train two to four reading specialists to implement the intervention. The veracity of this line of work entails assuming that treatments have fixed effects and that covariates used in the estimation models do not vary across reading groups or classrooms. Here the student is used as the sole unit of analysis, and issues associated with implementation where the teacher is also an analytic unit are temporarily dismissed as unimportant. In such cases, where the assumptions can be met, multilevel theory building is not necessary.

Reading theorists may also find it impractical to develop multilevel models for processes, relationships, and outcomes new to educational science. That is, when tackling phenomena previously unexplored in the reading literature, a theorist may find it helpful to act initially as if the phenomena occur at only one level of theory and analysis. In this way, a theorist temporarily restricts his or her focus, putting off consideration of multilevel processes for a period. It should be noted that, until recently, reading research mapped to this situation. That is, many reading researchers operationalized the design of their studies so that the student became the sole unit of analysis. This afforded enormous control to the researcher but assumed that treatment fidelity was not an issue. Here the researcher controlled the fidelity of implementation by varying the duration and intensity of the training process.

Use of this strategy when bringing the intervention to scale may be quite different, and depending on how tight the funding flow is, the strategy may be impossible to implement in practice. In studies at scale, the researcher has to evaluate whether the student is the only unit under study, and whether the effect to be estimated at the level of the teacher (or classroom) is always a fixed effect. This evaluation will cause the researcher to address the multilevel issues inherent in the proposed line of work. Consequently, the time may now be ripe for explicit multilevel theory building, for the scaling of reading interventions requires such theory.

Reading environments in schools are hierarchically nested (students in groups, groups in classrooms, classrooms in schools, and so on). The set of individual-level phenomena that are invariant across these contexts is likely to be smaller than we would like. Similarly, the set of group- or organization-level phenomena that are completely uninfluenced by lower levels is also likely to be small. Failure to account for such effects when they exist will, of course, yield incomplete or misspecified models.

What might good multilevel theory look like?

By definition, multilevel models are designed to bridge micro and macro perspectives. Such models must accurately specify relationships between phenomena at higher and at lower levels of analysis (for example, individuals and groups, groups and organizations, and so on). Accordingly, a multilevel theoretical model must specify how phenomena at different levels are linked. Links between phenomena at different levels may be top down or bottom up. Top-down theories include what teachers do in instructional groups and how this behavior varies across teachers (e.g., the number of groups used, their composition, and the allocation of time to phonics by group). Theories may include both top-down and bottom-up processes (and these will come to bear in the design of reading interventions). Analytically, however, the researcher is presently confined to top-down estimation procedures. These include Hierarchical Linear Modeling (Raudenbush & Bryk, 2002), Multilevel n (Goldstein, 1995), and the PROC MIXED procedure (Singer, 1998), to name but a few. Consequently, I focus attention solely on top-down processes. This is not to say that bottom-up, or emergent processes, do not play a critical role.

Rather, this focus highlights the state of the statistical art and our need to theorize, when relevant, in a way that optimizes the value of the analytic tools at our disposal. While we can theorize about the effects of bottom-up processes, we cannot estimate these effects statistically.

What outcomes?

The first step in the building of multilevel theory building is the "what." Put simply, on what should multilevel theory building and research focus? The possibilities are clearly endless, reflecting the full breadth of processes, behavior, and theory relevant to reading. For reading researchers involved in implementation studies, I believe that the focus of multilevel theory building must now begin to center on how children learn to read with comprehension and how we, as researchers, theorize, measure, and analyze this learning. The learning on the part of the research community must draw on studies of reading comprehension along with studies of decoding. As noted in the first paragraph, the measurement of reading comprehension is still a critical concern for the reading community (see Snow, 2002). The future work in the area will include not only the measurement of comprehension at fixed points in time but also the conceptualization and development of measures of comprehension that allow for the valid capture of change over time. When viewed in this manner, a further layer of multilevel theorizing can be considered, as multilevel estimating tools allow for the analysis of student growth (i.e., student development) over time.

Put simply, with good measures of comprehension that are sensitive to change over time, the research community will be better able to build and test multilevel theories that explicitly look at student developmental rates as the outcome measure of interest. In sum, student growth in reading comprehension can be analyzed from a multilevel perspective (Raudenbush & Bryk, 1988). Consequently, as noted above, the measurement of comprehension is critical to the future of reading research. Again, the development of these measures must serve two distinct purposes: They must be able to validly and reliably distinguish between students at fixed points in time, and they must also distinguish validly and reliably the developmental trajectories students take as they learn to read with comprehension (should such trajectories exist in reality).

The influence of context: Top-down processes

Each level of an educational system is embedded or included in a higher level context. Thus, in reading, individuals are embedded within groups and groups within classrooms. In addition, we find classrooms within schools (or school types) and schools within environmental niches and school districts. Top-down processes describe the influence of higher level contextual factors on lower levels of the system. Higher level units (e.g., teachers) may influence lower level units (e.g., students) in two distinct ways: (1) Higher level units may have a direct effect on lower level units, or (2) higher level units may shape or moderate relationships and processes in lower level units.

