EDUCATION RESEARCH

Instructional Complexity and the Science to Constrain It

School-researcher partnerships and large in vivo experiments help focus on useful, effective, instruction.

Kenneth R. Koedinger[1]*, Julie L. Booth[2], David Klahr[1]

Science and technology have had enormous impact on many areas of human endeavor but surprisingly little effect on education. Many large-scale field trials of science-based innovations in education have yielded scant evidence of improvement in student learning (*1*, *2*), although a few have reliable positive outcomes (*3*, *4*). Education involves many important issues, such as cultural questions of values, but we focus on instructional decision-making in the context of determined instructional goals and suggest ways to manage instructional complexity.

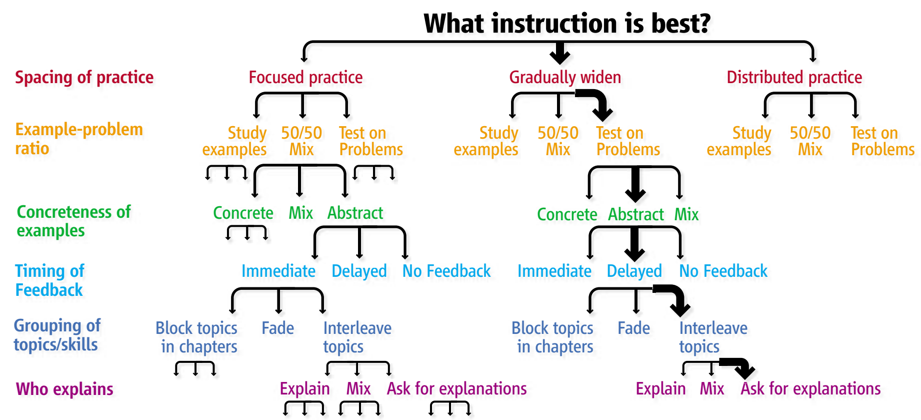
Ambiguities and Contexts in Instruction

Many debates about instructional methods suffer from a tendency to apply compelling labels to vaguely described procedures, rather than operational definitions of instructional practices (*5*, *6*). Even when practices are reasonably well defined, there is not a consistent evidential base for deciding which approach is optimal for learning. Empirical investigations of instructional methods, including controlled laboratory experiments in cognitive and educational psychology, often fail to yield consensus. For instance, controversy exists regarding benefits of immediate (*7*) versus delayed feedback (*8*), or use of concrete (*9*) versus abstract materials (*10*).

Further complicating the picture is that results often vary across content or populations. For example, instruction that is effective for simple skills has been found to be ineffective for more complex skills (*11*), and techniques such as prompting students to provide explanations (*12*) may not be universally effective (*13*). Effectiveness of different approaches is often contingent on student population or level of prior achievement or aptitude. Some approaches, for example, may be particularly effective for low-achieving students (*14*, *15*). Although specific instructional decisions may be useful at the level of the individual student (e.g., will this student learn better right now if I give her feedback or if I let her grapple with the material for a while?), the search for general methods that optimize the effectiveness, efficiency, and level of student engagement is more challenging.

Complexity of Instructional Design

Of the many factors that affect learning in real-world contexts, we describe three of particular importance: instructional technique, dosage, and timing. How choices on one dimension can be independently combined with choices on other dimensions to produce a vast space of reasonable instructional choice options is shown in the figure (Fig. 1).



Title. Different choices along different instructional dimensions can be combined to produce a vast set of instructional options. The path with thicker arrows illustrates one set of choices within a space of trillions of such options.

Instructional techniques. Many lists of learning principles suggest instructional techniques and point to supporting research (*12*, *16*). Each list has between 3 and 25 principles. In-depth synthesis of nine such sources yielded an estimate of 30 independent instructional techniques (see the table and table S1) (Table 1).

