# What do we mean? On the importance of not abandoning scientific rigor when talking about science education

### David Klahr<sup>1</sup>

PNAS

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213

Edited by Baruch Fischhoff, Carnegie Mellon University, Pittsburgh, PA, and accepted by the Editorial Board January 31, 2013 (received for review September 26, 2012)

Although the "science of science communication" usually refers to the flow of scientific knowledge from scientists to the public, scientists direct most of their communications not to the public, but instead to other scientists in their field. This paper presents a case study on this understudied type of communication: within a discipline, among its practitioners. I argue that many of the contentious disagreements that exist today in the field in which I conduct my research—early science education—derive from a lack of operational definitions, such that when competing claims are made for the efficacy of one type of science instruction vs. another, the arguments are hopelessly disjointed. The aim of the paper is not to resolve the current claims and counterclaims about the most effective pedagogies in science education, but rather to note that the assessment of one approach vs. the other is all too often defended on the basis of strongly held beliefs, rather than on the results of replicable experiments, designed around operational definitions of the teaching methods being investigated. A detailed example of operational definitions from my own research on elementary school science instruction is provided. In addition, the paper addresses the issue of how casual use of labels-both within the discipline and when communicating with the public-may inadvertently "undo" the benefits of operational definitions.

The 2012 Sackler Colloquium on the "Science of Science Communication" attracted >350 attendees and many more (~1,000) who viewed the presentations on the web, in real time and subsequently. [As of mid-January 2013, the number of hits on the individual talks ranged from ~100 to several thousand (www.youtube.com/user/sacklercolloquia?feature=results\_main).] Clearly, there is substantial interest in science communication among both scientists and their various publics. Although it is of utmost importance for scientists to be able to communicate their knowledge, their discovery and validation processes, and their uncertainties as clearly as possible to the public, it is equally important for communication within the community of scientific practitioners to be maximally lucid because there are undesirable ramifications for both scientific practice and public understanding when scientists within a discipline fail to communicate unambiguously among themselves.

One fundamental feature that makes science "science" is the operational definition, but, as I will demonstrate, many of the contentious disagreements that exist today in the field in which I conduct my research—early science education derive from a lack of operational definitions. Consequently, when competing claims are made for the efficacy of one type of science instruction vs. another, the arguments are hopelessly disjointed because of the frequent failure to provide clear definitions about the types of instruction that are being compared.

# When Scientists Talk to Scientists They Should Speak the Same Language

In the mature sciences—e.g., physics, chemistry, and astronomy there are extremely clear norms and conventions for talking about

procedures, methods, theories, underlying mechanisms, data presentation, protocols, etc. However, in the behavioral sciences, there is often less consistency and much wider variation with respect to the terminology used to describe theories, experimental paradigms, explanatory mechanisms, and research practices (1). This problem is particularly acute in the corner of the behavioral sciences that is related to education research, a field recently identified as the "education sciences"\*. In fact, the former Dean of the Harvard Graduate School of Education entitled her intriguing monograph on the history of Education Research "An Elusive Science." (2). The communication challenges associated with education research are probably best known in the areas of math and reading education, where the notoriously contentious disagreements about different pedagogical methodologies and different interpretations of the data have come to be known as "the math wars" (3) and "the reading wars" (4). However, in this paper, I will focus on the communication challenges within the very field where one would expect the clearest and least ambiguous terminology-within science education itself-and I will suggest that we are in the midst of an unfortunate enterprise that can be called, by analogy, "the science wars."

The basic problem was succinctly stated by Handelsman et al. (ref. 5, p. 521):

Why do outstanding scientists who demand rigorous proof for scientific assertions in their research continue to use and, indeed, defend on the basis of their intuition alone, teaching methods that are not the most effective?

The specific lament in Handelsman et al.is the claim that much of science education is based on a traditional form of didactic lecturing, which they argue is "not the most effective" way to teach science. Their paper goes on to advocate a more engaging, interactive "discovery-oriented" instructional approach. Unfortunately, nearly 10 y later, one could just as well use that very same statement to criticize the current enthusiasm for "inquiry approaches" to science education. For example, an influential National Research Council report on inquiry approaches to science education states that (ref. 6, p. 125):

<sup>1</sup>E-mail: klahr@cmu.edu.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "The Science of Science Communication," held May 21–22, 2012, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS Web site at www.nasonline.org/sciencecommunication.

