Res Sci Educ (2018) 48:389–435
DOI 10.1007/s11165-016-9573-3

Learning of Core Disciplinary Ideas: Efficacy Comparison of Two Contrasting Modes of Science Instruction

David Schuster[1] · **William W. Cobern**[2] · **Betty AJ Adams**[2] · **Adriana Undreiu**[3] · **Brandy Pleasants**[2]

Published online: 25 May 2017

Abstract Science curricula and teaching methods vary greatly, depending in part on which facets of science are emphasized, e.g., core disciplinary ideas or science practices and process skills, and perspectives differ considerably on desirable pedagogies. Given the multi-faceted nature of science and the variety of teaching methods found in practice, it is no simple task to determine what teaching approaches might be most effective and for what purposes. Research into relative efficacy faces considerable challenges, with confounding factors, ambiguities, conflations, and lack of controls being threats to validity. We provide a conceptual framework characterizing the many teaching strategies found in practice as being variants of two fundamental contrasting epistemic modes, and we disentangle conflations of terms and confusions of constructs in both teaching practice and research. Instructional units for two science topics were developed in parallel in the alternative epistemic modes, differing in concept learning paths but otherwise equivalent. We conducted a randomized controlled study of the comparative efficacy of the two modes for learning core disciplinary ideas, using operationally defined active-direct and guided-inquiry teaching methods. Five middle school teachers taught each unit in both modes over 4 years of classroom trials in an 8-day summer program for eighth grade students. Student understanding of core ideas was assessed using pre- and post-tests, and learning gains were analyzed by mode, teacher, topic, and trial year. Although routes to concept understanding were very different in the two modes, eventual student learning gains were similar, within statistical variation. Efficacy variations between and within teachers were greater than between modes, indicating the importance of teacher effects on student achievement. Findings suggest that teachers need not be bound to one mode throughout and can flexibly decide on the pedagogical approach for each concept and situation, on several grounds other than efficacy of core content acquisition alone.

✉ David Schuster
david.schuster@wmich.edu

[1] Physics Department and Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI 49008, USA

[2] Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI, USA

[3] University of Virginia's College at Wise, Wise, VA 24293, USA

Keywords Science education · Teaching methods · Inquiry instruction · Direct instruction · Epistemic mode · Core disciplinary ideas · Randomized controlled trial · Efficacy of instructional methods · Active engagement · Meaningful learning · Conceptual understanding · Teacher effects

Introduction

Science education standards, curricula, teaching methods, and assessments have varied considerably over time and across countries, at both policy level and in actual teacher practice. Some countries have national curricula and others decentralize to the state or even school district level. Some standards specify science content alone, others also include science practices and the nature of science, and some may advocate specific teaching approaches. Notwithstanding formal standards, documents, curricula, and approaches are often defined de facto by textbooks, teaching resources, and assessment systems. There are also issues of what facets of science are taught or not, for example, science as a product or body of knowledge (facts, concepts, principles, models, theories); science as a process or creative endeavor (practices, process skills, experimenting, scientific inquiry, etc.); or science as problem solving and application (practical uses, devices, engineering, technology). Correspondingly, science curricula and instruction may emphasize or privilege one or another of these facets. In the past, the emphasis was mostly toward the content side, certainly as commonly assessed.

In recent decades, however, the goals for science education have come to include much in addition to content knowledge. Over 40 years ago, West and Fensham (1974) commented on the rising interest in teaching science through what we now call hands-on laboratory activities and practices — the process side of science. In the USA, this expansion of goals is now explicit in science education policy and standards documents such as *A Framework for K-12 Science Education* (NRC 2012) and the *Next Generation Science Standards* (NGSS) (NRC 2013). These explicitly emphasize both content and process aspects of science by referring to Core Disciplinary Ideas, Science Practices, and Cross-Cutting Concepts. Still, West and Fensham argued that even with broadened science education scope, it remained crucial that students learn science's "highly developed content of knowledge" (p. 61). However, recent national assessments of student science knowledge show that inadequate concept understanding in science remains a persistent problem. For example, on National Assessment of Educational Progress (NAEP) tests, only a third of eighth grade students were rated "proficient" in science in 2011 (Fleming 2012; NCES 2011; NCES 2012; USDOE 2012).

In researching and comparing instructional approaches and their effectiveness, it is important to recognize and distinguish the facets of science that might be the main focus of instruction. One focus would be the core disciplinary ideas, with the goal being understanding content within the organized structure of a topic. Another could be scientific inquiry, investigative practices, and aspects of the nature of science. Not surprisingly, one finds that curricula and instruction vary markedly on which facets of science are the focus, how they are taught, and whether or not content and process are taught together. For any given curriculum there will also be different interpretations and emphases in textbooks' treatments, and variations in how teachers teach particular topics in their own situations. In addition, there are factors associated with assessment, and to what extent this aligns with objectives and instruction. Even where the curriculum has broader aims, content knowledge is often the main focus of tests and grades.

It is therefore no easy task to decide what aspects of science shall be the main emphasis of science education in practice, nor is it a simple question to ask what teaching methods may be

the most effective for teaching and learning "science." Consequently, trying to answer this question through research poses considerable challenges.

Core Disciplinary Ideas

Our study focuses on teaching and learning *core disciplinary ideas*, i.e., the central concepts, principles, laws, models, and theories of a science topic. We set out to compare the efficacy of two distinct instructional approaches for developing conceptual understanding of core ideas. Meaningful understanding of such ideas is indicated by the ability to apply them to explain, predict, answer conceptual questions, and solve problems, and this is how we assess student achievement in this study. Science practices and process skills are *not* the research focus of our current work, although some of them will naturally be present in engaging lessons that reflect aspects of doing science. We do not research the efficacy of methods for teaching practices, but some other research does so, for example Klahr and Nigam (2004), Dean and Kuhn (2006), and Matlen and Klahr (2013) for the control of variables strategy. One cannot assume that the same methods are necessarily best for teaching both content and process, nor that both of these facets must always be learned together at the same time in the same lesson. In fact, Hattie's (2009) synthesis of over 800 meta-analyses reports that different instructional methods have different degrees of effectiveness for learning science content or science process skills. We focus on the learning of core concepts in a coherent disciplinary structure, and not for example on investigation projects which emphasize science process skills and apply some science to answer particular questions.

Two Contrasting Epistemic Approaches

Given the diversity of science curricula, goals, teaching methods, and their multiple implementations in practice, it may seem daunting to try to evaluate and compare the effectiveness of instructional methods in a meaningful way. Nevertheless, for teaching and learning core disciplinary ideas, we argue that the many methods and strategies found in practice may all be seen as variants of two fundamental epistemic[1] approaches or modes. The essential distinction is that in one approach the core concepts and principles are presented and explained by the instructor as established science knowledge, while in the other the core ideas are co-developed by students and instructor in guided-inquiry fashion, through questions, observation, and exploration. In simple form, one may say that the difference between the two basic approaches lies in "how students come to the concept." Different learning paths are therefore taken in the two modes.

The two epistemic modes are almost inverses of each other in framing, sequencing, narrative, and epistemology, although they have the same end goal of conceptual understanding of core ideas. In the first epistemic approach, the core concepts, principles, and theory for a topic are explicitly presented, defined, and explained to learners, illustrated by demonstrations and example cases. This "theory" component of the topic is generally followed by a "practical" or "lab" component comprised of experimental activities where students aim to test and verify the theory, often working in groups. In this approach, the core ideas are treated as *ready-made science*, and students work with these ideas in tasks, activities, and problems. In the alternative epistemic approach a science topic or phenomenon is approached via focus

[1] Epistemic: of or relating to knowledge and cognition; of or relating to knowledge and the conditions for acquiring or validating it.

questions and exploration, and the relevant concepts or laws are developed to account for observations and evidence. In this way, the core ideas and laws are "invented" or "discovered" by students and teacher together, in a concept-formation process stemming from a perceived need. Concept names are ideally introduced only *after* the concepts themselves have been grasped. This approach casts learning as *science-in-the-making*, and aims to reflect not only what we know but how we come to know it. In terms of schema theory, these alternative learning paths reflect different ways for a learner to build schemata elements and connections for and around the concept involved. We might say an inquiry mode sequence lays cognitive and experiential groundwork for the introduction of the formal science concept. The two contrasting modes are also reflected in different teaching/learning scripts for a topic.

Note that both of the modes described above are intended to involve active student engagement with the subject matter. The *epistemic* distinction is not between passivity and engagement. Nor is it between "hands-on" or not, since practical activities can occur in both modes, though framed and sequenced differently. We also note th at inquiry is part of th eoretical science too. Both modes of instruction may take place in various formats, for example, whole-class learnng or small-group activities. Thus we see epistemic mode per se is not determined by class format, group size, or experimental work. These are all separate dimensions, and various combinations are possible. Furthermore, we do not consider one mode as intrinsically "teacher-centered" and the other "learner-centered" but prefer to think of both as learning-centered, involving teachers and students alike.

One way of clarifying the difference in concept learning paths is in terms of instructional narrative, i.e., how the "story" unfolds. The first mode goes early to the desired "end" of the story by providing the target knowledge upfront and explaining it, while the second starts at the "beginning" of the story and moves toward the target knowledge. Note that the distinctions are not primarily defined by specific classroom teaching techniques or methods that a teacher may use; these are best seen as supporting the chosen instructional narrative.

Terminology Issues

We have deliberately described the nature of the two epistemic approaches before "naming" them. With some trepidation, we will now refer to the first form of instruction as "direct mode" and the second as "inquiry mode." We did not do so at the outset because, in practice, people understand a wide range of different things by "direct" and "inquiry" instruction, and bring their own meanings to the terms, no matter our intent. These terms may also have positive or negative connotations depending on one's existing perspective and educational/ideological biases, and the distinction between them may not have been considered with reference to learning core science content versus science process skills. Single words like "direct" or "inquiry" certainly cannot convey the complexity and nuances of the constructs, let alone the range of meanings and methods. To exacerbate the terminology issue, a number of different constructs are commonly conflated under each name. Since the essential characteristics of each approach matter more than the names, we hope that by elaborating their nature in some detail later, and using operational definitions ourselves, we can obviate some of these confusions. Klahr (2013) in a paper titled *What do we mean*? describes how the casual use of labels for types of science instruction has led to hopelessly disjointed arguments for competing claims for teaching methods, based mainly on strong beliefs rather than valid comparative evidence, and he advocates strongly for the use of operational definitions in the field of science education.

Purposes of This Study

In light of the multiple facets of science and the great variety of teaching methods, we posed an overarching question: How does choice of fundamental epistemic mode affect the understanding that learners attain of core disciplinary ideas? Our research focus is thus on the relative efficacy of operationally defined direct and inquiry instructional approaches for the teaching and learning of core science concepts and principles. To this end, we designed instructional units for two science topics in parallel in both modes and implemented them in classroom instruction to compare the student conceptual learning gains achieved. One might anticipate three different possible outcomes. Either the direct mode or the inquiry mode proves to be superior for acquiring understanding of core concepts, or the two modes prove comparable in this regard. Any of these outcomes would be important to establish. Note that if the modes turn out to be comparable, this would not mean there is no reason to choose one over the other, but rather that the choice could be made on other grounds, that ideally do not detract from the desired learning of core disciplinary content. The project has implications for making informed choices of instructional modes and methods in teaching science.

Epistemic Modes and Instructional Approaches

The general epistemic distinction noted above between the alternative modes will be reflected in the lesson structure, sequencing, and teaching methods adopted for any given topic. This being the case, a useful way of illustrating and contrasting the modes of instruction is through topic examples. Below, we provide two cases, one for a law and one for a concept, comparing lesson designs in the alternative modes.

Examples of Instructional Designs in Alternative Epistemic Modes

Newton's second law of motion forms part of a topic unit developed for this research, so using it as an example will serve to illustrate the nature of instruction in each mode. To study the relation between force, mass, and motion, one needs a suitable object to which one applies a force in a situation where friction is minimal so the object is free to move without significant resistance. In our lesson design for a whole-class participatory activity, the object was a student sitting on a wheeled skateboard on a smooth floor in the hallway. Another student applied a continuing constant pushing force on the back of the skateboard rider and during the ensuing motion the rider dropped markers on the floor at equal time intervals to provide a visual record of positions during the motion. The other students were lined up along the hallway, observing the skateboard motion while clapping to the beat of the time intervals to facilitate marker dropping. The nature of the motion recorded in this way could be seen from the pattern of markers on the floor, with greater spacing indicating greater speed. Below, we describe how direct mode and inquiry mode approaches to learning the law might compare in practice, using this apparatus and activity.

In a direct mode approach, the instructor presents Newton's second law of motion upfront to the students, stating and explaining the relation between force, mass, and motion. This is written on the board in a form such as this: "A net force on an object results in acceleration, which is directly proportional to net force and inversely proportional to the mass of the object." The instructor also writes this algebraically as $a = F/m$ or $F = ma$. The law and its dependencies are carefully explained by the instructor and illustrated using examples and demonstrations. Students can ask questions on the material or the instructor may pose them to check understanding.

Thereafter, in the practical component, students test and verify the law and its dependencies using the appropriate skateboard marker-dropping activities. This is a deductive testing process.

In inquiry mode, by contrast, students first consider the question of what kinds of motions might ensue if a force is applied to an object free to move. They are then guided to explore this experimentally using the skateboard activity. Finding that marker spacing increases, students can infer that a constant applied force results in accelerated motion (rather than constant speed). In this way, they arrive (inductively in this case) at a knowledge claim (proposed law) based on observational evidence and discussion, with instructor framing and guidance. Students can then investigate the dependencies in the same way, finally arriving at Newton's second law relating force, mass, and motion.

In the first approach, Newton's law appears formally near the beginning of the lesson sequence, while in the alternative approach the law is developed during the lesson and so appears toward the end of the sequence. The experimental activity is used as a concept testing activity in one instructional mode, and as a concept development activity in the other. The instructional narratives are different for the two modes, as are students' cognitive trajectories for learning the concepts. Note that for this topic, the experimental procedure also involves control of variables as a science practice, though this was not our main focus. Besides showing the mode distinction, the example above also serves to show the active-engagement type of lessons that we produced in both modes for our research. Thus, any differences in efficacy of these lessons for learning the law cannot simply be ascribed to active learning or hands-on activity.

The second example is for teaching a concept rather than a law. Consider the abstract concept of *density*. In direct mode, it is relatively straightforward for a teacher to define and explain density and have students apply it to particular cases. It is more challenging to invent the concept of density by inquiry. One approach might be to start with suitable sets of blocks of various sizes and masses and made of various materials, and guide students toward a physical concept that would be the same for all objects of the same material, irrespective of size and mass. This would turn out to be the ratio of mass to volume, also understood conceptually as the mass of a unit volume of the material. Developing a concept in this way can be quite a challenge for teachers and students alike. Teachers need deep pedagogical content knowledge for the topic, and students need strong conceptual guidance. Note that "inventing" an abstract concept in class may be even harder for novices than "discovering" a law. Concept introduction or formation might be fitting terms when thinking processes are strongly guided.

In both approaches, once the core concepts have been acquired, there is an application phase where students apply the concepts to questions and problems. This is common to virtually all good science instruction for concept development.

Characteristics of Contrasting Epistemic Modes

The two modes constitute underlying epistemic approaches which may be implemented in lessons in various ways, depending on topic, nature of concept, stage of instruction, degree of guidance, specific methods and strategies, etc. Despite this variety, the essential distinction between direct and inquiry modes of instruction is typically evident. The question of interest for our research was to what extent the mode difference in topic framing and learning trajectory may affect meaningful student acquisition of core ideas. For the efficacy study, we compared two operationally defined modes which we call *active-direct* and *guided-inquiry*, differing in epistemic approach but both involving active student engagement. To investigate

using a fair comparison, we developed contrasting sets of lessons, designed in parallel, and cast in the alternative modes with contrasting narratives, but otherwise equivalent and both involving hands-on activities.

Various detailed mode characteristics informing instructional designs in the two contrasting modes are presented and discussed below. The alternative epistemic approaches have several characterizing features and differ along a number of dimensions. Several of these features may contribute to concept learning effectiveness, although not all may apply in every case. A general dissection and comparison of the nature of the two approaches will thus be useful, and we discuss the relevant characteristics below.