Dosage and implementation. Many instructional distinctions have multiple values or are continuous (e.g., the ratio of examples to questions or problems given in an assignment, the spacing of time between related activities). These dimensions are mostly compatible with each other—almost all can be combined with any other.

Intervention timing. The optimal technique may not be the same early in learning as it is later. Consider how novice students benefit from studying many worked examples in place of many problems, whereas shifting to pure problem-solving practice becomes more effective as students develop expertise (*17*). Many researchers have suggested that effective instruction should provide more structure or support early in learning or for more difficult or complex ideas and fade that assistance as the learner advances (*18*, *19*).

If we consider just 15 of the 30 instructional techniques we identified, three alternative dosage levels, and the possibility of different dosage choices for early and late instruction, we compute 3^{15*2} or 205 trillion options. Some combinations may not be possible or may not make sense in a particular content area, yet other factors add further complexity: Many techniques have more than three possible dosage levels, there may be more than two time points where the instructional optimum changes, different knowledge needs in different domains often require a different optimal combination. For example, it may be optimal to adjust spacing of practice continually for each student on each knowledge component (*20*). As another example, when the target knowledge is simple facts, requiring recall and use of knowledge produces more robust learning, but for complex problem-solving skills, studying a substantial number of worked example is better (*1*).

The vast size of this space reveals that simple two-sided debates about improving

[1]Carnegie Mellon University, Pittsburgh, PA 15213, USA.
[2]Temple University, Philadelphia, PA 19122, USA.
*Corresponding author. koedinger@cmu.edu

CREDIT:

learning—in the scientific literature, as well as in the public forum—obscure the complexity that a productive science of instruction must address.

Taming Instructional Complexity

We make five recommendations to advance instructional theory and to maximize its relevance to educational practice.

1. *Searching in the function space.* Following the Knowledge-Learning-Instruction framework (*21*), we suggest three layers of functions of instruction: (i) to yield better assessment outcomes that reflect broad and lasting improvements in learner performance, (ii) instruction must change learners' knowledge base or intellectual capacity and (iii) must require that learners' minds execute appropriate learning processes.

We specify different functions to be achieved at each layer. The most distal, but observable, functions of instruction are assessment outcomes: long-term retention, transfer to new contexts, or desire for future learning. More proximal, but unobservable, functions are those that change different kinds of knowledge: facts, procedural skills, principles, learning skills, or learning beliefs and dispositions. The most immediate and unobservable functions support learning processes or mechanisms: memory and fluency building, induction and refinement, or understanding and sense-making (*21*, *22*).

Functions at each layer suggest more focused questions that reduce the instructional design space (*23*): Which instructional choices best support memory to increase long-term retention of facts? Which are best for inducing general skills that produce transfer of learning to new situations? Which are best for sense-making processes that produce learning skills and higher learner self-efficacy toward better future learning? We can associate different subsets of the instructional design dimensions with individual learning functions. For example, spacing enhances memory, worked examples enhance induction, and self-explanation enhances sense making (see the table). The success of this approach of separating causal functions of instruction depends on partial decomposability (*24*) and some independence of effects of instructional variables: Designs optimal for one function (e.g., memory) should not be detrimental to another (e.g., induction). To illustrate, consider that facts require memory but not induction; thus, a designer can focus just on the subset of instructional techniques that facilitate memory.