Author contributions: D.K. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. B.F. is a guest editor invited by the Editorial Board.

<sup>\*</sup>Whitehurst G (2003) The Institute of Education Sciences: New Wine, New Bottles. American Educational Research Association, 2003 Annual Meeting Presidential Invited Session, April 22, 2003. Available at www2.ed.gov/rschstat/research/pubs/ies.pdf. Accessed September 26, 2012.

For example, studies of inquiry-oriented curriculum programs ... demonstrated significant positive effects on various quantitative measures, including cognitive achievement, process skills, and attitudes toward science.

This summary of prior research would seem to be clear evidence in support of inquiry approaches to science instruction, except for the fact that the report goes on to note, parenthetically, that "there was essentially no correlation between positive results and expert ratings of the degree of inquiry in the materials." Thus, we have an argument for the benefits of a particular pedagogy, but no consensus from experts about the "dose response," i.e., the extent to which different "degrees of inquiry" lead to different amounts of learning.

Criticism of the purported universal superiority of constructivist approaches to science teaching has been growing over the past decade (7, 8), as have arguments supporting it (9), culminating in an entire volume of pro and con perspectives on the issue (10). [In the interest of full disclosure, I have a chapter in that volume favoring "direct instruction" (11).] However, my aim in this paper is not to resolve the current claims and counterclaims, but rather to note that the assessment of one approach vs. the other is all too often defended, as Handelsman et al. (5) put it, "on the basis of ... intuition alone," rather than on the results of replicable experiments, designed around operational definitions of the teaching methods being investigated.

#### Needed: Operational Definitions of Instructional Practices

Not only is there a dearth of such information in the many comparative investigations of different approaches to science instruction, but even more troubling is the fact that some educational researchers with a constructivist perspective dismiss the very possibility that one can do rigorous comparative research on the topic of effective science teaching. For example, in a chapter in the Tobias and Duffy (10) compendium cited above, Jonassen (ref. 12, p. 29) argues that

high-quality research studies comparing the effectiveness of inquiry methods and direct instruction ... probably do not exist and cannot exist. Researchers examining the effectiveness of direct instruction begin with fundamentally different assumptions, evoke significantly different theory biases, and use different research methods than researchers examining informal or inquiry learning. Therefore, the questions they ask, the learning outcomes they seek and the research tools and methods they use are also quite different.

Such a perspective makes it impossible to determine the relative effectiveness of different methods of science teaching. Absent clear definitions, how are we to determine the "winner" from among the set of instructional "approaches" (in some instances, perhaps synonyms) that might include constructivism, explicit instruction, direct instruction, guided instruction, inquiry science, adaptive instruction, student centered instruction, authentic instruction, right-brain instruction, or hands-on instruction? To argue that different methods of science instruction are inherently incomparable removes the question of effective science education from the realm of science. However, it need not be so. One way to join the fragmented debate about optimal instructional methods in science education is to substantially clarify the relevant features of the different instructional manipulations; that is, the education research discourse must move from vague terminology and "isms" toward clear depictions of what is actually going on. This clarification process will require operational definitions of the instructional approaches being compared.

Absent such clarifications, it will remain extremely difficult to conduct the kinds of metaanalyses that are required if a signal is ever to emerge from the noise of the science wars. Those engaged in such endeavors consistently report a drastic reduction when going from the set of potentially relevant studies to those that are ultimately included in the metaanalysis, and a large proportion of this winnowing is caused by imprecise definitions of the instructional method. For example, two such recent metaanalyses of science teaching methods report reductions from 337 initial studies to 61 in the final statistical analysis (13) and from 59 down to 6 in another such metaanalysis of "inquiry based science teaching" (14).

# Operational Definitions and Shifting Terminology—A Cautionary Tale

In the next few paragraphs, I will provide a detailed example, from my own research, of operational definitions in an area of elementary school science instruction. In addition, I will address the issue of how casual use of labels may inadvertently "undo" the benefits of operational definitions.

What We Are Teaching. In the research to be described here, our instructional goal is to teach an important elementary-school science objective known in the literature on cognitive development as the control-of-variables strategy (CVS) (15). The procedural content of CVS instruction provides a method for creating experiments in which a single contrast is made between experimental conditions while "controlling" for all other potential causal factors. The conceptual content includes an understanding of the inherent indeterminacy of confounded experiments. In short, CVS is the basic procedure that enables children to design unconfounded experiments from which they can make valid causal inferences.