1. *Learning trajectories.* As noted above, the epistemic approaches differ in "how students come to the concept." In one approach they are "told" it by the teacher or textbook, while in the other they construct it, guided by the teacher. These constitute alternative learning paths or cognitive learning trajectories for concept acquisition. This difference in how things are encountered and learned leads to interesting questions. If learners are guided to come to the ideas themselves rather than being told, will they understand and retain them better? Will a concept make more sense to learners if they see its origin and the need for it? Irrespective, thinking it out could be an important kind of thinking skill in its own right. Conversely, might concept learning be clearer with less confusion and side tracks if the correct ideas are clearly presented and explained right from the start? We also keep open the possibility that alternative paths to learning a concept might make a difference in the concept introduction stages but little difference in eventual concept understanding.
2. *What we know and how we know.* An inquiry-based treatment by its nature addresses the question of "how we know" as well as "what we know." Direct treatments may or may not talk of "how we know." A potential concern is whether or not there is any trade-off between helping students experience "how we know" on every core idea, and helping students develop a thorough conceptual grasp of the core ideas themselves. There is also the issue of whether learning content and process at the same time can be cognitive overload. Again, it is possible that different approaches might be more or less appropriate or effective depending on the topic and the learning goal (Hattie 2009).
3. *The general and specific, the abstract and the concrete.* Direct instruction tends to go from the general to the specific, while the reverse is usually true for inquiry. Direct often goes from an abstract final form to concrete examples of it, while inquiry instruction often starts from concrete cases and generates the abstraction. It may be that exposure to specifics before generalities or vice versa affects concept learning, and might affect it differently for different learners.
4. *Deductive or inductive.* Going from the general to the specific is generally characteristic of deductive methods, with the opposite being true of inductive methods; however, this alone does not necessarily make instructional approaches purely deductive or inductive. For some topics this may be the case, for example, when observation of specific instances leads by simple induction to a generalization. However, not all science concepts are of this nature, nor arrived at this way, so a simple deductive vs. inductive distinction may be misleading. Deductive and inductive modes of reasoning occur throughout science itself and often in combination, a fact perhaps less recognized than it should be. Scientific inquiry is certainly not just inductively seeking patterns in data. Often, the process is hypothetico-deductive, where conjectures or hypotheses arise from initial observation and exploration, i.e., inductively, while almost simultaneously their deductive consequences

are envisaged and explored. The latter aspect we might call "deductive inquiry." Nevertheless, most inquiry-based learning sequences include inductive aspects that are not present in direct instruction. However, ancillary knowledge surrounding a topic also plays into new knowledge construction, and the science itself likely was not arrived at purely inductively in the first place. For these reasons, it is unrealistic to expect students to construct core knowledge purely inductively, even if guided, and especially not for every aspect of every science topic.

5. *Cause and effect (where applicable).* For science topics involving cause and effect, direct mode instruction often implicitly proceeds from cause to effect, or at least states the scientific cause right after showing an effect or phenomenon. Inquiry mode on the other hand often starts with an observed phenomenon and tries to understand the cause or mechanism. An example is in the contrasting approaches to the reasons for the seasons in one of our instructional units.
6. *Teacher techniques and actions.* The distinction between epistemic approaches is primarily about instructional design rather than teaching techniques. Teacher actions are seen as strategies to implement a chosen approach, but in themselves do not constitute the approach. For example, the issue is not whether a teacher asks a lot of questions or handles a student question by posing another. Neither is it about "eliciting prior knowledge" nor addressing common student conceptions. The direct or inquiry character of an approach lies primarily in how concept development is structured and the nature of the cognitive learning path. Specific teaching methods should then be such as to support mode.
7. *Content and process aspects.* One cannot assume that a particular mode is equally effective for teaching both the content and process aspects of science. Nor can it be assumed that content and process must necessarily be taught together in every lesson. However, these seem to be assumptions of some advocates for using inquiry-based methods throughout and thus obtaining "added value." Hattie (2009) cites studies showing that inquiry methods prove more effective for teaching process than for content. Our study specifically focuses on efficacy for learning core disciplinary content, although our active-engagement lessons naturally involve science process aspects.
8. *Mode narratives and rhetoric*. Direct and inquiry approaches to a topic will have different teaching narratives. The phrases "ready-made-science" and "science-in-the-making" (Latour 1987) help reflect the distinction. In direct mode students engage mostly with *knowledge comprehension and application*, while in inquiry mode they initially engage with knowledge *production*. The underlying narratives are different; the two approaches tell different stories, portray knowledge differently, and may send different messages about what science is like. A direct approach presents established knowledge at the beginning and may or may not give evidence or reasoning as to how we know, while inquiry tends to work from evidence and reasoning before developing a knowledge claim at the end. As noted previously, in many ways, these are inverse narratives. In some textbooks, one finds dry expository treatments that hardly mention context or basis for the factual knowledge presented. Schwab (1978) characterized this style of presentation as a "rhetoric of conclusions." However, it need not be this way. Ideally, a direct treatment can provide a good explanatory narrative, including purpose, context, interpretation, and implications. For inquiry approaches, the framing and rhetoric of the narrative will usually be that of discovery, whether or not the knowledge is developed by students' own hands-on investigations. In either mode, interesting stories can be told.

The various characterizing features listed above are not all independent of each other and may not all be present in any particular lesson episode. Several tend to go naturally together, and individually or as a whole they potentially affect efficacy for concept learning. Clusters of such characteristics are at work in the lesson designs for the topic units developed for this study.

There are various arguments one could put forward for each epistemic mode of learning core concepts. For example, as noted, if students develop a scientific concept themselves, they might understand it better and retain it longer, in which case an inquiry approach may work best. A counterargument might be that having students try to "invent" demanding concepts in short order while also carrying out practical activities may be unrealistic, in which case explicit instruction could be less confusing and more effective. Direct instruction is often seen as more "efficient" than inquiry for teaching content knowledge, but also as missing a sense of "how we know." Inquiry is seen as potentially more interesting and engaging, but also more time-consuming and covering too little content. In practice, one finds a diversity of methods and implementations under direct and inquiry labels, ranging across a spectrum from didactic direct presentation at one extreme to unguided open discovery at the other. Neither of the extremes are considered good educational practice, having research support.

We next elaborate on historical and current debates regarding direct and inquiry instruction, and the challenges that face research and that motivated our current study.

Background and Conceptual Framework

The Direct /Inquiry Debates

For more than a hundred years there have been both educational and political debates over the relative merits of so-called direct and inquiry approaches to science education, with strong opinions on both sides. However, exactly what is meant by each of these terms varies considerably, in both teaching practice and research studies. 'Direct' and 'inquiry' are words people use readily, even casually or vaguely without elaboration, but may actually have in mind variety of different things. Ambiguities and diverse views about the nature and purposes of the two approaches continue to confound both instruction and research. Nevertheless, one can say broadly that while forms of direct instruction have mostly dominated past practice, inquiry-based science instruction has become widely advocated worldwide in recent years through various national and state science education standards (e.g., AAAS 1990; Committee on Prospering in the Global Economy 2005, 2010; Michigan State Board of Education 2004; NRC 1996, 2012, 2013). In practice, there is a great range of teaching methods within each of these categories. Thus, one finds a wide variety of teaching practices nominally called "inquiry," though they may differ considerably in approach focus and degree of guidance, and may also involve different facets of science as objectives. A similar variety in practice is found for forms of "direct" or "explicit" instruction, ranging, for example, from purely didactic "lecture" exposition of content to structured lesson cycles with sequenced active learning stages.

In science, there are different kinds of endeavors with different aims, and aspects of science education can reflect this. The "discovery" aspect of science aims to *produce* new knowledge, i.e., to *develop* the basic science itself. An inquiry-based instructional approach that guides learners to develop core ideas through exploration and concept invention therefore reflects

discovery science. Another aspect of scientific work is the *application* of established science to answer specific questions, test predictions, or explain phenomena (McGrew et al. 2009). This is "normal science." A direct instructional approach can reasonably reflect this, in that after students have learned the theory they can do experiments to test it or obtain results, and use it to solve problems. Unfortunately, the distinctions between these different kinds of real science are not sufficiently recognized when referring to science instruction.[2] The question of the merits of direct and inquiry modes of instruction in science is therefore far from straightforward and unambiguous.

Conflations and Confusions

Various conflations of terminology and confusions of constructs regarding teaching methods plague both instruction and research. "Direct" has been tacitly conflated with "passive," "rote," or "lecture," and "inquiry" has been conflated with "hands-on" or "active learning" (Alozie et al. 2010; Anderson 2003; Bonwell and Eison 1991; Brickman et al. 2009; Dori et al. 2007; Kanter and Konstantopoulos 2010; Marx et al. 2004; Nock 2009). One also finds epistemic modes confused with particular classroom teaching techniques, or even identified with them. Conflations can lead to problems with both instructional design and research studies, and thus lead to results that are ambiguous, conflicted, and difficult to interpret. They are all the more pernicious when they go unrecognized. Klahr (2013) writes specifically of such issues and the problems they cause in both instruction and research.

Ausubel noted such conflations long ago, and argued that both "reception learning" and "discovery learning" (as he called them at the time) can lead to *meaningful* learning. Mental engagement is essential for meaningful learning in either mode of instruction. Direct instruction is not equivalent to didactic exposition with passive reception, nor is inquiry equivalent to "hands-on" (Anderson and Smith 1987; Ausubel 1961a, b, 1963; Ausubel et al. 1986; Heppner et al. 2006; Novak 1976, 1979). Effective use of either mode requires coherent instructional designs that lead students to engage with the important ideas, integrate them into a meaningful cognitive structure, and relate them to real phenomena through hands-on and minds-on experiences. Teaching that fails on these points is simply poor teaching, whatever the mode. Nevertheless, in some eyes, direct instruction is identified simply with didactic presentation of content by teachers or textbooks, and inquiry has become identified largely with process, for example, hands-on experimental activities that sometimes leave core content as secondary. Neither of these confusions of methods with mode are accurate or defining depictions.

Fundamental Epistemic Modes

Some might suggest that the direct/inquiry instructional mode distinction is posed too starkly and dichotomously, given the great diversity of science instructional methods found in practice. However, at root level, most of these pedagogical variations can be seen as having either a direct mode or inquiry mode underlying the epistemic approach. True, complete

[2] Current documents such as *A Framework for K-12 Science Education* (National Research Council 2011) make the point there can be different pedagogies for different purposes; however, current documents in science education are silent on the idea of pedagogies involving "an applied investigation" to obtain a "result," although many experimental activities found in classroom practice are actually of this nature rather than for science concept development.

lessons in the classroom will always be composites of elements serving various instructional purposes, but nevertheless each new science concept will usually be approached via either a direct or inquiry route, and one can identify and distinguish these two epistemic categories even in complex and diverse lessons.

Active Engagement

Inquiry-based pedagogies by their nature almost always involve some form of student activity. Therefore, the question arises whether it is the presence of such activities rather than the epistemic character of an approach that is mainly responsible for any improved acquisition of concept understanding. Furthermore, active student engagement in ill-designed inquiry lessons does not necessarily translate into meaningful learning. This is especially true if a curriculum or lesson plan is activity-based rather than concept-based, and in the extreme, hands-on activities can degenerate into "activity-mania" (Moscovici and Nelson 1998) that develops little science but takes time. Discriminating the nature of the epistemic approach from the degree of active engagement in various teaching methods is thus important for characterizing and evaluating instruction. In our research, we compare two contrasting epistemic modes which *both* involve active student engagement, referring to these as *active-direct* and *guided-inquiry*, while specifying the nature of each by operational models.

Ausubel's Conceptual Framework for Learning and Instruction

Ausubel's theory of learning as it relates to type of instruction provides a useful conceptual framework for instructional design as well as for analyzing and characterizing research studies on the effectiveness of various instructional types. Ausubel (1961a, b, 1963) and Novak (1976, 1979) argued that the important learning goal was "meaningful learning" as opposed to "rote learning," *whatever the type of instruction*. Ausubel provides a two-dimensional model representing *type of instruction* along a vertical axis and *type of learning outcome* along a horizontal axis (Fig. 1), thus separating the two constructs. Instructional type can range from reception to discovery, terms which are reasonably well reflected today by direct and inquiry. Learning outcome types can range from rote to meaningful. Besides memorization, rote learning would include fragmented learning of facts rather than building

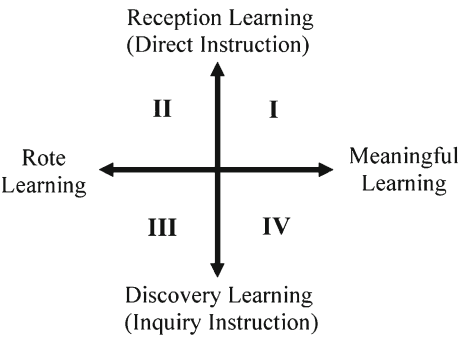


Fig. 1 Ausubel's axes form quadrants that relate type of instruction to nature of learning

coherent connections. Meaningful learning is such that new knowledge becomes integrated into and enriches the learner's conceptual schemata. On the orthogonal axis diagram in Fig. 1, the four quadrants I, II, III, and IV represent various possible combinations of learning outcome types and instructional types. In this framework, both reception and discovery learning can be either meaningful or rote. Quadrants I and IV both represent meaningful learning outcomes, attained with different types of instruction.

Ausubel and Novak believed that reception learning could be meaningful with appropriate instructional design. Novak referred to this as "direct facilitation of concept learning," and developed tools such as advance organizers and concept mapping for fostering meaningful reception learning (Mayer 1979; Novak 1976; Stone 1983; Trowbridge and Wandersee 2005). Research on conceptual change (Duit and Treagust 2003; Thorley and Stofflett 1996), explanatory analogies (Dagher 1995, 2005), bridging analogies (Clement 1982, 1998; Clement et al. 1989), and combining verbal learning with visual learning (Clark et al. 2011; Culatta 2012) may involve forms of direct instruction that can facilitate meaningful conceptual learning, and thus reside in quadrant I. Inquiry-based instruction ideally aims at quadrant IV, but an inquiry activity producing little meaningful learning would fall in quadrant III (rote/fragmented learning).

We draw on Ausubel's teaching/learning model in our research for a number of reasons. First, it is a model that focuses on both teaching and learning and discriminates between them. A model that contains both is particularly useful when working with learning outcomes from alternative pedagogies. Second, it is a model that provides a perspective and language for characterizing any teaching/learning endeavor along two orthogonal dimensions and for evaluating the nature of research studies into teaching and learning. Consider a study that compares the outcome from an inquiry pedagogy that includes student engagement activities with that for a direct pedagogy based solely on transmission and passive reception. This could be described as a quadrant IV vs. quadrant II comparison. Unfortunately, researchers too often make such straw man comparisons and misinterpret and misattribute the results they get. The proper comparison to set up would be between Q-IV and Q-I cases, if the aim of each method is meaningful learning. Taking this two-dimensional viewpoint allowed us to identify where gaps, confusions, and conflations existed in the literature, and to understand what kinds of comparisons were actually being made. We reviewed the inquiry studies that were used in the Furtak et al. (2012), Minner et al. (2010), and Schroeder et al. (2007) meta-analyses, and also reviewed the studies cited in *Taking Science to School* (2006) and *A Framework for K-12 Science Education* (2011). We found *no* Q-I vs. Q-IV (direct/meaningful vs. inquiry/meaningful) comparative studies. Comparative studies are typically Q-IV vs. Q-II (inquiry/meaningful vs. direct/rote) studies.

Another reason we find this teaching/learning framework still pertinent today is that it does not involve a priori assumptions or restrictions on the particular teaching methods that may be used to achieve learning goals. Therefore, we were able to design a fair, unconfounded comparison between Q-IV and Q-I cases, with instruction in both cases designed for active engagement and meaningful learning.

What Ausubel called "discovery" learning, as advocated by Bruner (1961, 1971) and others (e.g., Guthrie, 1967), subsequently developed into today's inquiry-based instruction (NRC 2000b). Unfortunately, much of the literature since the late 1980s tends to collapse Ausubel's two-dimensional framework of separate orthogonal constructs into a single dimension, with direct instruction implicitly identified with rote

learning and inquiry instruction with meaningful learning.[3] Although research into various forms of meaningful reception learning continue today (e.g., Clark et al. 2011; Klahr 2000; Matlen and Klahr 2013; Sweller 2009), the rote/meaningful learning dimension tended to be forgotten as the direct/inquiry instructional dichotomy became the focus. In 2000, the widely referenced book *How People Learn* (NRC 2000a) advocated active learning and inquiry instruction, with no mention of other meaningful learning options. Research on science instruction has focused predominantly on how to make inquiry instruction more effective, either by the professional development of science teachers (e.g., Oliveira 2010; Schneider 2011) and reducing teacher resistance to inquiry (e.g., Costenson and Lawson 1986; Robertson 2006; Roehrig and Luft 2004), or by innovations in inquiry instruction (e.g., Kanter and Konstantopoulos 2010; Lee et al. 2010; White and Frederickson 1998). However, the extensive research literature since the 1960s on the effectiveness of inquiry instruction, while generally positive, provides ambiguous and sometimes conflicting results, and we argue this stems largely from the way the issues are conceptualized and the research is designed.