Classroom teachers have a direct effect on the behavior of their students when they determine the accepted patterns of student interaction and work behavior in a reading group. The same teachers have a moderating effect on lower level relationships when the relationship between two lower level constructs changes as a function of reading intervention. For example, the relationship between students' individual levels of engagement and performance may vary across classroom contexts as a function of a chosen and well-executed implementation strategy (e.g., see Fletcher, 2005; Slavin, 2004). In contexts that provide autonomy and equal access to rich intellectual resources, student engagement may be associated with performance. However, in contexts that are low on autonomy and on access to intellectual resources, there are likely to be constraining effects of student engagement on performance, weakening the proposed relationship. Being specific about these forms of influence when bringing a reading implementation to scale is critically important. These cross-level influences provide theoretical and testable insight on the life of the intervention only when they are explicitly acknowledged and tested.

What is the role of time?

Time is rarely an overt theoretical consideration in either single-level or multilevel models of educational interventions (Sloane, 1993). This is as true in reading research as it is in other areas of educational research. Yet it is clearly the case that student (and teacher learning) is influenced and shaped by time. Here I explore two ways in which time may be incorporated into a multilevel model to increase

the rigor and effectiveness of the multilevel theory-building process.

Time as a boundary condition or moderator

Many grouped phenomena have a unidirectional effect on higher or lower level organizational phenomena, but multilevel relationships may not always be so simple. Instead, over time the relationship between phenomena at different levels may change in direction. As researchers, we may ask ourselves this question: Does motivation to succeed come before achievement? Alternatively, we could ask it the other way around: Does high achievement produce higher motivation? These alternative renderings of the research question highlight the critical issue of causal order. This is also true for higher level (macrolevel entities) and the linkages across these levels. Put simply, a given theoretical phenomenon may appear to originate at a higher or lower structural level according to the theorist's assumption about the current time point in a stream or cycle of events.

Paralleling the above example, the role a reading intervention plays on the relationship between individual motivation to read and a student's achievement in reading comprehension will likely vary over time. For example, the mediating role of motivation on reading achievement for the good reader may be very different from that for the struggling reader. The researcher needs to be theoretically explicit about the role of the proposed intervention (i.e., a group-level variable) in mediating the estimated relationship between motivation and achievement that occurs for individuals within groups.

A failure quite common to intervention research is to assume that the mediating effects, if estimated at all, are the same for all readers. When effects are presented in this manner, the researcher assumes that temporality is not an issue. In a unilevel model (e.g., with students as the analytic entity), one would generally present the measure of motivation as a covariate (introduced to reduce between-person differences) rather than as a mediating variable. Consequently, it is important that the researcher be as clear as possible in outlining the role of the chosen variables and the temporal scope of the intervention. Where possible, the theorist must explicitly specify the temporal reference points with respect to the multiple groups the intervention is meant to assist.

Variations in the time scale across levels

Differences in time scales affect the nature of links among levels (Simon, 1973). Lower level phenomena tend to have more rapid dynamics than higher level and, for that matter, emergent phenomena. This makes it easier to detect change in lower level entities. For example, models of school change take more time to develop than does individual learning as a consequence of said change (Slavin, 2004). The timing of evaluations that try to simultaneously capture school-level, teacher, and student measures must employ research designs that are sensitive to the temporal requirements of the multilevel theory being tested. This is one likely reason why top-down models predominate the survey literature on schools and the process of schooling. An interesting implication of this effect of time scale is that phenomena at different levels may manifest at different points in time. Researchers must be overt about their theoretical tradeoffs, and research designs must be sensitive to the temporal requirement of the proposed theory.

Conclusion

Argument by assertion is invariably a poor strategy for theory building. Argument by logical analysis and persuasion—argument that explains why—is always preferable. Being explicit about the models we theorize, build, and test is important. In multilevel theory building, explaining why is not merely preferable but essential. Educational multilevel theory and reading interventions conducted in nested settings draw on multiple subdisciplines (e.g., in education, cognition, neuroscience, psychology, and sociology). Therefore, the unstated assumptions in a multilevel theory may be obvious to the members of one subdiscipline but not to the members of another, who are equally interested in the new multilevel theory. One could argue that, among other concerns, assumptions such as these separate the work and harden the debate between cognitivists and socioculturalists found in the reading research literature.

In addition, multilevel data analysis has been the subject of considerable and continual debate. Many of the controversies and problems associated with multilevel research are based on misspecifications or misalignments among the theoretical level of constructs, their measurement, and their representation for analysis. The nature of these misalignments is well documented elsewhere (Burstein, 1980;

Firebaugh, 1979; Freeman, 1980: Robinson, 1950; Thorndike, 1939). For example, the blind aggregation of individual-level measures to represent group-level constructs can be problematic and needs to be thought through theoretically. Misalignments degrade construct validity and generate concerns about generalizability. To build theoretical models that are clear and persuasive, scholars must explicate the nature of the constructs. Precise explication lays the foundation for sound measurement.