CVS transcends specific scientific domains (e.g., physics, biology, chemistry, etc.), because the underlying procedural and conceptual aspects of an unconfounded experiment are domaingeneral. Thus, its mastery is a critical step in the development of scientific reasoning skills because experimental methods play an essential role in constraining the space of potential causal factors (16, 17). CVS mastery is considered a central objective from a wide variety of educational perspectives (18-21), and it is invariably included in high-stakes science assessments such as the Trends in International Science and Mathematics Study (TIMMS) (22) and the National Assessment of Educational Progress (NAEP) (23). Its importance is indicated by the fact that NAEP has recently developed some very sophisticated interactive assessment methods for measuring children's knowledge of CVS at different grade levels (http://nationsreportcard.gov/science 2009/ ict tasks.asp). Clearly, anyone (child or adult) who has not mastered CVS will have difficulty understanding science communication about results based on experimental methods.

Of course, when actually teaching CVS, one has to situate it in a specific physical context that is relevant for a particular domain. Thus, one may be teaching the principles of good experimental design in the chemistry laboratory or in the physics laboratory, even though the primary goal at that point is for students to master CVS rather than to acquire some knowledge about the domain. All of the instructional experiments in my laboratory have used simple physical materials (such as balls on ramps, springs and weights, pendulums, or objects sinking in water).

Three Types of Instruction: Operational Definitions. To summarize: Our goal is to teach CVS, not domain knowledge, and the independent variable in our research is the method of instruction. In one of our first CVS studies (24), we compared the relative effectiveness of three different types of instruction for teaching CVS to third to fifth grade students. The three types ranged from explicit, teacher-directed, instruction to more open-ended learner-directed discovery. Note that in the previous sentence, I have necessarily used the very terminology ("teacher-directed," "learner-directed") that I criticized earlier for its inherent ambiguity. However, there is a solution to this problem, and that is to be extremely explicit about the criteria that the researcher is using to classify specific instructional procedures into each category. Furthermore, one can remove the baggage-laden terms and describe the three different instructional methods simply as types A, B, and C. The essential features of each of the three types of instruction are depicted in Fig. 1, where each column corresponds to one of the three types of instruction and each row describes a particular feature and its value. (In our full scientific report on this study (24), of course, each of the cell entries in Fig. 1 was augmented by a detailed "script" for how that component of the instruction was actually implemented, so that it could be replicated in other laboratories. Thus, Fig. 1 should be viewed as a schematic for the full operational definition of each type of instruction.)

Each column in Fig. 1 represents the essential features of a particular type of instruction. As indicated in the first two rows of the figure, for all types of instruction, children dealt with the same materials. For example, Fig. 2 depicts the ramps apparatus that we used; each ramp had four features that could take on one of two values. In all cases, (i) children were presented with the same goal: to design a "good experiment" (i.e., "Can you set up the ramps to find out for sure whether the height of the ramp makes a difference in how far the ball rolls?"), (ii) this goal was provided by the teacher, not generated by the student, and (iii) all conditions used "hands-on" instruction, as children manipulated the materials. (In addition to the ramps, during the training phase of the study, some children worked with springs, and others with sinking objects. On the "transfer" phases of the study, children switched to the two domains on which they were not initially trained. For example, children trained on springs were next asked to design experiments with ramps, and then with sinking objects. See ref. 24 for details.)

At this point, the different types of instruction began to diverge. In type A instruction, the teacher presented explicit instruction regarding CVS (i.e., how to design an unconfounded experiment by varying the "focal variable," such as the surface of the ramp) while making sure that all of the other variables (ramp height, type of ball, and length of the run) were held constant on each ramp. In types B and C instruction, the student, not the teacher, designed the experiment. Next, in types A and B instruction, students were presented with probe questions: "Is this a smart way to find out whether the surface of the ramp makes a difference?" "Can you 'tell for sure' from this experiment whether <the variable being tested> makes a difference in the outcome?" "Why are you sure or not sure?" As shown in Fig. 1, there was no corresponding probe question in type C instruction. Other crucial features of instruction, and their presence or absence in each particular type of instruction, are indicated in the



## Three types of instruction

**Fig. 1.** Key features of each of the three types of instruction used in ref. 22. Each column represents an instructional type, and each row corresponds to an important feature of that type of instruction.

remaining rows of Fig. 1. As mentioned, the description above and the figure are substantially condensed from the descriptions and details in our paper (24). However, the point is clear: Each column, and the associated elaboration of what its contents mean, provides an operational definition of the three types of instruction being contrasted in this study.