Instructional Methods and Research Studies of Effectiveness

In this section, we review and discuss relevant literature concerning direct and inquiry methods, including some of their history and the research studies into effectiveness.

Direct Instruction

Good direct instruction aims at clarity of explanation and demonstration, with students cognitively engaged (Adams and Engelmann 1996; Randolph-Mason Women's College 2003; Schwerdt and Wuppermann 2011). In his theory of meaningful learning as related to instruction, Ausubel (1963) argued that directly presented, actively received and processed information could be meaningfully learned through integration into cognitive structures. Common methods of direct instruction include lectures, illustrations, demonstrations, audio-visual presentations, and of course textbook expositions, and associated laboratory activities are usually highly structured. However, note that the defining characteristic of epistemic mode is learning path rather than particular method and usually involves a number of learning phases. There are various models for effective direct mode instruction that describe important components, stages, and active student tasks (Archer and Hughes 2011; CSSP 2002; DataWORKS Educational Research 2012; Haak et al. 2011; Hassard 2003; Lawson 2010; Peterson 2011; Rosenshine 2008; Wright 2013).

Many of these direct instruction models provide carefully structured learning stages for use in lesson planning and classroom practice. An influential direct instruction model was developed in the 1960s by Engelmann (Adams and Engelmann 1996) and used in an extensive federally-funded research and implementation program called Project Follow Through, in

[3] Unfortunately, some of the support for inquiry instruction is based on an incorrect picture of expository or direct instruction. For example, discussions about Benchmarks for Science Literacy and the National Science Education Standards noted their "focus on inquiry-based science—rather than memorization" (Brady 2008, p. 607). This implies there are only two options: inquiry-based instruction or memorization, an opinion echoed by the Committee on Prospering in the Global Economy of the 21st Century (2009) and by the editorial pages of *Science* (e.g., Alberts 2011). The subtext here is that direct instruction, as the opposite of inquiry instruction, must necessarily be instruction for rote learning and memorization (e.g., Hein 2004).

which Direct Instruction System for Teaching Arithmetic and Reading (DISTAR) gained prominence. Instruction in science was not involved, though the same instructional principles would apply. A model developed by Hunter (Mueller 2013) was based on observation of the actual practices of successful teachers, and includes sequenced lesson elements, such as review, anticipatory set, objectives, topic input, modeling, checking understanding, guided practice, monitoring, closure, and independent practice. Archer and Hughes (2011) give a detailed account of Explicit Instruction in their book by that name. The various elements of these models reflect aspects of successful teaching practice, and most educators, inquiry-oriented or not, would hardly disagree with of them. Well-designed direct instruction is more than just content exposition for reception by passive students, although unfortunately this is a common caricature. Nevertheless, a criticism of much direct instruction is that it portrays science mainly as a final product—a body of knowledge. The narrative tends to be a "rhetoric of conclusions," to use Schwab's phrase, so that direct instruction may present "what we know" while neglecting "how we know."

In our experience and observation, most science teachers use some degree of direct instruction at various times in lessons, even if they aim to teach mainly by inquiry, and some teachers are more comfortable and effective with direct instruction. It is widely believed that direct instruction is more time-efficient than inquiry, and that its clear structure benefits students.

There is considerable research support for direct instruction. A 1996 meta-analysis of research studies on direct instruction found the average effect size per variable studied was about 0.75 (Adams and Engelmann 1996). VanLehn et al. (2005) found that direct explicit training in physics problem-solving was successful at helping students set up free-body diagrams and write and solve equations, and Chi and Van Lehn (2007) found this accelerated students' learning and transferred to new areas. Support is also found in Anastasiow et al. (1970), Chen and Klahr (1999), DataWORKS Educational Research (2012), Education Consumers Foundation (2011), Egan and Greeno (1973), Holliday and McGuire (1992), Klahr (2000, 2002), Klahr and Nigam (2004), Klauer (1984), Mayer (1979), Shuell (1986), Tennyson and Cocchiarella (1986), Walberg (1991), Wright and Nuthall (1970), and Yeany and Miller (1983). In *A Time for Telling*, Schwartz and Bransford (1998) argue that direct exposition and explanation can be very effective cognitively once students are adequately prepared to actively process and accommodate the information presented. Direct instruction also finds support in areas of education other than science, such as findings from DISTAR (Adams and Engelmann 1996), the American Federation of Teachers (2003), and Finn and Ravitch (1996), which suggest that effective instruction must be more teacher-led than student-directed. In 2002, the recipient of the Award for Education Research of the Council of Scientific Society Presidents was Siegfried Engelmann, cited for his research on direct instruction.

Leaving aside core disciplinary ideas for the moment to mention science *practices*, it is worth noting the Klahr and Nigam (2004) research, which showed showing that explicit (direct) instruction is superior to unguided discovery for learning about control of variables, an important science practice. This was questioned by Dean and Kuhn (2006) but affirmed by a more nuanced study by Matlen and Klahr (2013) involving different degrees and sequencing of guidance, which found that explicit instruction and high guidance throughout produced the best results. Thus, one cannot simply assume, as some do, that direct instruction will be inferior to inquiry for science practices and process skills, and there are arguments and evidence

to the contrary. It may be that practices are best shown rather than figured out, but this is not part of our work, which focuses on core concept acquisition.

Inquiry Instruction

Although there is great diversity in practice, inquiry-based instruction in science generally refers to approaches that aim to reflect the investigative attitudes, techniques, reasoning, and reliance on evidence that scientists use to construct new knowledge, i.e., the processes of scientific inquiry. Ideally, inquiry instruction allows students to develop science concepts and principles through teacher-scaffolded (White and Frederickson 1998) explorations of phenomena. This approach is the basis, for example, of the Investigative Science Learning Environment (Etkina and Van Heuvelen 2007). An inquiry approach to content will at the same time model certain process aspects of scientific inquiry. This may be considered a potential "added-value" benefit of inquiry-based instruction. The notion of "Scientific Teaching" (Ebert-May and Hodder 2008; Handelsman et al. 2004) posits that "the teaching of science should be faithful to the true nature of science by capturing the process of discovery in the classroom" (Yale University 2012). This may sound plausible but needs to be strongly qualified: students learning science in classroom sessions and scientists doing real science over time are different activities, in different situations, by different actors, with different goals. Not to recognize this is to make a category error. Thus, while the focus of the statement quoted above seems to be that the classroom situation should closely resemble real science, a broader perspective from learning is that an inquiry approach represents a cognitive and experiential learning trajectory towards understanding a science concept.

Inquiry-based instruction implicitly or explicitly aims to address three kinds of learning objectives simultaneously: content, process, and nature of science. The balance varies greatly; recently, there has been emphasis on hands-on practices, either in their own right or as a vehicle to teach content or nature of science, though sometimes at the expense of core ideas. The merits or otherwise of expecting students to learn three aspects of science at the same time in the same activity may be debated.

Inquiry-based approaches to science teaching have a long history, though they have only become prevalent in recent times. The movement to bring science into the school curriculum began in the late 1800s when advocates envisioned science instruction based on experience with the physical world, gathering of data, rational argument, and drawing of inferences from evidence. Thomas Huxley spoke of scientific training as "practising the intellect in the completest form of induction; that is to say, in drawing conclusions from particular facts made known by immediate observation of Nature" (DeBoer 1991, p. 11). The origins of the modern day concept of inquiry science teaching lie with the 1960s NSF-funded curriculum projects (Anderson 2003; DeBoer 1991; Krajcik et al. 2001; Rudolph 2002; Schwab 1962). Laboratory activities became ubiquitous in science instruction on the basis that effective pedagogy must reflect the nature of a discipline. Science as a discipline is not only content but also inquiry, "the warp and woof of a single fabric" (Rutherford 1964, p. 83). It was reasoned that science instruction must therefore be more than the clear explication of information; it must include the investigative processes and thinking that lead to the development of concepts. Instructional designs aiming to reflect scientific inquiry in topic teaching and concept development are often based on "learning cycle" models, which have stages devised to represent aspects of scientific inquiry (AAAS 1990; Eisenkraft 2003; Lawson 2001; Renner and Marek 1990), with concept development a central stage.

In recent years under National Research Council and AAAS leadership, the USA has developed a commitment to the teaching of science *as inquiry* across the K-12 grades (American Association for the Advancement of Science [AAAS] 1990; NRC 1996, 2000b, 2012).[4] The science education community, including the National Science Teachers Association (NSTA), the National Association for Research in Science Teaching (NARST), and the Association of Science Teacher Educators (ASTE), has overwhelmingly adopted an inquiry pedagogy perspective for science education, and an emphasis on inquiry has become prevalent internationally as well.

Many educators feel that inquiry instruction is more in keeping with cognitive constructivism, i.e., the tenet that meaningful knowledge cannot simply be transmitted and absorbed; understanding needs to be actively constructed by learners. Thus, Llewellyn (2007) states: "For many teachers, the principles of constructivism lay the foundation for understanding and implementing inquiry-based learning" (p. 53). However, constructivism is a theory of learning rather than of instruction, and learners must process input and construct understanding whatever the nature of instruction. There are also potential affective benefits to inquiry; students may become more curious and interested in the topic, with increased intrinsic motivation and intellectual satisfaction.

Despite the potential benefits of inquiry instruction, there are some problems with its use. If inquiry is too open-ended, students have difficulty forming suitable questions to explore, choosing variables to work with, linking hypotheses and data, and drawing correct conclusions from experiments (de Jong et al. 2005). Students can become lost and frustrated, and unguided naïve intuitions can lead to misconceptions (Brown and Campione 1994). As a result, teachers may spend considerable time scaffolding students' content and procedural skills together (Aulls 2002). It may also be unrealistic to expect students to be able to "invent" fairly demanding concepts by inquiry in a short lesson, when historically it may have taken scientists many years. Detractors see inquiry as inefficient and ineffective. Furthermore, as Padilla (2013) recently pointed out, "The key, often forgotten, aspect of inquiry is that it is an intellectual endeavor," noting that students can sometimes be "physically but not intellectually engaged in science" (2013, p. 26).

There is also the issue of the degree of instructional guidance. Inquiry and the National Science Education Standards (NRC 2000b) describes five features of inquiry-based instruction and for each lists a spectrum of possible "levels" of inquiry practice, depending on the degree of teacher-directedness. The tendency toward less-guided science teaching methods is criticized by Sweller et al. (2007) in their paper *Why minimal guidance during instruction does not work*, which gives both a theoretical basis and empirical support for more explicit instruction. A response by Kuhn (2007) does not address this but shifts the question to what we should teach, arguing that the body of scientific knowledge is not so important compared to inquiry and argumentation skills, and that activities should instead center on developing these skills. She suggests that students should acquire only "some rudimentary understanding of the physical and biological world around them" (p. 111). This position seems to conceive science content as an ever-expanding accumulation of facts, rather than recognizing the central role of core principles and coherent theory in the scientific enterprise. Sweller et al. (2007) give an effective response, and the Next Generation Science Standards reflect the crucial role of core disciplinary ideas.

[4] This preference for inquiry instruction continues with the Next Generation Science Standards.

A practical problem is that teachers and researchers alike have a wide range of notions about what actually constitutes inquiry, for what purposes, and what methods are appropriate. Varied interpretations and practices notwithstanding, inquiry is omnipresent in the language of science education.[5] Anderson (2002) asserted that research regarding the teaching of science had matured and "tended to move away from the question of whether or not inquiry teaching is effective, and has become focused more on understanding the dynamics of such teaching and how it can be brought about" (p. 6).[6] However, we believe that the question remains open, perhaps phrased briefly as: should one teach content through inquiry, and what does this actually mean in practice? What science teaching methods may be most effective for what purposes?

In recent decades, literally thousands of articles have been published on inquiry instruction in the sciences. Proponents of inquiry argue that the evidence to date provides support for its effectiveness in improving content learning, science process skills, and student attitudes (e.g., Kanter and Konstantopoulos 2010; Marbach-Ad and Claassen 2001; Marx et al. 2004; Secker 2002; Secker and Lissitz 1999; Timmerman et al. 2008; Tretter and Jones 2003; Udovic et al. 2002; White and Frederickson 1998). This research is typically cited by policy documents such as *A Framework for K-12 Science Education: Practices*, *Crosscutting Concepts*, *and Core Ideas* (NRC 2011). The findings, however, are mixed, and a few even negative, with respect to the efficacy of inquiry instruction for conceptual understanding of core ideas (e.g., Anastasiow et al. 1970; Ausubel 1961a, b, 1962; Craig 1956; Ivins 1985; Kersh 1962; Shulman and Keislar 1966; Tai and Sadler 2009; Wittrock 1964). Given the large number of research studies and the varied results, researchers have used meta-analyses to try to reach a conclusion. For example, Shymansky et al. (1983) and Shymansky et al. (1990) found support for inquiry by using meta-analysis techniques on the effectiveness of NSF-funded curricula. More recently, meta-analyses by Furtak et al. (2012), Minner et al. (2010), and Schroeder et al. (2007) provide reasonable support overall for the effectiveness of inquiry instruction. Many of the meta-analyses, however, acknowledge difficulties, in part because of assumptions they had to make about the original research studies, including ignoring issues with research designs. Minner et al. (2010) specifically noted this problem, stating "the rigor over this 18-year time span of the synthesis studies indicate a small but statistically significant trend toward a decrease in the methodological rigor with which the studies were conducted…." (p. 14). Critics such as Klahr (2000) and Sweller (2009) view very little of the research on inquiry as unconfounded (see also Kirschner et al. 2006; Mayer 2004). Too many studies on inquiry lack operational definitions of type of instruction, use vague or ambiguous terminology, conflate various constructs (Furtak et al. 2012; Klahr 2013), and are not comparative or adequately controlled.

Comment on Research Studies on Efficacy

Given all the above considerations about instructional modes and methods, it is important for science educators and instructional developers to understand the nature and relative efficacy of the two fundamental epistemic approaches for teaching and

[5] Although *A Framework for K-12 Science Education* (National Research Council 2011) cautions that earlier debates about dichotomies between direct and inquiry instruction are simplistic, the document clearly endorses inquiry instruction in its multiple forms.

[6] Implicit is the assumption that the case in favor of inquiry is now firmly established. Brady (2008) stated that, "We know how to teach science" implying that we know to teach science by inquiry. The only question left to ask was "when will we do it?" (p. 607). Similar views have been expressed in the editorial pages of *Science* (e.g., Allende 2008). Settlage (2007) commented that it would be "unlikely today that any methods book would have a chance in the marketplace if 'inquiry' was not prominent" (p. 464). Nevertheless, we believe that some assertions about the effectiveness of inquiry and the evidence behind it stretch the fabric of the research results.

learning science. Both have substantial bodies of literature that support their use. Evaluative studies of various forms of either "direct" or "inquiry" instruction report degrees of success for each approach. Hattie's (2009) synthesis of over 800 meta-analyses of instructional studies regarding contributions to learning from various teaching approaches provides a comprehensive account of research results for a great many approaches and strategies, including direct and inquiry methods in science education. Meta-analyses of research studies indicate varying degrees of success for both these methods. However, few of these studies are controlled *comparisons* of effectiveness of the two methods, operationally defined. More commonly, studies report performance results obtained for a particular method or implementation. Many involve curriculum innovations along with pedagogy. In such comparisons, many factors besides basic instructional mode are at play, for example, different topics, curricula, teachers, students, classroom conditions, objectives, assessments, or teacher preparation. In many cases, only the innovation but not the control is specified in any detail, and the comparison may be against a straw man foil reflecting poor instruction. We do not consider such studies to be meaningful comparisons of pedagogies if they involve multiple confounding factors and are not conducted under well-specified treatment and control conditions. There is also the question of what is assessed: in some research studies, the assessments did not align very well with all the objectives and claimed benefits of the instructional method.