In addition, conflicts exist regarding the best way to analyze multilevel models (see Kreft & DeLeeuw, 1998). Consequently, research design and analytic strategies need to be aligned with the levels inherent in these models. However, these debates are often quelled in the presence of carefully and fully explicated theoretical models that make the choice of analytical strategy clear (Raudenbush & Bryk, 2002). Thus multilevel reading researchers must specify not only what, how, where, and when, but also when possible and why. For example, why are relationships in the model conceptualized as top down rather than bottom up? Earlier, I noted that this could be due to software constraints. Why are predictors assumed to have immediate rather than long-term consequences for the outcomes of interest? The list goes on.

Nearly as important as the question of why, and perhaps even more interesting, is the question of why not. Why might bottom-up processes *not* yield a group-level property? That is, why might members' perceptions *not* converge to form a shared unit of measure (e.g., aggregates for school-level reading achievement in whole-school interventions like that of Success for All)? Why might top-down processes *not* constrain relationships in a reading group? If implementation requires a high degree of teacher knowledge or training for fidelity, then explicitly measuring and testing for the degree of fidelity is critical to the theory, the intervention, and to later implementations of the reading intervention. Why might predictors hypothesized to be influential over time prove instead to have immediate consequences? In exploring why not, reading researchers will likely refine their models to incorporate important insights and nuances related to the practice of implementation. This extra effort will help add clarity and depth to their multilevel theory building and, in parallel, to the building of a science for the study of how reading interventions that depend on classroom teachers work to effect student learning.

Finally, one might ask, what happens to the individual reading researcher if the proposed model for understanding the scaling of reading interventions should come to pass? The answer is quite simple. The role of the individual independent researcher will not go away. Basic research will still need to be conducted. It is clear that there is much measurement ahead (e.g., particularly in the area of reading comprehension). Moreover, the necessary small-scale studies to support the development of interventions need to be conducted. However, critical questions and answers regarding the implementation of this basic research in classroom settings will be conducted in an interdisciplinary manner and serve to provide important and critical feedback to the ongoing work of the single investigator. This feedback loop should allow for a more seamless process between basic laboratory work and classroom intervention work before the scaling of interventions are undertaken (Brown, 1992). The proof of the pudding will, of course, be in the eating.

FINBARR C. SLOANE completed his PhD at the University of Chicago. He is a program director at The National Science Foundation, Division of Research, Evaluation and Communication. His research interests include student learning of mathematics, multilevel theory related to the scaling of educational interventions, and statistical modeling. He may be contacted at The National Science Foundation, Division of Research, Evaluation and Communication, Room 855S, 4201 Wilson Boulevard, Arlington, VA 22230, USA, or by e-mail at fsloane@nsf.gov.

REFERENCES

BARR, R., & DREEBEN, R. (1983). *How schools work*. Chicago: University of Chicago Press.

BROWN, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences*, *2*, 141–178.

BURSTEIN, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, *8*, 158–233.

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

FIREBAUGH, G. (1979). Assessing group effects: A comparison of two methods. *Sociological Methods and Research*, *17*, 384–395.

FLETCHER, J.M. (2005). Retrieved August 2004 from http://drdc.uchicago.edu/community/projects/fletcher.shtml

FREEMAN, J. (1980). The unit problem in organizational research. In W.M. Evans (Ed.), *Frontiers in organization and management* (pp. 59–68). New York: Praeger.

GOLDSTEIN, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.

KREFT, I., & DELEEUW, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage.

RAUDENBUSH, S., & BRYK, A. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, *15*, 423–76.

RAUDENBUSH, S.W., & BRYK, A.S. (2002). *Hierarchical linear models* (2nd ed.). Newbury Park, CA: Sage.

ROBINSON, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–257.

SIMON, H.A. (1973). The organization of complex systems. In H.H. Pattee (Ed.), *Hierarchy theory* (pp. 1–27). New York: Braille.

SINGER, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*, 323–355.

SLAVIN, R. E. (2004). Education research can and must address "what works" questions. *Educational Researcher*, *33*(1), 27–28.

SLOANE, F. (1993). *Multilevel models for the study of learning.* Invited address University Council for Educational Administration annual meeting, New Orleans, LA.

SNOW, C. (2002). *Reading for understanding: Toward a research and development program in reading comprehension.* Santa Monica, CA: RAND.

THORNDIKE, E.L. (1939). On the fallacy of imputing the correlations found in groups to the individuals or smaller groups composing them. *American Journal of Psychology, 52*, 122–124.

AUTHOR'S NOTE

The views expressed in this article do not necessarily reflect the views of The National Science Foundation or those of the Interagency Education Research Initiative (IERI) program. I would like to thank Donna Alvermann, Anthony Kelly, and Gabe Della-Piana for their encouragement, commentary, and critical feedback. An earlier version of this article that focused on multilevel theory in mathematics education was presented to the faculty of St. Patrick's College of Education, Dublin, Ireland, in the fall of 2004. I thank the faculty of the College in general, along with Sean Close and Hugh Gash, in particular, for their verbal feedback.

IERI is a program of research whose purpose is to study the scaling of educational interventions in reading, mathematics, and science. Three funding agencies participate as equal partners in the funding and oversight of awards. These agencies are The National Science Foundation, The National Institutes for Health, and the U.S. Department of Education.