The results of this training experiment revealed that (i) only type A instruction led to immediate gains in children's mastery of CVS, and (ii) when tested on different physical materials several days later (such that children initially trained with ramps were now asked to design experiments with springs, and so on), children were able to transfer their CVS knowledge to materials with completely different physical dimensions. Other studies like this have shown that children presented with type A instruction remembered and used what they learned about CVS in substantially different contexts (i.e., they transferred their CVS knowledge) (25), and they retained it for several months, and even several years, after their instruction (26–28). Moreover, the explicit steps involved in our type A instruction have been transformed into a computer tutor that is as effective as a human tutor for CVS (29).

**Time Matters.** The challenge of creating effective operational definitions about instructional methods is exacerbated in contexts in which there are radically different theoretical perspectives. If one theory ignores a potentially causal factor that another theory views as causal, then there will be disagreements about the sufficiency of the operational definition. The controversies surrounding methods of science instruction exemplify this problem, because although the duration of any particular instructional method must certainly have an impact on its effectiveness, the temporal properties of instructional treatments are often inadequately described, or not mentioned at all. Thus, to the extent that different learning processes are involved when students are engaged in a 10-min lesson vs. a 10-mo curriculum, an operational definition of an instructional intervention that does not include temporal information is seriously deficient.

Indeed, there are deep theoretical reasons for assuming that the temporal duration of an instructional intervention plays a central role in what students learn from it. Allen Newell (30) proposed a categorization of psychological processes based on the range of time scales they traversed, and Nathan and Alibali (ref. 31, pp. 34–35) revisited time scale in the domain of instruction:

A time-scale analysis ... shows how disparate research traditions ... can be conceptualized within a unifying framework for the study of learning and complex behavior. At time scales below 10 ms, intellectual behaviors are at biological (primarily neural) levels of operation. In the next band of our human time scale, from 100 ms to 10 s, behaviors transition into the cognitive band, and include perceptual and motor processes, as well as basic and complex mental processes ranging from word and object recognition to brief communicative exchanges.

The next band, from minutes to ... hours, addresses behavior that is more planful, ... and task oriented. ... Human behavior at the further reaches of the next band (hours to days) is characterized primarily by social and developmental operations, such as experiences with classroom or on-the-job training over whole class periods or training units spanning several days.

Given this perspective, it is clear that a 15-min lesson toward the "direct instruction" end of the spectrum is highly likely to have different effects than several hours, or weeks or months, of the same type of instruction. Accordingly, the important theoretical constructs used to make predictions and account for the effects of instruction are likely to be quite different at these two points on the instructional duration time scale.

**Mea Culpa.** Indeed, in many of my reports about our research on teaching CVS, there is no explicit mention of the duration of the different types of instruction being compared. Consequently,

S A Z C



Fig. 2. Ramps. On each of the two ramps, children can vary the height of the ramp, its, surface, and its length, and the type of ball. In the completely confounded experiment depicted here, the contrast is between a golf ball on a steep, smooth, short ramp (A) and a rubber ball on a low, rough, long ramp (B). Reproduced with permission from ref. 24. Copyright 1999 by the Society for Research in Child Development, Inc.

someone attempting to replicate or modify our instructional treatments would have no explicit guidance on how long they should take. They might infer (correctly in this instance) from the detailed descriptions of the procedures, as well as from their own familiarity with instructional procedures with young children, that the three treatments were of approximately equal duration of ~15–20 min each. However, it is not always so easy to make such assumptions about the temporal duration of "head-to-head" comparisons of instructional treatments. Moreover, one might expect different results from comparisons of two different 20-min instructional treatments and from two different 20-wk treatments having the same contrasting features. Thus, the absence of explicit temporal information renders a full operational definition of instructional methods incomplete.

What's in a Name? One important part of any definition, and particularly an operational definition, is the name given to the construct being defined. In sciences that are still in the process of developing unambiguous operational definitions, the name may carry unintended excess baggage over and above the specifics of the operational definition. Moreover, to the extent that the terminology may be widely used in everyday language, it may be interpreted in different ways by different people. Of course, even physics has this problem with the distinction between welldefined terms and their casual use in everyday language: compare "mass" vs. "weight," "velocity" vs. "acceleration," and "heat" vs. "temperature." However, there is no confusion within the discipline about the operational definitions of these constructs.