Regarding degrees of success reported for a broad range of educational strategies, Hattie (2009) remarks that "almost everything works" (p. 15). He notes that 95 % of effect sizes for all the things we do in education are positive; virtually all strategies, whatever their nature, are reported to "work" in having some positive effect on achievement (especially if the bar for learning is only set at zero rather than a meaningful effect size of around 0.4). The real question, however, is how well does a method work compared to others and for what purposes. Nevertheless, across all the meta-studies, Hattie's synthesis indicates that direct instruction is associated with a greater effect size on achievement (Cohen's $d = 0.59$) than is inquiry ($d = 0.33$) (p. 120). This synthesis of research on instructional effectiveness is for all subjects, not only science. Reported outcomes also depend on different kinds of objectives that may be the goal of instruction, for example, whether the main focus is on process (practices) or on content (core ideas). Inquiry potentially targets both of these, and in this regard Bredderman (1983) reported that inquiry had a greater effect on process ($d = 0.52$) than on content ($d = 0.16$), and similarly Shymansky et al. (1990) reported effect sizes for science inquiry of $d = 0.4$ for process skills and $d = 0.26$ for content. A systematic review by Marzano (1998) concluded that the "best way to teach organizing ideas—concepts, generalizations, and principles—appears to be to present those constructs in a rather direct fashion" and then have students apply these concepts to new situations" (p. 106).

However, we note again that even within this large number of studies and meta-analyses, there are few unconfounded controlled studies comparing instructional modes directly, using the same topics, teachers, learning objectives, assessments, and classroom conditions, and examining where each mode may provide better concept learning outcomes.

Current Situation

We argue that weaknesses of theoretical conceptualization, conflations, and flaws in method have much to do with the current situation of inconsistent or conflicted research results regarding efficacy of instructional methods under direct and inquiry labels. As noted earlier, in much research on instruction for science concept learning, the two-dimensional framework

of orthogonal teaching/learning axes has been ignored or implicitly collapsed, thus associating reception (direct) mode with rote learning at one pole, and discovery (inquiry) mode with meaningful learning at the other. When a study makes a comparative claim against a "control" referred to simply as "traditional science teaching," it might also be a sign that the researchers have recognized neither the distinctions nor orthogonality (e.g., Harris et al. 2012; Tretter and Jones 2003). Many studies compare effectiveness of methods located in quadrants II and IV (direct/rote vs. inquiry/meaningful), but a more valid and useful comparison would be between Q-I and Q-IV cases (direct/meaningful vs. inquiry/meaningful), i.e., comparing worthy alternative methods for achieving meaningful learning. From the teaching/learning framework perspective, the varying or even inconsistent research results found for the effectiveness of inquiry-based methods is perhaps not surprising, and some research claims for the superiority of inquiry may be suspect.

Weaknesses of conceptualization or theory and lack of control for confounding factors can lead to problems with research design and methodology and hence questionable conclusions. For example, Lee et al. (2010) compared inquiry lessons, which also included embedded visualizations, to a "traditional teaching" control without such visualizations, and concluded that the inquiry approach was superior. However, from such a design, it is impossible to know whether the instructional mode or the visualizations (or both) were responsible for the outcome. A study by Blanchard et al. (2010) provided 6 weeks of professional development in inquiry teaching for a new unit to a treatment group of teachers, but provided none to the control group. The findings of the study, interpreted as favoring inquiry, are confounded by the professional development disparity.

With all the above issues in mind, the central question we addressed was not whether cognitively engaged, experiential learning of science is more effective than passive, non-experiential learning. That question has been unequivocally answered in favor of the former. Rather, our research asked whether a direct mode or inquiry mode epistemic approach to active, experientially based learning is more effective for science conceptual development, when lessons in both modes are expertly designed and well taught. We therefore compared two active-engagement methods whose fundamental difference was in how science concepts were approached in instruction and encountered by the learner. The goal of our present study was to determine the relative efficacy of *active-direct* and *guided-inquiry* epistemic modes for science conceptual development in a controlled comparative study at eighth grade level. Instructional modes were operationally defined, and topic units designed in parallel, with both modes containing hands-on active-engagement activities. Learning objectives were identical for the two modes, as were the assessments. Our work pays particular attention to research design and methodology for a fair and unconfounded comparative study, attempting to identify and obviate many of the ambiguities, conflations, methodological problems, and threats to validity that crop up in research on instructional effectiveness (Kelly and Lesh 2000a, 2000b; NRC 2002; Schneider 2004).

Project Goals and Research Design

The project had both instructional development and research goals as follows:

Instructional development goals

(a) To characterize features of alternative epistemic modes of instruction.

(b) To devise operational models for active-direct and guided-inquiry instructional modes.
(c) To design and develop two science units, each in both direct and inquiry modes, and produce materials for students and guides for teachers.
(d) To develop formative and summative assessments for student conceptual understanding of core disciplinary ideas.

Research goals

To conduct a randomized controlled experimental study comparing student outcomes for learning core disciplinary ideas via active-direct and guided-inquiry instructional modes, focused on the following research questions:

(a) Does one or the other instructional mode lead to better learning gains for students' understanding of core ideas, or do both lead to similar gains?
(b) Do students' learning gains vary significantly between teachers and between topics?

Research Structural Design

Our focus was the comparative efficacy of two alternative epistemic approaches to the teaching and learning of core disciplinary ideas. To determine this, we designed topic units in the alternative modes, having either direct or inquiry learning paths to the core concepts, but otherwise equivalent. Five experienced middle school teachers implemented the units in an eight-day summer program, over 5 years of trials with over 500 eighth grade students. From the performance assessment data, we compared student learning gains achieved in the two modes, for each topic and for each teacher. The experimental study was a randomized controlled trial of the comparative efficacy of operationally defined active-direct and guided-inquiry instructional treatments.

The structural design of the research is shown in Table 1. This shows the successive stages of the project: development of instruction and assessment; piloting trials; professional development; and four annual research trials. It also shows the crossover research design involving teachers switching modes after the first two trials, so that each teacher taught two trials in one mode and two in the other. Partial early findings before mode switching were described in a previous article (Cobern et al. 2010). Students were randomly assigned to active-direct or guided-inquiry classes, hereafter referred to simply as direct or inquiry.

During the development phase of the project, suitable topics were chosen, learning objectives formulated, instructional units designed and developed in each mode, lab equipment obtained, and student and teacher materials produced. Conceptual assessments were developed for each topic, aligned to the learning objectives and instruction. The development phase was followed by a professional development and piloting year. We worked with teachers preparing them to teach the units in the assigned modes, and ran a full-scale pilot study where teachers implemented the units in an initial summer program, on the basis of which we made improvements. For the next 4 years, we conducted four research trials in four successive summers. The experimental design involved the two science topic units, labeled A and B in the table, taught in each of the 4 years, by four teachers, with two teaching by direct mode and two by inquiry mode. A fifth teacher and fifth class of students were included to handle enrollment overflow and the possibility of losing a teacher over the duration of the project. Since no teacher loss occurred, the fifth data set could be included in the study. The crossover research

Table 1 Structural design of the study

Year	Unit	Project activity				
Development	A	Dynamics (force & motion) unit and assessment development				
	B	Light (climate & seasons) unit and assessment development				
		Teacher 1	**Teacher 2**	**Teacher 3**	**Teacher 4**	**Teacher 5**
Pilot year	A	• Direct	• Direct	□ Inquiry	□ Inquiry	□ Inquiry
	B	• Direct	• Direct	□ Inquiry	□ Inquiry	□ Inquiry
Trial 1	A	• Direct	• Direct	□ Inquiry	□ Inquiry	□ Inquiry
	B	• Direct	• Direct	□ Inquiry	□ Inquiry	□ Inquiry
Trial 2	A	• Direct	• Direct	□ Inquiry	□ Inquiry	□ Inquiry
	B	• Direct	• Direct	□ Inquiry	□ Inquiry	□ Inquiry
Trial 3	A	□ Inquiry	□ Inquiry	• Direct	• Direct	• Direct
	B	□ Inquiry	□ Inquiry	• Direct	• Direct	• Direct
Trial 4	A	□ Inquiry	□ Inquiry	• Direct	• Direct	• Direct
	B	□ Inquiry	□ Inquiry	• Direct	• Direct	• Direct

design ensured that each teacher taught each of the units A and B in both direct and inquiry modes.

Operational Models for Contrasting Instructional Modes

The instructional development stage of the project involved designing instructional units in direct and inquiry modes for teaching and learning of core content in two chosen topics. We used *active-direct* and *guided-inquiry* modes, which we define operationally below, informed by ideas from Cobern et al. (2012), Haak et al. (2011), NRC (2000b), and Renner and Marek (1990). These two active-engagement options can both potentially lead to meaningful learning, and thus are located in quadrants I and IV of Ausubel's framework. We devised *instructional models* for each mode, specifying stages of instruction and the nature of each. On these models, we based the detailed designs of our direct and inquiry lessons. The distinction we made between modes was in framing and sequencing and hence in the cognitive and experiential learning paths toward concepts. Other than that, we strove to keep everything else the same between the two methods for a fair comparison. In either mode, the importance of clear lesson structure was recognized for effective learning, with different structures for the two modes. The contrasting models are described below.

Model of Active-Direct Instruction Our perspective on direct instruction is informed by research showing that neither didactic direct teaching nor "cookbook" lab activities

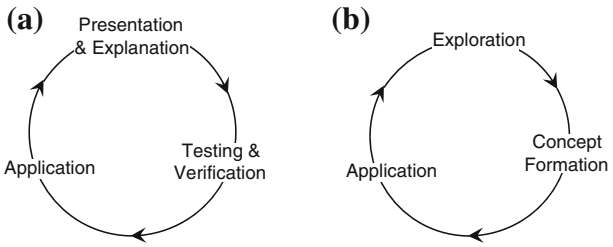


Fig. 2 Active-Direct Learning Cycle (**a**) and Karplus Guided Inquiry Learning Cycle (**b**)

are effective approaches for meaningful learning (NRC 2000a). Therefore, our "active-direct" mode of instruction has the following components: presentation and explanation, verification/replication, and application (Fig. 2a). The instructor uses exposition and examples to present and explain concepts and principles explicitly to the students, as established knowledge to be learned and understood. This also serves as a basis for subsequent student laboratory activities to test and verify theory, thus acting as a form of advance organizer (Ausubel et al. 1986; Mayer 1979) for the experimental part. In the application phase, students apply the concepts and principles in questions, problems, and explanations.

Model of Guided-Inquiry Instruction Our model for guided-inquiry instruction is based on the Karplus learning cycle (Abraham and Renner 1986; Atkin and Karplus 1962; Karplus 1977), informed by subsequent research and cognitive learning theory, and expanded in models such as the BSCS 5E learning cycle (Brown and Abell 2007; Bybee et al. 2006; Eisenkraft 2003; Lawson 2001). The Karplus cycle is intended to reflect scientific inquiry in teaching and learning, and has three major phases: exploration, concept formation, and application (Fig. 2b). In the exploration and concept formation phases, teachers guide students in experimental and cognitive activities toward forming the target concepts and principles. In the application phase, students apply the concepts and principles in questions, problems, and explanations. We use the Karplus cycle as the basis for inquiry instruction because it is at the heart of the other cycles and embodies the desired mode characteristics.

Design and Development of Instructional Units and Assessment Items

Overall curriculum coherence is important for achieving educational as well as research goals (Bybee 2003; Li et al. 2006). Therefore, our project involved a substantial instructional development component, including developing learning objectives, teaching models and methods, instructional units, experiments, demonstrations, student workbooks, teacher guides, and assessments, and locating lab equipment. The various aspects are described below.

Science Topics and Instructional Units

We chose two areas of science that occur in virtually all science standards: *Force and motion* and *Earth temperatures and seasons*. Both involve core concepts, laws and scientific models,

and are of substantial conceptual demand. Students are known to have difficulties with them, with various "alternative conceptions" being common. Each topic was developed into imaginative new instructional units, formulated in both epistemic modes. The units were as follows.

A. *It's Dynamic!* The relation between force, mass, and motion.

This is a conceptual unit on introductory dynamics involving concepts of force, net force, position, velocity, acceleration and mass, and the relation between force, mass, and motion (Newton's laws). It was restricted to straight-line cases for project purposes. The unit is hereafter referred to simply as *Dynamics*.

B. *It's Illuminating!* The relation between sunlight and temperature variations on Earth (climate and seasons).

This is a conceptual unit on temperature variations on Earth that arise from radiant solar energy and the various earth-sun geometries involved. The unit comprises a foundation of basic science (intensity dependencies on angle, distance, and time) as a basis for understanding temperature variations on Earth by location (latitude) and by time of year (seasons). The treatment is from both ground-based and space-based observational perspectives. The unit is hereafter referred to simply as *Light*.

These topics are of different types. Unit A is "pure" science, about fundamental concepts and laws. Unit B uses core ideas and geometrical models to explain the observed temperature variations on earth by latitude and time of year. Both types of science are important, and each can be taught in direct or inquiry modes. We wanted both in our research in case the nature of the topic made a difference to instructional mode efficacy.

For each topic, broad goals and detailed concept learning objectives were formulated. These were consistent with National and State standards (MSBOE 2004; NRC 1996) and are also in accord with the subsequently released Next Generation Science Standards. The units were developed as 8-day modules each involving an hour of instructional time per day, including a welcome and pre-test session the first day, and a post-test and award session the last day. The central focus was on coherent development of the core ideas rather than on process-oriented activities. Treatment of the science was mainly conceptual but included basic mathematics where appropriate. Since student interest and engagement is important in learning, the units had interesting storylines and activities. Following concept acquisition, students applied the concepts in conceptual problems. Since meaningful understanding of a concept involves the ability to use it (Krathwohl 2002; Mintzes et al. 2000; Mintzes et al. 2005), which was the basis of our assessment items where students apply core ideas in solving problems, answering questions, explaining cases, and making predictions.

The instructional units were developed in parallel in the two modes. Parallel design ensured that both versions were produced with equal attention and were equivalent in all aspects other than mode. This does not often happen when innovations are compared with "existing" instruction. The topic example given earlier, for teaching Newton's second law, illustrates the distinctions and commonalities between modes.

For our eighth grade students, who had little experience of concept development in inquiry mode, most inquiry episodes of the units were strongly guided. Appropriate degree of

guidance will generally depend on the nature and complexity of the topic and the background and needs of the learners. Too little guidance risks confusion and unproductive use of time, and may prompt undesirable naïve conceptions. Guidance is planned but dynamically adapted in class. Even strongly guided lessons remain inquiry lessons in terms of framing, sequencing, and cognitive learning path.

Direct lessons are generally easier for teachers to plan and execute than inquiry, which might also play a role in the relative success of each method in the classroom. Nevertheless, effective direct instruction is demanding of good pedagogical content knowledge for each phase of instruction, such as the nature and phrasing of explanations. Thus, a conceptual explanation of Newton's first law of motion, for example, might be couched in terms of the notion of "coasting" and relating this to students' experiences of "keeping going" on a bicycle without pedaling, with a real demonstration. Such motion can usefully be phrased as the "natural" motion of objects when there is no net forceacting, helping to make intuitive physical sense of the law. This sequence is designed to cue in students a phenomenological primitive or p-prim (DiSessa 1988, 1993), which we might call 'persisting' or more specifically 'keep on coasting', as a useful resource to build on in learning. Such explanations and illustrations are important for meaningful understanding, beyond a clear formal statement of the law.

Development of Instructional Materials

Student Booklets For each unit, we wrote student booklets in both modes for use in class, mostly in worksheet outline form. These also provided a guiding structure for students and teachers during lessons, and helped to maintain fidelity to both curriculum and mode.

Teacher Booklets We also produced teacher booklets for each unit in both modes. These contained the student materials and corresponding teacher guide materials: teaching notes, sample lesson narratives, equipment lists, projection slides, and wall posters. The narratives were to help lesson preparation, not to serve as scripts to be used verbatim in class. They also supported professional development for the units by illustrating the intended meanings of modes in specific cases.

The booklets are available from the authors.

Development of Assessment Items and Instruments

Instruments comprising sets of assessment items were created to assess student understanding of core ideas for each topic. They were closely aligned with the specific learning objectives and instruction. Items were at middle school level and consistent with district and state grade-level standards. Items were selected-response conceptual assessments of core ideas, administered pre- and post-instruction. The nature and quality of assessment was important since it provided the basis of measurement of learning gains.

Assessment items embodied our criterion for understanding: the ability to apply the basic science correctly in new situations (Anderson and Krathwohl 2001). The format was multiple choice with four response options providing plausible conceptual alternatives, including distracters representing common "alternative conceptions." Items were case-based and problem-based, and conceptual rather than numerical or factual. They tested understanding of core ideas and were equally appropriate whether the instruction was direct or inquiry. Cognitive

demand was at Bloom's Taxonomy levels 2 and 3 (Comprehension and Application). Level 2 items were usually variants of cases seen in class, while level 3 items involved application to new or unfamiliar cases.