	Principle	Description of Typical Effect
Memory/Fluency	Spacing	Space practice across time > mass practice all at once
	Scaffolding	Sequence instruction toward higher goals > no sequencing
	Exam expectations	Students expect to be tested > no testing expected
	Testing	Quiz for retrieval practice > study same material
	Segmenting	Present lesson in learner-paced segments > as a continuous unit
	Feedback	Provide feedback during learning > no feedback provided
Induction/Refinement	Pretraining	Practice key prior skills before lesson > jump in
	Worked example	Worked examples + problem-solving practice > practice alone
	Concreteness fading	Concrete to abstract representations > starting with abstract
	Guided attention	Words include cues about organization > no organization cues
	Linking	Integrate instructional components > no integration
	Goldilocks	Instruct at intermediate difficulty level > too hard or too easy
	Activate preconceptions	Cue student's prior knowledge > no prior knowledge cues
	Feedback timing	Immediate feedback on errors > delayed feedback
	Interleaving	Intermix practice on different skills > block practice all at once
	Application	Practice applying new knowledge > no application
	Variability	Practice with varied instances > similar instances
Sense-making/Understanding	Comparison	Compare multiple instances > only one instance
	Multimedia	Graphics + verbal descriptions > verbal descriptions alone
	Modality principle	Verbal descriptions presented in audio > in written form
	Redundancy	Verbal descriptions in audio > both audio & written
	Spatial contiguity	Present description next to image element described > separated
	Temporal contiguity	Present audio & image element at the same time > separated
	Coherence	Extraneous words, pictures, sounds excluded > included
	Anchored learning	Real-world problems > abstract problems
	Metacognition	Metacognition supported > no support for metacognition
	Explanation	Prompt for self-explanation > give explanation > no prompt
	Questioning	Time for reflection & questioning > instruction alone
	Cognitive dissonance	Present incorrect or alternate perspectives > only correct
	Interest	Instruction relevant to student interests > not relevant

Table 1**Instructional design principles**. These address three different functions of instruction: memory, induction, and sense-making (see table S1).

Theoretical work can offer insight into when an instructional choice is dependent on a learning function. Computational models that learn like human students do demonstrate, for instance, that interleaving problems of different kinds functions to improve learning of when to use a principle or procedure (*25*), whereas blocking similar problems types ("one subgoal at a time") improves learning of how to execute (*26*).

2. *Experimental tests of instructional function decomposability.* Optimal instructional choices may be function-specific, given variation across studies of instructional techniques where results are dependent on the nature of the knowledge goals. For example, if the instructional goal is long-term retention (an outcome function) of a fact (a knowledge function), then better memory processes (a learning function) are required; more testing than study will optimize these functions. If the instructional goal is transfer (a different outcome function) of a general skill (a different knowledge function), then better induction processes (a different learning function) are required; more worked example study will optimize these functions. The ideal experiment to test this hypothesis is a two-factor study that varies the knowledge content (fact-learning versus general skill) and instructional strategy (example study versus testing). More experiments are needed that differentiate how different instructional techniques enhance different learning functions.

3. *Massive online multifactor studies.* Massive online experiments involve thousands of participants and vary many factors at once. Such studies (*27*, *28*) can accelerate accumulation of data that can drive instructional theory development. The point is to test

hypotheses that identify, in context of a particular instructional function, what instructional dimensions can or cannot be treated independently.

Past studies have emphasized near-term effects of variations in user interface features (*27*, *28*). Designing massive online studies that vary multiple instructional techniques is feasible, but convenient access to long-term outcome variables is an unsolved problem. Proximal variables measuring student engagement and local performance are easy to collect (e.g., how long a game or online course is used; proportion correct within it). But measures of students' local performance and their judgments of learning are sometimes unrelated, or even negatively correlated, with desired long-term learning outcomes (*29*).

4. *Learning data infrastructure.* Massive instructional experiments are essentially going on all the time in schools and colleges. Because collecting data on such activities is expensive, variations in instructional techniques are rarely tracked and associated with student outcomes. Yet, technology is increasingly providing low-cost instruments to evaluate the learning experience for data collection. Investment is needed in infrastructure to facilitate large-scale data collection, access, and use, particularly in urban and low-income school districts. Two current efforts include LearnLab's huge educational technology data repository (*30*) and the Gates Foundation's Shared Learning Infrastructure (*31*).

5. *School-researcher partnerships.* Ongoing collaborative problem-solving partnerships are needed to facilitate interaction between researchers, practitioners, and school administrators. When school cooperation is well-managed and most or all of an experiment is computer-based, large well-controlled "in vivo" experiments can be run in courses with substantially less effort than an analogous lab study.