Thus, in our first report on our three types of training, we deliberately used somewhat inelegant terminology that sought to capture the essence of the operational difference between the three types of training, without being overly cumbersome. We dubbed type A, B, and C instruction "training-probe," "notraining-probe," and "probe," respectively. However, in several subsequent studies, in which we simplified the instructional contrasts to just two types (types A and C), we began to call type A "direct instruction" and type C "discovery learning." The consistent finding in several variations on our basic training study design was that in these studies, direct instruction about CVS produced more learning than discovery learning. For example in one study (28), after a 20-min training session, 75% of the students in the direct instruction condition mastered CVS, whereas only 25% of the students in the discovery condition did so. We also found that, when challenged to transfer their CVS knowledge to an "authentic" task-judging science fair posters created by other children—the children who had mastered CVS in the training phase were much better judges than those who had not mastered CVS—regardless of how they had been instructed. That is, when asked to make broader, richer scientific judgments, the many children who learned CVS via direct instruction performed as well as those few children who discovered the method on their own. There was no particular advantage to having "discovered" CVS rather than having been "told" CVS.

Nevertheless, although these results seemed to indicate that we had identified an effective instructional procedure for teaching voung children how to master CVS, the everyday labels we had begun to use led to substantial disagreement within the field about which of our conditions was really direct instruction, which was really discovery learning, and whether one or the other was a parody of the corresponding method (32, 33). For example, it was suggested that, although the "direct instruction" label is acceptable for an approach in which the teacher designs and summarizes the experiment (as in our type A instruction), that label should not be used in a situation that also includes probe questions (and student replies) as in our type A instruction (Fig. 1). Critics argued that because such interactive engagement with students begins to move from the "talking head" approach often associated with direct instruction toward a type of guided discovery, our type A instruction involves more engagement with the student than is commonly allowed in "pure direct instruction." Such criticisms often suggested that a fair test of direct instruction would have included a fourth condition in the Chen and Klahr (24) study that eliminated probe questions, explanations, and summaries, thus creating a more widely accepted example of "pure" direct instruction. Similarly, with respect to our referring to type C instruction as discovery learning, critics have argued that even the most extreme forms of discovery would include some type of probe questions and guided summaries of student learning, whereas our type C has none of those features. As noted earlier, disagreements about what the "correct" label should be for different types of science instruction abound. Given that each type of instruction shown in Fig. 2 has several attributes, it is clear that there is a large space of potential experimental contrasts, and much research remains to be done to discover the instructional efficacy of different points in the space. However, the point of this paper is that no matter what the findings, they need to be reported in terms of operational definitions, rather than vague labels.

Although I have attempted to convey the serious challenges to advances in our field because of widespread terminological ambiguities, I see indications that others are beginning to argue for the necessity of operational definitions and to use them in their scientific reports. For example, in the introduction to an extensive study of the relative effectiveness of different approaches to CVS instruction, Cobern et al. (ref. 34, p. 8) say the following:

We aimed to describe the various components of the research project fully enough that readers or researchers would in principle know specifically what was done, how, and why, and thus could check or replicate aspects of the study if desired. For example, we do not use vague verbal tags like "inquiry" or "direct" for our instruction. Single words or short phrases cannot possibly encapsulate all aspects and variants of an educational concept or setting, and different people will ascribe different meanings and interpretations to such terms, leading to miscommunication and confusion, often unrecognized. Instead we provide operational definitions or models of exactly what we mean and what we did. We do the same for the assessment and its alignment with objectives and instruction.

#### **Communicating with the Media**

S A Z C

In addition to the challenges of within-discipline communication about our science, there is the question of how to describe the methods, results, and implications of our research to the media. My own experience has taught me that the correct reply to the question of how to communicate with the media is "with great caution." Because of their relevance to some of the ongoing disputes about inquiry approaches to science education, our studies of how to best teach CVS (a small—albeit nontrivial—part of the overall science curriculum) attracted some media attention (35–43).