Items were tested and refined during the pilot year. Items were checked by independent experts in both subject matter and assessment, for content validity, construct validity, cognitive demand and clarity, and also to confirm that they did not in some way favor either direct or inquiry instruction. The complete instruments were also evaluated for reasonably balanced coverage of learning objectives for the topics. The piloting phase allowed us to obtain statistical data using subjects from the actual eighth grade target population. Information obtained on item difficulty and discrimination enabled us to replace or modify items that appeared too easy or had low discrimination, which would reduce the power of the study to detect mode differences. On some items, the scores before instruction were higher than expected, likely due to prior knowledge, which would leave less room for instruction-induced gain, so items with the highest pre-scores were later replaced with more challenging items, and items showing little discrimination were eliminated or modified. To avoid the possibility of inadvertent or deliberate "teaching to the test," the teachers were not involved in either item development or test administration, and were thus blind to the summative tests. They did, however, use other items of this nature formatively as "concept checks" during instruction.

Four examples of assessment items are given in Fig. 3, two items from each topic unit. These examples give a concrete idea of the nature of the assessment and the level of conceptual understanding that was the desired outcome of instruction.

Comment on Assessment Characteristics

Assessment Creation and Alignments We created all assessments ourselves, specifically designed for these instructional units. Assessment was closely aligned with learning objectives and instruction, and reflected the type of conceptual understanding we wished students to attain. Such alignment is crucial for valid research assessment of student understanding arising from instruction, but is not generally the case if one seeks an "existing" external instrument that might somehow serve as an indicator even if not a good match. We piloted and refined the instruments in a pilot year. To ensure balanced concept coverage, we drew up tables of alignments, showing which assessment items addressed each learning objective, which learning objectives were addressed by each assessment item, and which grade-level content standards applied. The objectives, assessments, and alignment tables are available from the authors.

Item Format and Characteristics We used an objective selected-response item format in this research for a number of reasons. Carefully designed and worded selected-response items are well suited to assessing conceptual understanding of core ideas if posed as qualitative problems with plausible conceptual response options. The distracters involved common alternative conceptions and naïve intuitions associated with the physical concepts. Although multiple-choice questions commonly found in external high stakes testing are often (justifiably) criticized as involving mostly declarative knowledge, isolated facts, or formula plugging, this is certainly not the case for our conceptual problem-based items, which are demanding of meaningful understanding of core ideas. Conceptual items of this type are useful individually as formative "concept checks" during instruction, and summatively in an

1. Example items from the Dynamics unit

Here are two example items which test conceptual understanding of the relation between force and motion (Newton's first and second laws respectively)

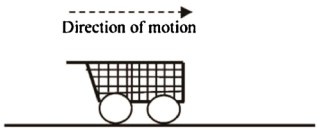
1.1 Ice skater

An ice skater is skating on an indoor ice rink. She first gets up to a fast speed, then stands on one skate and keeps going steadily across the ice without any apparent effort.
The reason for her keeping moving like this without effort is that . . .

A. the ice must be sloping slightly downwards.
B. the air is pushing her forward.
✔C. this is the natural behavior of objects with no net force on them.
D. the force that got her moving is still acting on her.

1.2 Shopping cart

A shopping cart is given an initial push to get it moving on a carpeted floor. The cart then travels along the carpet on its own for a while, gradually slowing down.



What horizontal force(s), if any, are acting on the shopping cart while it is still moving forward but slowing down?

A. Only a forward force, diminishing with time.
✔B. Only a backward force.
C. A forward force and a backward force, but the forward force is bigger.
D. A forward force and a backward force, but the backward force is bigger.

2. Example items from Light, Climate & Seasons

These examples are two of the more challenging conceptual questions for this unit, testing understanding of how various factors cause temperature variations on Earth's surface

2.1 Reasons for the seasons

In Canada it is **colder** in January than it is in July. People suggest various possible reasons for this difference, as follows:

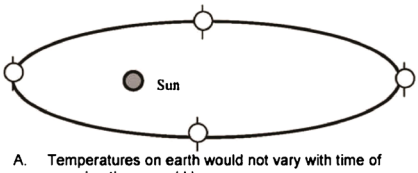
x. The sun's rays strike the ground at more of a slant in January than July in Canada.
y. Daytime is shorter in January than July in Canada
z. The Earth is farther from the Sun in January than July.

Which of the above are correct reasons for the temperature difference?

✔A. x and y only
B. z only
C. y and z only
D. x, y, and z

2.2 If the Earth orbited differently

Imagine that instead of being nearly circular, the earth's orbit was a long ellipse, with the sun nearer to one end of it than the other as shown. Then as the earth orbited, its distance from the sun would vary a lot. Also imagine that the earth's axis has *no tilt*. This is all shown in the diagram.
With this *long elliptical orbit* and *no tilt*, which one of the following statements would be true?



A. Temperatures on earth would not vary with time of year. i.e. there would be no seasons.
B. Temperatures would vary during the year, with seasons being opposite in the northern and southern hemispheres.
✔C. Temperatures would vary during the year, with seasons being the same in the northern and southern hemispheres.
D. It is not possible to say without more information.

Fig. 3 Examples of assessment items from the two science content units

instrument to assess concept mastery overall. Assessments of similar nature are widely used in various "concept inventory" instruments such as the force concept inventory (FCI), a method for assessing conceptual understanding of force across multiple situations, and which has been much used in evaluating aspects of mechanics courses.

Comparative Assessment While more detailed information could have been obtained by the addition of graded written responses, note that the research goal was comparative: we sought possible differences in assessed learning gains between two instructional modes. Since assessment was identical for both modes, any reservations or preferences about assessment format are applicable to both modes, and are therefore less pertinent in comparative studies. The selected-response format allowed us to include more students and have more assessment items than would have been practical with graded written responses, and we could also use standard statistical analyses on the items and responses, providing information at a number of levels. Objective selected-response assessments have both strengths and limitations, but the fairly demanding conceptual nature of the questions and response options with respect to core ideas should be clear from the examples in Fig. 3.

Research and Methods

Setting, Subjects, and Implementation

To conduct the experimental comparison of instructional mode efficacy, we organized an annual summer science program for middle school students about to enter eighth grade, recruited from several urban, suburban, and rural school districts. A summer program allowed for the random assignment of students to direct or inquiry classes for research purposes. We elected to work with eighth grade because the middle school years are transitional between elementary and high school and begin the formal study of science, thus being important to students' future success. School districts sent out advance program announcements to parents, and student participation was a family decision. Over a hundred students each year participated in classroom trials over five summers from 2006 to 2010 including the pilot phase. The program ran for 8 days over 2 weeks, Monday through Thursday each week. One-hour lessons in each of two topics were taught each day by our five experienced middle school teachers. In the opinion of our teachers, the composition of students attending the trials was not noticeably different from that of their regular middle school classes with regard to academic ability, interest, and behavior.

A summer program has various advantages and disadvantages for research compared to conducting it in schools. Running our own program enabled us to assign students randomly to treatment groups, something difficult or impossible to do in the regular school situation. We could also assign teaching modes to classes and topics, and control pace and duration of lessons. Other factors in the program were the freshness of the situation for students, in having new teachers, peers, and locale, and the relatively short time in class each day. Students were thus not locked into existing "rituals" of their regular daily schooling, which might have affected how they responded to the new units and pedagogies. We wished to use a classroom format but also minimize pre-existing expectations, routines, and habits from their usual environments, which would likely not be the same for the different students. Nevertheless, a voluntary summer program format has drawbacks and limitations. A possible disadvantage is that it does not count for formal grades, arguably reducing extrinsic motivation and incentive (though students did get completion certificates). Also in a voluntary summer program, it is unrealistic to assign homework and reading. However, this might be seen as an advantage for research purposes, since it avoided unknown external influences out of class. Learning gains achieved were dependent upon in-class student engagement with lessons in the two modes. Notwithstanding these possible advantages and disadvantages, it is important to note that these aspects were the same for both treatments, and our interest was in the comparative learning gains between contrasting instructional modes.

Teachers, Mode Assignments, and Preparation

Teacher Recruitment We recruited five experienced middle school teachers, referred to as Ann, Joe, Liz, Sam, and Tom, who were already familiar with the broad topic domains of the units. As experienced teachers with good classroom control, they would be able to focus on instruction and fidelity to mode, rather than worry about how to ensure discipline and attention.

Mode Allocations Teachers were initially allocated to one of the two treatment modes according to their preferences, i.e., the way they felt most comfortable teaching. Allocating to the other mode at the start could have introduced a confounding factor for some teachers but not others, involving switching natural style and thereby affecting instructional quality, at least initially. For comparing mode efficacy, we needed to control for possible "teacher effects"; therefore, after trials 1 and 2, the teachers switched modes for trials 3 and 4. The crossover also provided some information on the nature and magnitude of teacher effects, although a limitation is the small number of teachers in the current study. An alternative crossover research design would be to have each teacher teach in both modes on the same day; this is attractive in theory, but we were concerned that daily mode switches would be hard for teachers to handle and make it difficult to maintain fidelity to mode.[7]

Professional Development and Teaching Fidelity Teachers and researchers worked together in multiple sessions during the pilot year to ensure that mode characteristic and distinctions were understood, and that lessons taught would be as close as possible to the intended curriculum and instructional mode. Toward the same end, teachers could use the detailed teacher booklets in both preparing and teaching their lessons. Nevertheless, we could not simply assume that the carefully designed lessons in each mode would be implemented as intended, so observers evaluated teacher fidelity to mode and curriculum. Reasonable fidelity expectations for teachers need to allow for the flexibility inherent in good teaching. Teaching involves interacting with students and shaping things dynamically as the lesson proceeds, with a personal style and natural degree of personalization of the narrative. All classrooms have variability, due to variations in students, teachers, and events. Therefore, our operational criterion for sufficient fidelity was that qualified independent observers were able to identify instructional type within natural background variation, and assign a fidelity rating of at least 5 on a 7-point scale. Independent observers[8] evaluated two lessons per teacher per unit, and each teacher was seen by two observers. Observers were initially blind to teacher mode assignments, but because fidelity to mode was reasonably good, they quickly identified the direct and inquiry teachers. Therefore, in subsequent sessions, they had the appropriate unit materials and could score teachers on fidelity to mode and lesson plan. Qualitative and quantitative fidelity findings are included with the study results. Teachers posted journal notes each day on how teaching went, how students responded, and where they may have deviated from intended lessons. The researchers monitored these journals for signs of any problems. All lessons were videotaped and could be reviewed, both to monitor fidelity and for teacher development purposes.

Student Performance Measurements

Assessment instruments consisted of 21 selected-response items for the dynamics unit and 24 for the light unit, and were compiled from a larger bank of items we created, tested, and

[7] This decision was later validated by the opinions of our teachers when they switched modes after 2 years. They were more comfortable handling the mode switch after having longer experience with the units. Even then, two of our teachers were initially unable to effect the switch with acceptable fidelity (trial 3), even though they had time to prepare for it.

[8] For the independent fidelity assessment, the project contracted a team specializing in observing and evaluating science instruction, the Science and Mathematics Program Improvement (SAMPI) group, nationally known for its expertise at science/math curriculum and instruction evaluation (Jenness and Barley 1999).

refined. The instruments were administered pre- and post-instruction, so that learning gains could be ascertained rather than just final level of understanding, thereby taking into account possible differences in starting knowledge. The tests were administered by the external evaluators. Students circled their chosen responses on the question paper. The data were entered electronically by project assistants to obviate errors that may occur if students fill in scantron sheets. Data from the assessments represented 2 science topics/units, 2 modes, 5 teachers, 20 classes, 409 students, and 4 trial years. Processing of performance data would enable us in principle to analyze it by topic, mode, teacher, class, student, trial year, whole instrument, and individual items.

Results and Analyses

We present results for the following aspects of the research: teaching fidelity; pre- and post-test data; performance gains and normalized gains; and comparative analyses across modes, topics, teachers, and trial years.

Teaching Fidelity and Quality

For the first two trials, all five teachers met the fidelity standard. In the third trial, where teachers switched modes, one direct instructor (Sam) and one inquiry instructor (Ann) fell below our "fidelity to mode" criterion; hence, their data from that trial were not included in the comparative analyses below. Preparation for the fourth trial included having teachers review the videos of their teaching, and these two teachers were able to identify their difficulties. As a result, in the fourth trial, all five teachers met the fidelity criterion.

The median teacher "fidelity to mode" rating of 6 on a 7-point scale is arguably adequate for research purposes, while remaining realistic with respect to inevitable variation in actual science classrooms. Fidelity scores were somewhat higher for direct instruction than inquiry. The independent evaluation results gave us confidence that the difference between treatments (instructional modes) was sufficiently clear in both the construction and the implementation of lessons.

We became aware that teachers' existing conceptions of direct and inquiry teaching methods interacted with how they interpreted the intended epistemic mode, even with written teacher guides. They tended to see things from a teacher point of view, perhaps not surprisingly, and to conceive named methods in terms of teacher actions rather than learning paths. One teacher, on first switching from inquiry to direct, said it was "such a relief not to have to draw the lesson out of the students." Another teacher was not clear initially how much to "allow" questions and discussion in direct mode. All benefited from thinking about their ideas and practices in the light of the project.

Pre-Test and Post-Test Data and Score Distributions

As noted earlier, quantitative data on student conceptual understanding of each topic unit were obtained by administering sets of conceptual multiple-choice items. Items tested student ability to apply the core science concepts to new situations. There was a range of difficulty; some items were relatively easy and/or resembled cases seen in instruction, while others were more demanding and/or novel. The assessments were the same for both modes of instruction. Student responses were scored and analyzed

for percentage of items answered correctly using standard MCQ analyses. As an example of the score data we obtained, Fig. 4 shows the distribution of pre- and post-test scores for the light unit over all five teachers and all 4 years. Normal curves are also fitted, but note that post-test scores may not in fact be distributed normally due to the score ceiling; the curves are used simply to depict visually the means, widths, and shift of mean (gain). Average pre-test scores on the multiple choice assessment instruments of around 50 % were higher than expected, given that guessing would potentially lead to pre-test scores around 25 %. However, middle school students do have some prior exposure at lower grade levels to certain aspects of the topics. Standard deviations on both pre- and post-tests were around 20 %. Student scores on the pre-tests indicated that randomization of students across classrooms was effective, in that any variation in pre-scores between classes was consistent with that expected by chance for class sizes of 20 to 25 students.

Gain and Normalized Gain

From the pre- and post-test performance data, we calculated both raw percentage gain (post-score minus pre-score) and normalized percentage gain. The latter is defined as the ratio of actual gain to maximum possible gain for a given pre-score. It has become fairly common practice to use normalized gain as a measure of pre-post improvement, as a way to take into account different pre-test scores, since higher pre-scores offer less potential gain. The defining equation for normalized gain is g = (post-score − pre-score) / (max score − pre-score). Normalized gains are thus ratios between 0 and 1, with 1 being the maximum achievable. To minimize unusual distortions that might occasionally occur with this definition (if a pre-score is high and the post-score lower), we used the concept of normalized change in subsequent calculations (Marx and Cummings 2007). Normalized change is the gain or loss over the maximum possible gain or loss, respectively.

Effect sizes (Cohen's d) for overall raw percentage gain over four trials/years were 0.71 for the dynamics unit and 0.81 for the light unit. Correlations (Pearson) between pre-scores and post-scores were typically positive and significant. Raw gains showed consistent but negative correlations with pre-test scores, but normalized gains did not (normalization was effective).

Mean normalized gains were just over 0.2 for the dynamics unit and over 0.3 for the light unit, for both modes. These are of the same order as typical normalized gains on the well-known force concept inventory (FCI), which Hake (1998) reports as ranging from about 0.2 for traditional courses to about 0.35 for courses involving active engagement. Gains were less than expected, but note that our assessment items are conceptually demanding, involving not mere knowledge recall but application of core concepts to conceptual problems and new situations, and most students in the program would not have been used to this.[9]

[9] Given that our program ran during the summer without homework or grade incentives, our gain scores compare well with other conceptual assessment results on topics known to be challenging. However, when average gains are modest (e.g., less than a standard deviation), differences in gains between instructional modes will also be small, especially in such "real-world" classroom contexts, given the score spreads and classroom variations observed.