A lab-derived principle may not scale to real courses because nonmanipulated variables may change from the lab to a real course, which may change learning results. In in vivo experiments, these background conditions are not arbitrarily chosen by the researchers but instead are determined by the existing context. Thus, they enable detection of generalization limits more quickly before moving to long, expensive randomized field trials.

School-researcher partnerships are useful not only for facilitating experimentation in real learning contexts but also for designing and implementing new studies that address practitioner needs (*32*, *33*).

In addition to school administrators and practitioners, partnerships must include critical research perspectives, including domain specialists (e.g., biologists and physicists); learning scientists (e.g., psychologists and human-computer interface experts); and education researchers (e.g., physics and math educators). It is important to forge compromises between the control desired by researchers and the flexibility demanded by real-world classrooms. Practitioners and education researchers may involve more domain specialists and psychologists in design-based research, in which iterative changes are made to instruction in a closely observed, natural learning environment in order to examine effects of multiple factors within the classroom (*34*).

Our recommendations would require reexamination of assumptions about the types of research that are useful. We see promise in sustained science-practice infrastructure funding programs, creation of new learning science programs at universities, and emergence of new fields (*35*, *36*). These and other efforts are needed to bring the full potential of science and technology to bear on optimizing educational outcomes.

References and Notes

1. M. Dynarski *et al.*, *Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort* [Report provided to Congress by the National Center for Education Evaluation, Institute of Education Sciences (IES), Washington, DC, 2007].
2. Coalition for Evidence-Based Policy, Randomized Controlled Trials Commissioned by the IES since 2002: How Many Found Positive Versus Weak or No Effects; http://coalition4evidence.org/wp-content/uploads/2013/06/IES-Commissioned-RCTs-positive-vs-weak-or-null-findings-7-2013.pdf.
3. J. Roschelle, N. Shechtman, D. Tatar, S. Hegedus, B. Hopkins, S. Empson, J. Knudsen, L. P. Gallagher, Integration of Technology, Curriculum, and Professional Development for Advancing Middle School Mathematics: Three Large-Scale Studies. *Am. Educ. Res. J.* **47**, 833–878 (2010).
4. J. F. Pane, B. A. Griffin, D. F. McCaffrey, R. Karam, *Effectiveness of Cognitive Tutor Algebra I at Scale* (Working paper, Rand Corp., Alexandria, VA, 2013); www.rand.org/content/dam/rand/pubs/working_papers/WR900/WR984/RAND_WR984.pdf.
5. D. Klahr, J. Li, Cognitive Research and Elementary Science Instruction: From the Laboratory, to the Classroom, and Back. *J. Sci. Educ. Technol.* **14**, 217–238 (2005).
6. D. Klahr, What do we mean? On the importance of not abandoning scientific rigor when talking about science education. *Proc. Natl. Acad. Sci. U.S.A.* **110** (Suppl 3), 14075–14080 (2013).
7. A. T. Corbett, J. R. Anderson, *Proceedings of ACM CHI 2001* (ACM Press, New York, 2001), pp. 245–252.
8. R. A. Schmidt, R. A. Bjork, New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychol. Sci.* **3**, 207–217 (1992).
9. A. Paivio, Abstractness, imagery, and meaningfulness in paired-associate learning. *J. Verbal Learn. Verbal Behav.* **4**, 32–38 (1965).
10. J. A. Kaminski, V. M. Sloutsky, A. F. Heckler, Learning theory. The advantage of abstract examples in learning math. *Science* **320**, 454–455 (2008).