Although it would be disingenuous to claim that I did not find this attention gratifying, it was also troubling because it raised the challenge of ensuring that the translation from the laboratory to the public maintained at least a modicum of accuracy and coherence. Sometimes, such clarity can be achieved by working with highly accommodating reporters and reviewing the accuracy of their drafts. However, the realities of deadlines, as well as the professional practices and constraints of the media, often work against such collaborative refinement of media reports, even for science writers of the highest skill and integrity. Another tactic is to request space in the publication for review and reassessment of the material in the media report. That is very rare, although I did have one chance to do that in *Education Week* in response to

- Bazerman C (2000) Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science (Univ of Wisconsin Press, Madison).
- Lagemann EC (2000) An Elusive Science: The Troubling History of Educational Research (Univ of Chicago Press, Chicago).
- 3. Klein D (2007) A quarter century of US 'math wars' and political partisanship. *Journal* of the British Society for the History of Mathematics 22(1):22–33.
- Anderson K (2000, June 18) The reading wars: Understanding the debate over how best to teach children to read (book review) LA Times.
- Handelsman J, et al. (2004) Education. Scientific teaching. Science 304(5670): 521–522.
- 6. National Research Council (2000) Inquiry and the National Science Education Standards: A Guide for Teaching and Learning (National Academies Press, Washington).
- Sweller J, Kirschner PA, Clark R (2007) Why minimally guided teaching techniques do not work: A reply to commentaries. *Educ Psychol* 42(2):115–121.
- Mayer RE (2004) Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *Am Psychol* 59(1):14–19.
- Gil-Pérez D, et al. (2002) Defending constructivism in science education. Sci Educ 11(6): 557–571.
- Tobias S, Duffy TM, eds (2009) Constructivist Theory Applied to Instruction: Success or Failure? (Taylor and Francis, London).
- Klahr D (2009) "To every thing there is a season, and a time to every purpose under the heavens": What about direct instruction? Constructivist Theory Applied to Instruction: Success or Failure? eds Tobias S, Duffy TM (Taylor and Francis, London).
- Jonassen D (2009) Reconciling a human cognitive architecture. Constructivist Instruction: Success or Failure? eds Tobias S, Duffy TM (Taylor and Francis, London), pp 13–33.
- Schroeder CM, Scott TP, Tolson H, Huang T-Y, Lee Y-H (2007) A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. J Res Sci Teach 44(10):1436–1460.

some articles and subsequent letters to the editor about our research involving the assessment of science fair posters (ref. 44, p. 36). I wrote:

[Critics] ... argue that the procedure followed in what we called the "discovery learning" condition is not representative of what is really recommended by discovery-learning advocates. ... [But] ... is it really so different? Our discovery condition presented the experimental apparatus to the children. It presented them with a goal, "see if you can set up the ramps to see if the height of the ramp makes a difference," and then students were free to explore, in a hands-on fashion, various kinds of arrangements, run the experiments, observe the results, and finally, under teacher suggestion, move on to another goal, such as "see if you can set up the ramps to see if the surface of the ramp makes a difference in how far the ball rolls." I would venture that this is not so far from what passes for discovery learning in many elementary school classrooms. . . . [another letter] . . . reiterates our argument for the need ... to make more precise use of terminology before moving on to policy decisions. Indeed, it is surprising that science educators so often abandon one of the foundations of science — the operational definition — when they engage in heated debates about discovery, inquiry, hands-on, and the rest. No science can advance without clear, unambiguous, operationally defined procedures. Neither can education science.

## **Approach Avoidance**

The terminological proliferation in the area of science education is daunting. It includes such "approaches" as: constructivism, explicit instruction, Piagetian approach, inquiry science, direct instruction, adaptive instruction, student-centered instruction, authentic instruction, hands-on instruction, didactic instruction, drill and kill, minds-on instruction, etc. However, these imprecise slogans convey little of substance because they are so loosely and multiply defined and interpreted. Specifying a "Newtonian approach" does not get you very far on the journey to Mars. Only a determined and consistent effort to better define instructional methods will ensure advances in the Education Sciences.

ACKNOWLEDGMENTS. I thank Robert Siegler, Stephanie Siler, Bryan Matlen, Sharon Carver, and Cathy Chase for their comments on earlier drafts. The research reported here was supported in part by Department of Education Institute of Education Science Grant R305A100404 (to Carnegie Mellon University) and in part by National Science Foundation Grant SBE-0836012. The opinions expressed are mine, and do not necessarily represent views of either of the granting agencies or of the colleagues who read and commented on earlier drafts.