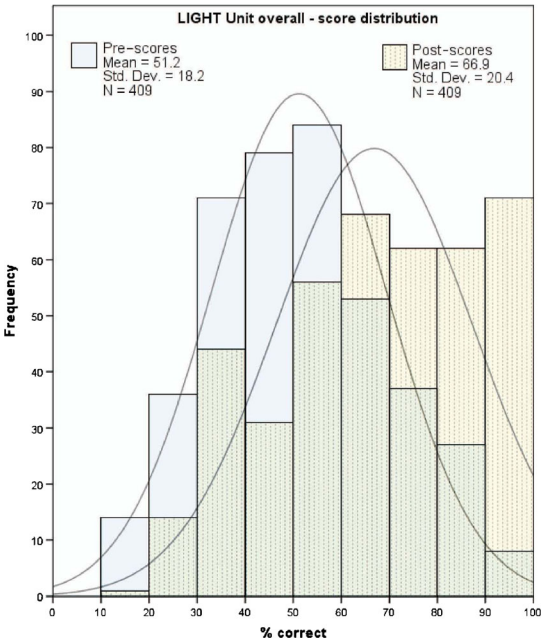


Fig. 4 Overlaid histograms showing overall pre-score and postscore distribution for the light unit

Comparative Analyses by Unit Topic, Teacher, Mode, and Trial Year

In the following sections, we present and discuss results for students' science content understanding (scores and gains), grouped by the two unit topics, the two instructional modes, the five teachers, and the four trial years. To address our main research question comparing modes of instruction, we first report comparisons within subsets of the data as it accumulated over time in each unit, then present an overall comparison between direct and inquiry instructional modes based on aggregated data from the entire project.

For each raw gain and normalized gain/change value, we calculated standard deviation and determined to what extent the differences in gains were statistically significant under the conditions of our program, using standard ANOVA and/or two-tailed *t* tests and an alpha level of 0.05. Given that randomization was at the student level, the student was taken as the unit of analysis to allow for future analyses with respect to student characteristics. The performance data collected also allowed us to analyze at the detailed level of individual assessment items, though we are not reporting results of such analyses here.

Within both the dynamics unit and the light unit, data from trials 1 and 2 are shown aggregated, as are the data from trials 3 and 4 (after teachers' switched modes). We justified this on the basis that within teacher/class variance within each pair of trials was calculated to be no more than variance expected by chance. Therefore, the tables below are grouped by pairs of trial years within each unit topic. Each shows results for the five teachers and for the two instructional modes, direct and inquiry. Class average pre-scores, post-scores, and raw gains are displayed in tabular form (along with normalized change) as numerical means with standard deviations.

Results Across Trials for the Dynamics Unit

Results for the dynamics (force and motion) unit are shown in Table 2. All results and conclusions regarding instructional mode are based on the classes/teachers who met the fidelity standard, therefore the third trial data for Ann and Sam are not included.

In the first pair of trials of the dynamics unit, there were no statistically significant differences between modes or between teachers on raw gain or normalized gain/change. The second two trials in dynamics yielded statistically significant differences for normalized change between Liz and Ann ($t(59) = 2.311$, $p = 0.024$; effect size $d = 0.68$), between Liz and Joe ($t(91) = 2.375$, $p = 0.020$; $d = 0.50$), and between Liz and Tom both teaching within direct mode ($t(91) = 2.081$, $p = 0.040$; $d = 0.44$). There was a smaller but statistically significant overall difference between inquiry and direct on normalized change ($t(175) = 2.010$, $p = 0.046$; $d = 0.32$), but not on raw gain.

Results Across Trials for the Light Unit

Results for the light (climate and seasons) unit are shown in Table 3. All results and conclusions regarding instructional mode are based on the classes/teachers who met the fidelity standard, therefore the third trial data for Ann and Sam are not included.

In the first pair of trials of the light unit, the only statistically significant difference found was between teachers Ann and Tom on raw gain ($t(73) = 2.132$, $p = 0.036$; $d = 0.61$), though not on normalized change ($t(73) = 1.857$, $p = 0.067$). For the second pair of light trials, similar to the dynamics results, we found statistically significant differences between one direct teacher (Liz) and three other teachers on raw gain as well as normalized change: Ann ($t(65) = 2.683$, $p = 0.009$; $d = .73$), Joe ($t(92) = 3.030$, $p = 0.003$; $d = 0.63$), and another direct teacher, Sam ($t(67) = 2.692$,

Table 2 Dynamics unit results by teacher and by mode

Trials 1 and 2 (2007-2008)

Means (SD)	Ann (N=39)	Joe (N=32)	DIRECT (N=71)	INQUIRY (N=106)	Liz (N=35)	Sam (N=35)	Tom (N=36)
pre-score	54.9 (15.0)	48.1 (18.0)	51.8 (16.7)	48.3 (17.4)	46.0 (15.8)	51.6 (16.8)	47.4 (19.3)
post-score	66.7 (14.5)	58.2 (18.5)	62.8 (16.9)	58.1 (18.9)	55.4 (17.1)	62.3 (21.6)	56.6 (17.4)
% gain	11.7 (12.1)	10.1 (13.9)	11.0 (12.9)	9.8 (12.9)	9.4 (8.8)	10.7 (13.7)	9.3 (15.5)
normalized change	26.1 (25.0)	19.6 (28.8)	23.2 (26.8)	20.4 (27.5)	18.7 (19.1)	26.3 (33.5)	16.3 (27.6)

(This analysis omits 3 cases where the post-score was below 60% of the pre-score.)

Trials 3 and 4 (2009-2010) (teachers switched modes)

Means (SD)	Ann (N=16)	Joe (N=48)	INQUIRY (N=64)	DIRECT (N=113)	Liz (N=45)	Sam (N=20)	Tom (N=48)
pre-score	33.7 (11.3)	32.7 (13.7)	32.9 (13.1)	32.7 (12.9)	35.1 (10.8)	36.1 (15.8)	29.1 (12.7)
post-score	54.6 (18.2)	50.6 (18.1)	51.6 (18.1)	46.1 (15.4)	44.6 (12.3)	48.9 (17.8)	46.3 (16.9)
% gain	20.9 (21.5)	17.9 (19.0)	18.7 (19.5)	13.4 (16.0)	9.6 (14.8)	12.8 (15.7)	17.2 (16.5)
normalized change	29.4 (30.0)	24.8 (26.8)	25.9 (27.5)	17.8 (24.8)	12.4 (23.3)	18.1 (26.3)	22.8 (24.9)

(This analysis omits 9 cases where the post-score was below 60% of the pre-score, and excludes '09 Ann and '09 Sam)

p = 0.009; d = 0.71). We found a smaller but statistically significant difference between inquiry and direct on normalized change (t(156.6*) = 2.692, p = 0.008; d = 0.40), but not on raw gain. (*Equal variances not assumed, Levene's test.)

Overall Results Across Direct and Inquiry Instructional Modes

Our central research goal was to compare student learning gain outcomes for two epistemically distinct instructional modes, i.e., active-direct and guided-inquiry instruction. Findings over the four trial years in this respect can be summarized as follows.

Comparisons Within Each Unit Within the dynamics unit over all four trial years, the differences between direct and inquiry modes in raw gain and normalized change were not statistically significant. In the dynamics unit overall, Tom's direct mode scores were higher to a statistically significant degree than his inquiry scores (t(82) = 2.238, p = 0.028; p = 0.50). Similarly, looking at the light unit over all 4 years, the difference in raw gain between modes was not statistically significant; however, the difference in normalized change was somewhat in favor of the inquiry mode (t(361) = 2.143, p = 0.033) (mean difference 7.2, standard error of difference 3.4, effect size d = 0.23).

Figure 5 provides a graphical summary of the overall results for dynamics, whereby one can visually compare mean gains between and within different modes, teachers, and years. Figure 6 illustrates graphically the overall results for the light unit. The bar heights in Figs. 5 and 6 show the mean overall raw student gains for each mode, and the average gain scores across both modes (all teachers combined) are represented by dashed lines traced across each graph. Overall means per teacher within mode are shown by markers emphasizing whether taught in direct mode (solid circle) or in inquiry mode (empty diamond). Confidence interval (95 %) error bars are shown for teacher means per mode, and along side these "lines" the average gain

Table 3 Light unit results by teacher and by mode

Trials 1 and 2 (2007-2008)

Means (SD)	Ann (N=40)	Joe (N=32)	DIRECT (N=72)	INQUIRY (N=107)	Liz (N=36)	Sam (N=36)	Tom (N=35)
pre-score	59.4 (17.9)	50.1 (16.5)	55.3 (17.8)	53.8 (20.1)	55.2 (18.3)	53.0 (23.0)	53.2 (19.3)
post-score	70.5 (16.3)	65.5 (18.6)	68.2 (17.4)	68.3 (21.9)	67.6 (22.2)	65.5 (22.2)	71.8 (20.5)
% gain	11.0 (13.8)	15.3 (13.2)	12.9 (13.6)	14.4 (16.1)	12.4 (15.3)	12.5 (15.8)	18.6 (16.8)
normalized change	29.9 (30.8)	32.5 (28.3)	31.1 (29.5)	35.7 (35.2)	30.0 (35.4)	33.6 (36.7)	43.6 (32.9)

(This analysis omits 1 case where the post-score was below 60% of the pre-score.)

Trials 3 and 4 (2009-2010) (teachers switched modes)

Means (SD)	Ann (N=20)	Joe (N=47)	INQUIRY (N=67)	DIRECT (N=117)	Liz (N=47)	Sam (N=22)	Tom (N=48)
pre-score	49.5 (18.9)	51.5 (17.8)	50.9 (18.1)	47.1 (16.5)	47.6 (15.6)	46.3 (18.8)	46.9 (16.4)
post-score	71.0 (18.8)	70.0 (19.2)	70.3 (18.9)	63.1 (21.3)	60.1 (19.8)	68.6 (24.3)	63.5 (21.2)
% gain	21.4 (14.1)	18.5 (11.0)	19.4 (12.0)	16.0 (16.2)	12.5 (14.5)	22.3 (15.0)	16.6 (17.6)
normalized change	46.4 (26.8)	43.6 (27.9)	44.5 (27.4)	32.4 (32.4)	25.5 (30.1)	46.8 (31.8)	32.5 (33.2)

(This analysis omits 4 cases where the post-score was below 60% of the pre-score, and excludes '09 Ann and '09 Sam)

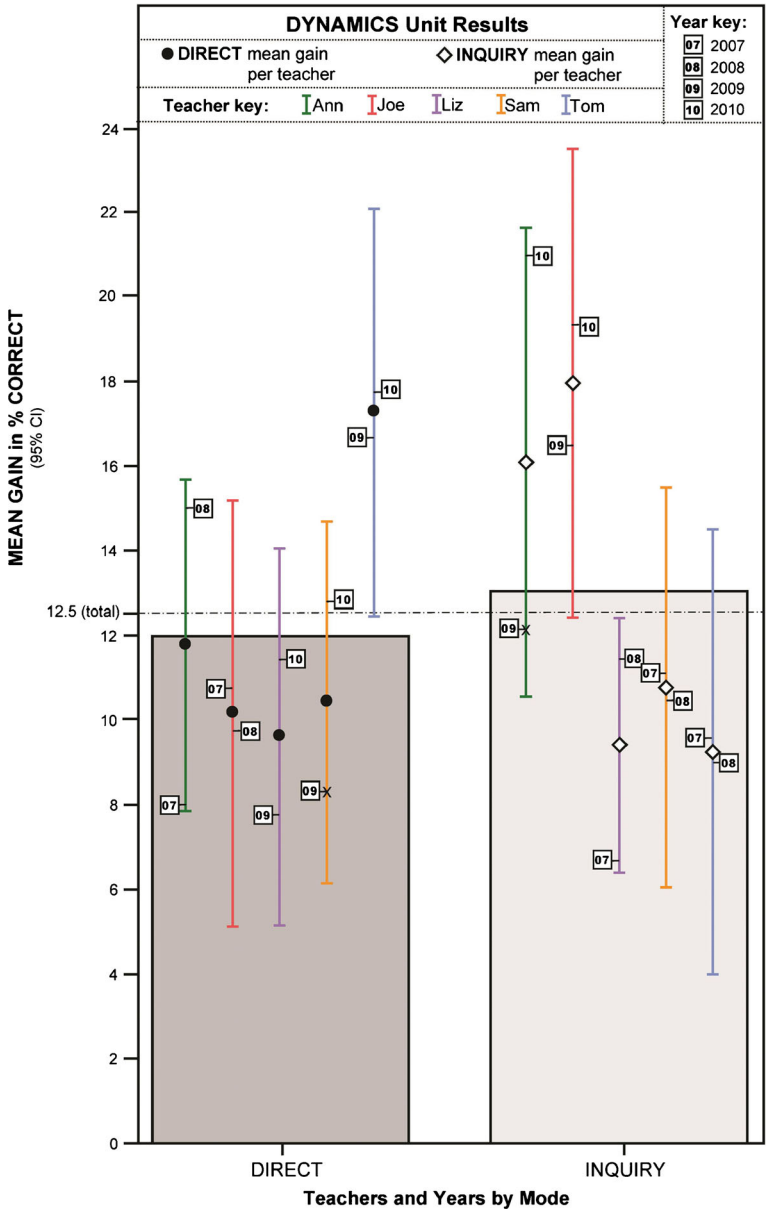


Fig. 5 Resulting gains for the dynamics unit by mode, and by teacher within mode and within year. (x indicates lack of fidelity to mode, hence data not included in statistical comparisons by mode)

scores per teacher per single trial/year are marked by the year itself, 07, 08, 09, and 10. The two classes which did not meet the threshold for fidelity to mode are indicated by an X.

Comparison Overall Combining the results of both the dynamics and light units over all four trial years (Table 4), using only teachers/classes meeting our fidelity to mode

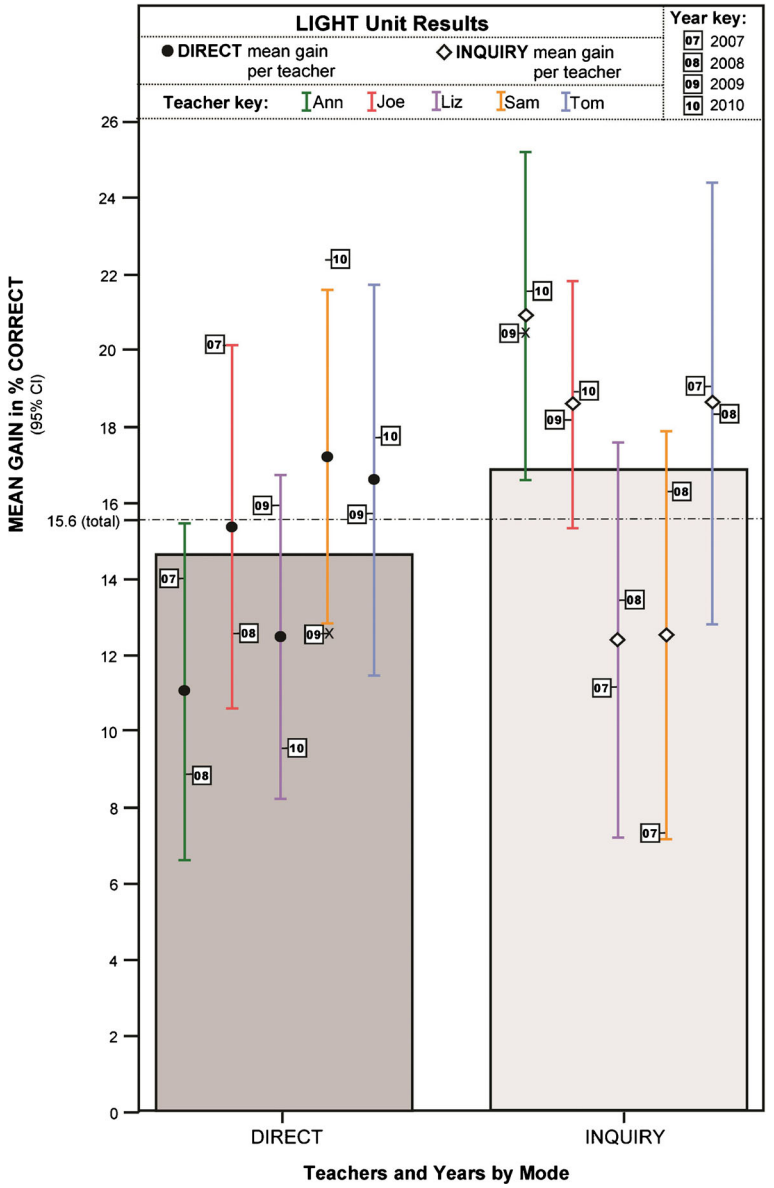


Fig. 6 Resulting gains for the light unit by mode, and by teacher within mode and within year. (x indicates lack of fidelity to mode, hence data not included in statistical comparisons by mode)

threshold, the difference in student score raw gain between direct and inquiry modes was not statistically significant (mean difference 1.1, standard error of difference 1.1, effect size $d = 0.07$). The difference in normalized change between direct and inquiry modes was statistically significant ($t(715) = 2.167$, $p = 0.031$), but with a fairly small effect size (mean difference 4.9, standard error of difference 2.3, effect size $d = 0.16$).