11. G. Wulf, C. H. Shea, Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon. Bull. Rev.* **9**, 185–211 (2002).
12. H. Pashler *et al.*, *Organizing Instruction and Study to Improve Student Learning* (National Center for Education Research 2007-2004, U.S. Department of Education, Washington, DC, 2007).
13. R. Wylie, K. R. Koedinger, T. Mitamura, *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (CSS, Wheat Ridge, CO, 2009), pp. 1300–1305.
14. R. E. Goska, P. L. Ackerman, An aptitude-treatment interaction approach to transfer within training. *J. Educ. Psychol.* **88**, 249–259 (1996).
15. S. Kalyuga, Enhancing Instructional Efficiency of Interactive E-learning Environments: A Cognitive Load Perspective. *Educ. Psychol. Rev.* **19**, 387–399 (2007).
16. K. A. Dunlosky, E. J. Rawson, E. J. Marsh, M. J. Nathan, D. T. Willingham, Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychol. Sci. Public Interest* **14**, 4–58 (2013).
17. S. Kalyuga, P. Ayres, P. Chandler, J. Sweller, The Expertise Reversal Effect. *Educ. Psychol.* **38**, 23–31 (2003).
18. R. L. Goldstone, J. Y. Son, The Transfer of Scientific Principles Using Concrete and Idealized Simulations. *J. Learn. Sci.* **14**, 69–110 (2005).
19. P. A. Wo niak, E. J. Gorzela czyk, Optimization of repetition spacing in the practice of learning. *Acta Neurobiol. Exp. (Warsz.)* **54**, 59–62 (1994).
20. P. I. Pavlik, J. R. Anderson, Using a model to compute the optimal schedule of practice. *J. Exp. Psychol. Appl.* **14**, 101–117 (2008).
21. K. R. Koedinger, A. T. Corbett, C. Perfetti, The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cogn. Sci.* **36**, 757–798 (2012).
22. L. Resnick, C. Asterhan, S. Clarke, Eds., *Socializing Intelligence through Academic Talk and Dialogue* (American Educational Research Association, Washington, DC, 2013).
23. G. Bradshaw, in *Cognitive Models of Science*, R. Giere and H. Feigl, Eds. (University of Minnesota Press, Minneapolis, 1992), pp. 239–250.
24. H. A. Simon, *Sciences of the Artificial* (MIT Press, Cambridge, MA, 1969).
25. N. Li, W. Cohen, K. Koedinger, Problem Order Implications for Learning Transfer. *Lect. Notes Comput. Sci.* **7315**, 185–194 (2012).
26. K. VanLehn, *Artif. Intell.* **31**, 1 (1987).
27. D. Lomas, K. Patel, J. Forlizzi, K. R. Koedinger, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, 2013), pp. 89–98.
28. E. Andersen, Y. Liu, R. Snider, R. Szeto, Z. Popovi , *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, 2011), pp. 1275–1278.
29. R. A. Bjork, J. Dunlosky, N. Kornell, Self-regulated learning: beliefs, techniques, and illusions. *Annu. Rev. Psychol.* **64**, 417–444 (2013).
30. LearnLab, Pittsburgh Science of learning Center; www.learnlab.org/technologies/datashop/index.php.
31. InBloom, www.inbloom.org.
32. Strategic Education Research Partnership, www.serpinstitute.org/about/overview.php.
33. IES, U.S. Department of Education, Researcher-Practitioner Partnerships in Education Research, Catalog of Federal Domestic Assistance 84.305H; http://ies.ed.gov/funding/ncer_rfas/partnerships.asp.
34. S. A. Barab, in *Handbook of the Learning Sciences*, K. Sawyer, Ed. (Cambridge Univ. Press, Cambridge, 2006), pp. 153–170.
35. Society for Research on Educational Effectiveness, www.sree.org/.
36. International Educational Data Mining Society, www.educationaldatamining.org/.

Acknowledgments: The authors receive support from NSF grant SBE-0836012 and Department of Education grants R305A100074 and R305A100404. The views expressed are those of the authors and do not represent those of the funders. Thanks to V. Aleven, A. Fisher, N. Newcombe, S. Donovan, and T. Shipley for comments.

Supplementary Materials

www.sciencemag.org/content/vol/issu/page/suppl/DC1

10.1126/science.1238056