- Furtak EM, Seidel T, Iverson H, Briggs DC (2012) Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Rev Educ Res* 82: 300–329.
- Case R (1974) Structures and strictures: Some functional limitations on the course of cognitive growth. Cognit Psychol 6:544–573.
- Klahr D (2000) Exploring Science: The Cognition and Development of Discovery Processes (MIT Press, Cambridge, MA).
- Klahr D, Simon HA (1999) Studies of scientific discovery: Complementary approaches and convergent findings. *Psychol Bull* 125:524–543.
- 18. DeBoer GE (1991) A History of Ideas in Science Education: Implications for Practice (Teachers College Press, New York).
- Duschl R (1990) Restructuring Science Education: The Importance of Theories and Their Development (Teachers College Press, New York).
- Murnane RJ, Raizen SA, eds (1988) Improving Indicators of the Quality of Science and Mathematics Education in Grades K-12 (National Academy Press, Washington).
- 21. National Research Council (1995) National Science Education Standards (National Academy Press, Washington).
- National Center for Education Statistics. (1999) Overview and key findings across grade levels. *Highlights from TIMSS: The Third International Mathematics and Science Study* (Dept of Educ, Washington). Available at http://nces.ed.gov/pubs99/1999081. pdf. Accessed September 26, 2012.
- National Center for Education Statistics, National Assessment of Educational Progress. Available at http://nces.ed.gov/nationsreportcard/. Accessed September 26, 2012.
- Chen Z, Klahr D (1999) All other things being equal: Children's acquisition of the control of variables strategy. *Child Dev* 70(5):1098–1120.
- Matlen B, Klahr D (2013) Sequential effects of high and low instructional guidance on children's acquisition and transfer of experimentation skills. *Instr Sci* 41(3): 621–634.

- Toth EE, Klahr D, Chen Z (2000) Bridging research and practice: A cognitively-based classroom intervention for teaching experimentation skills to elementary school children. Cogn Instr 18(4):423–459.
- Strand-Cary M, Klahr D (2008) Developing elementary science skills: instructional effectiveness and path independence. Cogn Dev 23:488–511.
- Klahr D, Nigam M (2004) The equivalence of learning paths in early science instruction: Effect of direct instruction and discovery learning. *Psychol Sci* 15(10): 661–667.
- Siler SA, Klahr D, Price N (2013) Investigating the mechanisms of learning from a constrained preparation for future learning activity. *Instr Sci* 41(1):191–216.
- 30. Newell A (1990) Unified Theories of Cognition (Harvard Univ Press, Cambridge, MA).
- 31. Nathan MJ, Alibali MW (2010) Learning sciences. WIREs Cogn Sci 1(3):329-345.

DNAS

SANC

- Klahr D (2010) Coming up for air: But is it oxygen or phlogiston? A response to Taber's review of constructivist instruction: Success or failure? *Educ Rev* 13(13): 1–6.
- 33. Taber KS (2010) Constructivist pedagogy is superior—it is a matter of definition. ADL Newsletter for Educators and Educational Researchers. (October). Available at http:// research.adlnet.gov/newsletter/academic/201010.htm#Klahr2010. Accessed September 26, 2012.

- Cobern WW, et al. (2012) Active Learning in Science: An Experimental Study of the Efficacy of Two Contrasting Modes of Instruction. Available at www.wmich.edu/ way2go/docs/Article-way2goWEBSITE.pdf. Accessed September 26, 2012.
- Adelson R (2004) Instruction versus exploration in science learning. APA Monit 35(6):34.
  Begley S (December 10, 2004) The best ways to make schoolchildren learn? We just don't know. Wall Street Journal, Section B, p. 1.
- Begley S (December 17, 2004) To improve education, we need clinical trials to show what works. *Wall Street Journal*, p B1.
- 38. Bower B (2009) Think like a scientist. Sci News 175(13):20.
- 39. Brown, E.A. (August 30, 2011) Experts review psychology of science teaching. *Educ Daily* 44(149):1, 3.
- Cavanagh S (2004) NCLB could alter science teaching. *Educ Week* 24(11): 1, 12–13.
  Crane E (March 1, 2005) The science storm: Some say direct instruction works best,
- while others argue for the discovery learning method. Surprisingly, the two camps are very close to being on the same team. *District Admin* 41(3):46.
- Dittmann M (2004) Studying ways to improve education: A two-year program has selected 13 fellows to conduct school-based research with leaders in the field. APA Monit 35(8):52.
- 43. Ezarik M (2004) Stand and deliver...or let them discover? District Admin 40(11):59
- 44. Klahr D (2005) A science study's author responds to critics. Educ Week 24(16):36.