Over all four trials, with dynamics and light combined, Joe's inquiry mode student score gains were higher than his direct mode gains to a statistically significant degree

($t(157) = 2.297$, $p = 0.023$; $d = 0.37$), as were Ann's (excluding her 2009 data) ($t(53.0^*) = 3.020$, $p = 0.004$; $d = 0.69$).

On the other hand, Liz, Sam, and Tom all reached higher overall student gains in direct mode than in inquiry mode (Sam's differed by an effect size of 0.41, excluding his 2009 data, and Tom's differed by an effect size of 0.36).

Over all four trials, combining both units and all teachers meeting fidelity standards, the normalized change effect size between direct and inquiry of 0.16 was quite small relative to several differences between modes within teachers (above), and between teachers within modes (below). The small overall difference was not of *practical* significance, given the much greater variation both within teacher and between teachers. In summary, over all 4 years/trials, combining both units, and all teachers meeting fidelity standards, there was not a statistically significant difference between direct and inquiry modes of instruction with regard to gain/increase in percentage correct from pre-score to post-score on conceptual assessments (again, effect size $d = 0.07$).

Results Across Teachers

From Figs. 5 and 6, one can get a sense of the variation in results not only by mode but also by teacher. "Natural teacher variations" in personal teaching styles and practices were also clearly observable in the classrooms, as one might expect, even after professional development on the topics. As evident above, differences in student gains between modes were often smaller than differences in student gains between teachers, even within mode. Sometimes both the highest and lowest gain scores per trial were within the same mode (see Fig. 6, 2010, direct mode). Another potentially important observation is that across teachers, neither mode was consistently more effective for student learning.

Over all 4 years combined, within the direct mode for dynamics, Tom's students' gain scores were higher by a statistically significant amount than Joe's ($t(78) = 1.998$, $p = 0.049$; $d = 0.46$) and also than Liz's ($t(91) = 2.344$, $p = 0.021$; $d = 0.49$). Within the inquiry mode for dynamics, over all 4 years, Joe's students' gain scores were higher by a statistically significant amount than Liz's ($t(70.1^*) = 2.738$, $p = 0.008$; $d = 0.56$) and Tom's ($t(82) = 2.233$, $p = 0.028$; $d = 0.50$), and Ann's were also higher than Tom's ($t(50) = 2.215$, $p = 0.031$; $d = 0.68$). Due to lack of fidelity, the third trial

Table 4 Comparison of inquiry versus direct methods overall (dynamics and light units combined)

% mean gain/increase (SD) N	Ann	Joe	Liz	Sam	Tom	Total
Direct	**11.4** (12.9) N=79	**12.7** (13.7) N=64	**11.0** (14.6) N=92	**17.8** (15.9) N=42	**16.9** (17.0) N=96	**13.7** (15.1) N=373
Inquiry	**21.2** (17.5) N=36	**18.2** (15.5) N=95	**10.9** (12.5) N=71	**11.6** (14.7) N=71	**13.8** (16.7) N=71	**14.8** (15.6) N=344
Overall	**14.4** (15.1) N=115	**16.0** (15.0) N=159	**11.0** (13.7) N=163	**13.9** (15.4) N=113	**15.6** (16.9) N=167	**14.2** (15.4) N=717

data points for Ann's and Sam's classes did not contribute to comparisons within intended instructional modes.

Over all 4 years combined, within the direct mode for light, there were statistically significant differences in gain scores between teachers Sam and Liz ($t(67) = 2.599$, $p = 0.011$; $d = 0.68$), and Sam and Ann ($t(60) = 2.982$, $p = 0.004$; $d = 0.80$). Within the inquiry mode for light, over all 4 years, Liz's gain scores were lower by a statistically significant amount than Joe's ($t(61.0^*) = 2.045$, $p = 0.045$; $d = 0.48$) and Ann's ($t(54) = 2.179$, $p = 0.034$; $d = 0.62$). Ann's inquiry light gains were also significantly higher than Sam's ($t(54) = 2.102$, $p = .040$; $d = 0.60$). Again, for the light unit, due to lack of fidelity, the third trial data points for Ann and Sam did not contribute to comparisons.

Looking across teachers over all 4 years, with direct and inquiry instructional modes combined, and the light and dynamics units also combined, there were statistically significant differences in percentage gains between the classes of Liz and Ann ($t(318) = 2.541$, $p = 0.012$; $d = 0.29$), Liz and Joe ($t(320) = 3.144$, $p = 0.002$; $d = .35$), and Liz and Tom ($t(317.4^*) = 2.726$, $p = 0.007$; $d = 0.30$). (*Equal variances not assumed, Levene's test.)

Results Across Topics

Differences between results for the dynamics and light units on pre-score, post-score, normalized change ($t(781.4^*) = 7.359$, $p < 0.001$; $d = 0.52$), and raw gain (mean difference 3.1, standard error of difference 1.1) ($t(805) = 2.976$, $p = 0.003$; $d = 0.21$) were all statistically significant, but not greatly. The dynamics unit proved somewhat more difficult for students, in that there were lower gains overall. There is no reason to expect the same scores and gains on the two separate topic units, since they are on different content with different assessments, but the results indicate some relative consistency within the student sample pool regarding the challenges of these science topics. (*Equal variances not assumed, Levene's test.)

Results Across Different Trial Years

The "order" aspect can be viewed as a replication in four successive years with different student subjects, and with instructors switching modes halfway through. Results were similar across the 4 years of trials, with average performance data increasing slightly with year, which is not surprising as the teachers became more familiar with and adept at teaching the science content units. However, this gradual improvement over time was not statistically significant overall, nor was it relevant to our research questions. There were no statistically significant overall differences in average normalized gain/change between years, thus, this replication data could be aggregated for studying other factors. There was only one statistically significant difference found where year/trial was the only distinct variable, between trials 3 and 4, within the light unit, within one teacher (Sam), within one mode (direct) ($t(44) = 2.370$, $p = 0.022$; $d = 0.72$).

Lesson Time Comparisons Between Modes and Between Teachers

Overall, direct lessons took about 10 min less per nominal 1-h session than inquiry lessons, although this difference varied considerably according to the lesson involved and the particular

teacher. Time variations between teachers were at least as great as time variations between instructional modes.

Discussion and Conclusions

This experimental study compared two carefully designed, epistemically contrasting approaches to teaching and learning core disciplinary ideas in science. Operationally defined models of active-direct and guided-inquiry instruction were used to develop instructional units in two science topics, in both modes. These were taught by experienced middle school teachers to eighth grade students in classroom environments in an 8-day program, over 4 years of trials, in a controlled comparative study using a crossover research design. The two epistemic modes, each involving active-engagement activities, led to comparable learning gains for conceptual understanding of core ideas. Combining the results from the two units over all four trial years, using only data from teachers meeting a fidelity-to-mode threshold, there was a small difference in normalized change (but not raw gain) between active-direct and guided-inquiry instruction for conceptual understanding. This was marginally statistically significant, but the effect size (Cohen's d = 0.16) was small, and thus not indicative of any practical significance. To put the overall direct/inquiry effect size of 0.16 into perspective compared to teacher effects, note that between certain pairs of individual teachers, the differences found had effect sizes of 0.29, 0.30, and 0.35 on overall results, combining modes within teacher. Moreover, within the inquiry mode, we saw overall differences between pairs of individual teachers with effect sizes of 0.48, 0.50, 0.56, 0.60, 0.62, and 0.68; and within the direct mode, overall differences between teachers with effect sizes of 0.46, 0.49, 0.68, and 0.80. Another interesting point is that Table 4 shows that three of the five teachers, Liz, Sam, and Tom, had greater overall success with direct mode than inquiry, particularly Sam and Tom with effect sizes of 0.41 and 0.36.

It may not be surprising that one finds comparable learning gains for active-direct and guided-inquiry instruction, if both involve active engagement and are well taught in that mode. Learning is enhanced when students are engaged, and on this basis much existing inquiry instruction might certainly be more effective than more passive situations such as didactic lecture presentations or reading the textbook (as various Q-II vs. Q-IV studies have shown). The situation changes, however, when direct mode instruction is designed to facilitate active cognitive and experiential engagement. Our active-direct and guided-inquiry instructional approaches both involved student engagement, even though the concept learning paths were different. Another reason for eventual gains being similar in the two modes might be that after initial concept learning in either mode, understanding is consolidated in an application phase through problems and further discussion. This "spiraling back" on a newly learned concept occurs with all good instruction, and may tend to even out any initial differences in concept learning efficacy between modes. On the other hand, if a student does not learn a concept properly in one mode or the other, and then tries to proceed further on the basis of poor understanding, that student will probably not be able to cope well with the application phase either, so in fact good initial concept acquisition is important in either mode.

Another important reason it may be difficult to convincingly demonstrate practically significant differences between epistemic modes for concept learning in classroom situations

has to do with "natural teacher variation."[10] Although our teachers all received intensive professional development in both topics in both modes, our study revealed clear teacher effects. The teachers met fidelity standards and yet were still noticeably different from each other in conduct of lessons, emphasis, explanatory ability, personality, and pedagogical content knowledge. The teacher effects on student performance, as described above, highlight this. Mixed patterns of results were found both between teachers on the same topics and within teachers on different topics and different assessment items, underscoring the idea that teachers have natural proclivities, talents, strengths, and weaknesses for various aspects of instruction. One teacher's students had good gains with both approaches, while other teachers got better results with one approach than with the other. It is a reasonable conjecture from our study that, with respect to student learning of core disciplinary ideas, the teacher is at least as important as instructional mode, if not much more so. Furthermore, our results suggest a potentially strong interaction between teacher and mode that influences efficacy.

While students made statistically significant learning gains in both units, there was a rather wide range of scores, and hence large standard deviations. This suggests that for a study to show convincing statistical significance for mode differences, learning gains as well as gain differences between modes would need to be considerably larger than those observed, and/or standard deviations would need to be smaller. Although a larger-scale study could provide larger N-size, this would be at the cost of precision, since in practical terms it becomes far more difficult to prepare, control, and monitor all the instructional and classroom situations and factors that can erode fidelity to mode. One might invoke an "uncertainty principle" analogy here, in that as the number of teachers and students involved goes up, teacher and classroom fidelity to mode becomes increasingly uncertain. Following Cronbach (1975), it could instead be more informative for this type of study to be carefully repeated a number of times in different locales with different teachers and topics. Even so, small efficacy differences between modes, such as we observed, even when statistically significant, would likely not be of as much practical classroom significance as teacher effects.

Besides the learning of core disciplinary ideas, advocates of either direct or inquiry forms of instruction would suggest that beyond concept acquisition, one or the other mode may be preferable because it provides other benefits. *Taking Science to School* (Duschl et al. 2007) argues that "one may be tempted to ask 'Is inquiry better than direct instruction'… but the critical question is 'Better for what?'" (p. 252). Indeed, inquiry is commonly thought to better represent scientific practices and the nature of science, and potentially be more interesting to students. Hence, while both approaches may lead to similar levels of conceptual understanding, many will argue that forms of inquiry can in addition model aspects of scientific inquiry and may also engender positive student attitudes, so that inquiry might potentially provide "added value" beyond conceptual understanding. In that case, one would privilege inquiry; but only if it could be done so that trying to learn content and process together was not cognitive overload and confusing, especially for novices, detracting from concept learning rather than adding value. It is thus not clear whether various added-value educational objectives can or should be attained by learners simultaneously with learning core content, or whether it makes more sense to target these objectives through activities designed especially for the purpose.

With regard to positive student attitude and appreciation of the nature of science, if, as is commonly done, one sets up a quadrant II vs. quadrant IV comparison (direct/rote vs. inquiry/

[10] We suspect that there is also "natural student variation." However, as our study did not look at student differences we will refrain from speculating.

meaningful), the value of inquiry would likely be superior to passive forms of direct instruction. What has not been considered is the effect of active-direct instruction on student attitudes, nor the possibility that such instruction can be formulated so as to support the teaching of certain aspects of the nature of science. Comparative Q-I to Q-IV studies (direct/meaningful vs. inquiry/meaningful) are lacking in this regard as well.

At the outset, we noted three possible research outcomes of this instructional efficacy study: one or the other of the two epistemic modes could result in better conceptual learning gains, or else the modes would prove comparable. We found that although concept learning paths differed, student learning gains were similar for active-direct and guided-inquiry modes. The latter had marginally higher conceptual gains, but the difference was not of practical educational significance. It was overshadowed by efficacy differences between teachers, irrespective of mode, and sometimes even by variations within each teacher across mode. This suggests the importance of teacher effects on student achievement, including teacher pedagogical content knowledge, natural style, personality, class management, etc., as well as interaction with mode, which appears to be quite strong. The findings also suggest that some previous claims for the superiority of one instructional mode or the other may be overstated, or may result from inappropriate comparisons against straw foils, or from confounded research with questionable designs. While our findings may disappoint advocates one way or the other, they are nonetheless important to know, and have implications for informed instructional design decisions for teaching and learning core disciplinary ideas.

Knowing that concept understanding can be achieved either way with good instruction, teachers can be more confident using their professional judgment in deciding how to teach various components of a lesson, appropriate to the nature of the topic, rather than feeling obliged to stick to an "approved" mode throughout, especially when this may be contrary to their natural inclinations. This realization is likely to be enabling in contrast to prescriptive or limiting. Teachers will have a good degree of flexibility to decide on one mode or the other to achieve content and/or process goals.

Since the alternative approaches to *content* also reflect science *practices* differently, the latter can factor into teachers' instructional choices. They can use methods appropriate to particular topics and contexts, e.g., inductive or deductive, in the knowledge that each of these can reflect various facets of real science practices. Similarly, developers of teaching materials can use appropriate modes to suit a range of purposes and goals. The findings, along with our analyses and interpretations of the issues, could lead to reevaluation and expansion of the legitimate range of approaches used in teaching science. This opens up possibilities in the repertoires of instructional methods used by teachers and the approaches taken by textbook authors, and could likewise affect teacher preparation programs.

Although our research finds that the two contrasting epistemic modes led to similar gains in understanding of core ideas, in our own teaching, we prefer a guided-inquiry approach of one form or another, whether as the overall method or for certain stages, and certainly in the initial framing of a new topic. We believe that even when a topic is to be treated directly, it is still best to approach it with an inquiring frame of mind. To quote Paul Tillich: "The fundamental pedagogical error is to throw answers, like stones, at the heads of those who have not yet asked the questions." In this spirit, a direct approach does not have to be cast as a rhetoric of conclusions but can relate meaningfully to context, questions, and purpose. An inquiring attitude can also be expressed in a degree of meta-level comment during instruction, complementing the object-level treatment of content.

In conclusion, we reiterate that instruction for core concept development usually takes one or the other of two fundamental epistemic approaches: some form of direct mode instruction or some form of inquiry mode instruction. When both approaches involved active student engagement, conceptual learning gains for core ideas were similar for both. Our conclusion is that well-designed instructional units, sound active-engagement lessons, good pedagogical content knowledge for specific concepts, and good teaching are at least as important for concept learning as epistemic mode. Thus, mode and narrative can be chosen as appropriate to the nature of the concept or activity, the learning goals, the learners, the situation, and professional judgment. Findings suggest that teachers need not be bound to one mode throughout and can choose the pedagogical approach on several grounds other than efficacy of content acquisition alone. Given that the contrasting modes reflect science practices differently and use different instructional narratives, these can be considerations in lesson design for particular topics. This leads us to suggestions for the nature of future work. It will clearly be useful to pursue further research regarding instructional mode characteristics, implementations, and efficacy, but there also needs to be a research focus on interactions between teacher, instructional mode, and students. There could also be a focus on the design of learning trajectories for specific science concepts, as a component of pedagogical content knowledge for teaching and learning those concepts. Research could also provide a better understanding of teachers' natural variations, proclivities, strengths and talents, and help them become knowledgeable about effective types of instruction for particular objectives and situations. This could inform teachers' lesson designs and classroom practice, and benefit student learning, enjoyment, and achievement in science.

Compliance with Ethical Standards

Funding Funded by the National Science Foundation's Interagency Education Research Initiative (IERI/NSF 04-553) Award #0437655. Any opinions, findings, conclusions, or recommendations in this paper are those of the authors and do not necessarily reflect the views of the NSF.

References

Abraham, M. R., & Renner, J. W. (1986). The sequence of learning cycle activities in high-school chemistry. *Journal of Research in Science Teaching, 23*(2), 121–143.

Adams, G. L., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond Distar*. Seattle: Educational Achievement Systems.

Alberts, B. (2011). A new college science prize. *Science, 331*(6013), 10.

Allende, J. E. (2008). Academies active in education. *Science, 321*(5893), 1133.

Alozie, N. M., Moje, E. B., & Krajcik, J. S. (2010). An analysis of the supports and constraints for scientific discussion in high school project-based science. *Science Education, 94*(3), 395–427.

American Association for the Advancement of Science (AAAS). (1990). *Project 2061: science for all Americans*. Washington, DC: American Association for the Advancement of Science, Inc.

American Federation of Teachers. (2003). Direct Instruction. URL http://people.uncw.edu/kozloffm/aftdi.html. Accessed 6 January 2003.

Anastasiow, N. J., Sibley, S. A., Leonhardt, T. M., & Borich, G. D. (1970). A comparison of guided discovery, discovery and didactic teaching of math to kindergarten poverty children. *American Educational Research Journal, 7*(3), 493–510.

Anderson, R. D. (2002). Reforming science teaching: what research says about inquiry. *Journal of Science Teacher Education, 13*(1), 1–12.

Anderson, R. D. (2003). Inquiry in the Everyday World of Schools. URL http://www.enc.org/topics/inquiry/context/document.shtm?input=FOC-000708-index. Accessed 14 June 2003.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Anderson, C. W., & Smith, E. L. (1987). Teaching science. In V. Richardson-Koehler (Ed.), *Educators' handbook: a research perspective* (pp. 84–111). New York: Longman.

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction: effective and efficient teaching*. New York: The Guilford Press.

Atkin, J. M., & Karplus, R. (1962). Discovery or invention? *The Science Teacher, 29*, 45–47.

Aulls, M. W. (2002). The contributions of co-occurring forms of classroom discourse and academic activities to curriculum events and instruction. 94(3), 520–538.

Ausubel, D. P. (1961a). In defense of verbal learning. *Educational Theory, 11*(1), 15–25.

Ausubel, D. P. (1961b). Learning by discovery: rationale and mystique. *NASSP Bulletin, 45*, 18–58.

Ausubel, D. P. (1962). In defense of verbal learning: beams in the eyes of a critic. *Educational Theory, 12*(4), 230–233.

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. NY: Grune & Stratton.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1986). *Educational psychology: a cognitive view* (2d ed.). New York: Werbel & Peck.

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided-inquiry and verification laboratory instruction. *Science Education, 94*(4), 577–616.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: creating excitement in the classroom AEHE-ERIC higher education report no. 1*. Washington, DC: Jossey-Bass.

Brady, T. E. (2008). Science education: Cassandra's prophecy. *Phi Delta Kappan, 89*(8), 605–607.

Brickman, P., Gormally, C., Armstrong, N., & Hallar, B. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence. *International Journal for the Scholarship of Teaching and Learning, 3*(2).

Brown, P. L., & Abell, S. K. (2007). Examining the learning cycle. *Science and Children*, 58–59.

Brown, A. L., & Campione, J. C. (1994). *Guided discovery in a community of learners. Classroom lessons: integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*(1), 21–32.

Bruner, J. S. (1971). *Some Elements of Discovery*. New York: W.W. Norton & Company, Inc.

Bybee, R. W. (2003). The teaching of science: content, coherence, and congruence. *Journal of Science Education and Technology, 12*(4), 343–358.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: origins, effectiveness and applications*. Colorado Springs, CO: BSCS.

Chen, Z., & Klahr, D. (1999). All other things being equal: acquisition and transfer of the control of variables strategy. *Child Development, 70*(5).

Chi, M., & Van Lehn, K. (2007). Accelerated future learning via explicit instruction of a problem-solving strategy. In R. Luckin, K. R. Koedinger, & J. E. Greer (Eds.), *Proceedings of 13th International Conference on Artificial Intelligence in Education* (pp. 409–417). Amsterdam, Netherlands: IOS Press.

Clark, N., Eyler, G., Rivas, A., & Wagner, T. (2011) Direct Instruction vs. Scientific Inquiry: Evaluating Student Outcomes. URL emurillo.org/Classes/Class2/DirectvsScientific.docx. Accessed 25 June 2012.

Clement, J. J. (1982). Student preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66–71.

Clement, J. J. (1998). Expert novice similarities and instruction using analogies. *International Journal of Science Education, 20*(10), 1271–1286.

Clement, J. J., Brown, D. E., & Zietsman, A. (1989). Not all preconceptions are misconceptions: finding 'anchoring conceptions' for grounding instruction on students' intuitions. *International Journal of Science Education, 11*(5), 554–565.

Cobern, W. W., Schuster, D. G., Adams, B., Applegate, B., Skjold, B., Undreiu, A., Loving, C. C., & Gobert, J. D. (2010). Experimental comparisons of inquiry and direct instruction in science. *Research in Science & Technological Education, 28*(1), 81–96.

Cobern, W. W., Schuster, D. G., Adams, B., & Skjold, B. (2012). The Pedagogy of Science Teaching Test (POSTT). URL http://www.wmich.edu/science/inquiry-items/. Accessed 13 July 2012.

Committee on Prospering in the Global Economy of the twenty-first Century: An Agenda for American Science and Technology. (2005). *Rising above the gathering storm: energizing and employing America for a brighter economic future*. Washington, DC: National Academy of Sciences, National Academy of Engineering, Institute of Medicine.

Committee on Prospering in the Global Economy of the twenty-first Century: An Agenda for American Science and Technology. (2010). *Rising above the gathering storm, revisited: rapidly approaching category 5*. Washington, DC: National Academy of Sciences, National Academy of Engineering, Institute of Medicine.

Costenson, K., & Lawson, A. E. (1986). Why isn't inquiry used more in classrooms? *The American Biology Teacher, 48*(3), 150–158.

Craig, R. C. (1956). Directed versus independent discovery of established relations. *Journal of Educational Psychology, 47*(4), 223–234.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116–127.

CSSP. (2002) Scientific Community Recognizes Direct Instruction. URL http://www.nifdi.org/pdfs/PressRel_CSSP.PDF. Accessed 13 July 2012.

Culatta, R. (2012) Dual Coding Theory (A. Paivio). URL http://www.instructionaldesign.org/theories/dual-coding.html. Accessed 22 June 2012.

Dagher, Z. R. (1995). Analysis of analogies used by science teachers. *Research in Science Teaching, 32*(3), 259–270.

Dagher, Z. R. (2005). The case for analogies in teaching science for understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Teaching science for understanding: a human constructivist view* (pp. 195–211). San Diego, CA: Academic Press.

DataWORKS Educational Research. (2012) Research Supporting EDI. URL http://www.dataworks-ed.com/research/support. Accessed 20 June 2012.

de Jong, T., Beishuizenm, J., Hulshof, C., Prins, F., van Rijn, H., van Someren, M., Veenman, M., & Wilhelm, P. (2005). Determinants of discovery learning in a complex simulation learning environment. In P. Gardenfors & P. Johansson (Eds.), *Cognition, education, and communication technology* (pp. 257–283). Mawah, NJ: Erlbaum.

Dean Jr, D., & Kuhn, D. (2006). Direct instruction vs. discovery: the long view. *Science Education, 91*(3), 384–397.

DeBoer, G. E. (1991). *A history of ideas in science education: implications for practice*. New York: Teachers College Press.

DiSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Putall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2-3), 105–225.

Dori, Y., Hult, E., Breslow, L., & Belcher, J. (2007). How much have they retained? Making unseen concepts seen in a freshman electromagnetism course at MIT. *Journal of Science Education and Technology, 16*(4), 299–323.

Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: learning and teaching science in grades K-8*. Washington, DC: Committee on Science Learning, Kindergarten through Eighth Grade, National Research Council.

Ebert-May, D., & Hodder, J. (2008). *Pathways to scientific teaching*. Sunderland, MA: Sinauer Associates, Inc., Publishers.

Education Consumers Foundation. (2011). *Direct instruction: what the research says*. Arlington, VA: Education Consumers Foundation.

Egan, D. E., & Greeno, J. G. (1973). Acquiring cognitive structure by discovery and rule learning. *Journal of Educational Psychology, 84*(1), 85–97.

Eisenkraft, A. (2003). Expanding the 5E model. *The Science Teacher, 70*(6), 56–59.

Etkina, E., & Van Heuvelen, A. (2007). Investigative science learning environment - a Science process approach to learning physics. In E. F. Redish, & P. Cooney (Eds.), *Research based reform of University physics* (pp. 1-48). College Park, MD: American Association of Physics Teachers.

Finn Jr., C. E., & Ravitch, D. (1996). *Education reform 1995–1996*. Washington, DC: Thomas B. Fordham Foundation.

Fleming, N. (2012) NAEP Reveals Shallow Grasp of Science. URL http://www.edweek.org/ew/articles/2012/06/19/36naep.h31.html?tkn=VLPFYOoO%2Fh6K0gBMoWRnkBNKB%2B3NDBvfmvWl&cmp=ENL-EU-NEWS1. Accessed 20 June 2012.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research, 82*(3), 300–329.

Guthrie, J. T. (1967). Expository instruction versus a discovery method. *Journal of Educational Psychology, 68*(1), 45–40.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*(6034), 1213–1216.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S. M., & Wood, W. B. (2004). Education: scientific teaching. *Science, 304*(5670), 521–522.

Harris, C., Cheng, B. H., Gong, B., DeBarger, A. H., Korbak, C., Lundh, P., Moorthy, S., Snow, E., & Stanford, T. (2012). Project-Based Inquiry Science Efficacy Study. URL http://ctl.sri.com/projects/displayProject.jsp?Nick=pbis. Accessed 21 June 2012.

Hassard, J. (2003). Direct-interactive teaching model. URL http://scied.gsu.edu/Hassard/mos/7.3.html. Accessed 17 December 2003.

Hattie, J. A. C. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Oxon and New York: Routledge.

Hein, G. (2004) The challenge of constructivist teaching: adapted with permission. In E. Mirochnik & D. C. Sherman (Eds.), *Passion and pedagogy: relation, creation, and transformation in teaching* (pp. 197–214). New York: Peter Lang. URL http://www.lesley.edu/faculty/ghein/papers_online/challenge_construct_2002/challenge_construct_2002.html. Accessed 8 March 2009.

Heppner, F. H., Kouttab, K. R., & Croasdale, W. (2006). Inquiry: does it favor the prepared mind? *American Biology Teacher, 68*(7), 390–391.

Holliday, W. G., & McGuire, B. (1992). How can comprehension adjunct questions focus students' attention and enhance concept learning of a computer-animated science lesson? *Journal of Research in Science Teaching, 29*(1), 3–15.

Ivins, J. E. (1985). A comparison of the effects of two instructional sequences involving science laboratory activities (Verification, Discovery, Information Processing). Unpublished doctoral dissertation, University of Cincinnati.

Jenness, M., & Barley, Z. A. (1999). Observing teaching practices in K-12 classrooms: instruments and methods (Science - version B: for multiple classroom observations to improve programming). Kalamazoo, MI: Science and Mathematics Program Improvement (SAMPI).

Kanter, D. E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: the effects of teacher content and pedagogical content knowledge and inquiry-based practices. *Science Education, 94*(5), 855–887.

Karplus, R. D. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching, 14*(2), 169–175.

Kelly, A. E., & Lesh, R. A. (Eds.). (2000a). *Handbook of research and design in mathematics and science education*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Kelly, A. E., & Lesh, R. A. (2000b). Clinical methods. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of research and design in mathematics and science education* (pp. 513–732). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Kersh, B. Y. (1962). The motivating effect of learning by independent discovery. *Journal of Educational Psychology, 53*(2), 65–71.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(1), 75–86.

Klahr, D. (2000). *Exploring science: the cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Klahr, D. (2002) Paths of Learning and their consequences: Discovery Learning versus Direct Instruction in elementary school science teaching. URL http://www.lrdc.pitt.edu/supergroup/#previoustalks. Accessed 10 April 2009.

Klahr, D. (2013). What do we mean? On the importance of not abandoning scientific rigor when talking about science education. *Proceedings of the National Academy of Sciences, 110* (Supplement_3), 14075–14080.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science, 15*(2), 661–667.

Klauer, K. J. (1984). Intentional and incidental learning with instructional texts: a meta-analysis for 1970-1980. *American Educational Research Journal, 21*(2), 323–339.

Krajcik, J. S., Mamlok, R., & Hug, B. (2001). Modern content and the enterprise of science: science education in the twentieth century. In L. Corno (Ed.), *Education across a century: the centennial volume*. Chicago: University of Chicago Press.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory Into Practice, 41*(4), 212–264.
Kuhn, D. (2007). Is direct instruction the answer to the right question? *Educational Psychologist, 42*, 109–113.
Latour, B. (1987). *Science in action: how to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard University Press.
Lawson, A. E. (2001). Using the learning cycle to teach biology concepts and reasoning patterns. *Journal of Biological Education, 35*(4), 165–169.
Lawson, A. E. (2010). *Teaching inquiry science in middle and secondary schools*. Washington, DC: SAGE.
Lee, H.-S., Linn, M. C., Varma, K., & Liu, O. L. (2010). How do technology-enhanced inquiry science units impact classroom learning? *Journal of Research in Science Teaching, 47*(1), 71–90.
Li, J., Klahr, D., & Siler, S. (2006). What lies beneath the science achievement gap? The challenges of aligning science instruction with standards and tests. *Science Educator, 15*(1), 1–28.
Llewellyn, D. (2007). *Inquiry within: implementing inquiry-based science standards in grades 3–8*. Thousand Oaks, CA: Corwin Press.
Marbach-Ad, G., & Claassen, L. A. (2001). Improving students' questions in inquiry labs. *American Biology Teacher, 63*(6), 410–419.
Marx, J., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*, 87–91.
Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in the middle grades: assessment of learning in urban systemic reform. *Journal of Research in Science Teaching, 41*(10), 1063–1080.
Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: is it all in the timing? *Instructional Science, 41*(3), 621–634.
Mayer, R. E. (1979). Twenty years of research on advance organizers: assimilation theory is still the best predictor of results. *Instructional Science, 8*(2), 133–167.
Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19.
McGrew, T., Alspector-Kelly, M., & Allhoff, F. (2009). *Philosophy of science: an historical anthology*. Malden, MA: Wiley-Blackwell.
Michigan State Board of Education. (2004) Science v 1.09 Grade Level Content Expectations. URL http://www.michigan.gov/documents/mde/Complete_Science_GLCE_12-12-07_218314_7.pdf.
Minner, D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis. *Journal of Research in Science Teaching, 47*(4), 474–496.
Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2000). *Assessing science for understanding: a human constructivist view*. San Diego, CA: Academic Press.
Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2005). *Teaching science for understanding: a human constructivist view*. San Diego, CA: Academic Press.
Moscovici, H., & Nelson, T. H. (1998). Shifting from activity-mania into inquiry. *Science and Children., 35*(4), 14–17.
Mueller, J. F. (n.d.) Madeline Hunter's Lesson Plan Format. URL http://jfmueller.faculty.noctrl.edu/205/madelinehunter.htm. Accessed 27 August 2013.
National Center for Education Statistics. (2011). *The nation's report card: trial urban district assessment science 2009* (NCES 2011–452). Washington, D.C: Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
National Research Council (1996). *National Science Education Standards*. Washington, DC: National Research Council.
National Research Council (2000a). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academies Press.
National Research Council (2001a). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academies Press.
National Research Council (2001b). *Knowing what students know: the science and Design of Educational Assessment*. Washington, DC: National Academies Press.
National Research Council (2002). *Scientific research in education*. Washington, DC: National Academies Press.
National Research Council (2013). *Next generation science standards: for states, by states*. Washington, DC: The National Academies Press. doi:10.17226/18290.
National Research Council (Committee on Conceptual Framework for the New K-12 Science Education Standards) (2012). *A framework for K-12 science education: practices, crosscutting concepts, and Core ideas*. Washington, DC: The National Academies Press.
Nock, G. A. B. (2009). The effects on community college student physics achievement and attitudes about learning physics due to inquiry-based laboratory activities versus cookbook laboratory activities. Unpublished doctoral dissertation, The University of Mississippi, University, MS. http://adsabs.harvard.edu/abs/2009PhDT.....11N.

Novak, J. D. (1976). Understanding the learning process and effectiveness of teaching methods in the classroom, laboratory, and field. *Science Education, 60*(4), 493–512.
Novak, J. D. (1977). An alternative to Piagetian psychology for science and mathematics education. *Science Education, 61*(4), 453–477.
Novak, J. D. (1979). Position papers: the reception learning paradigm. *Journal of Research in Science Teaching, 16*(6), 481–488.
Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching, 47*(4), 422–453.
Padilla, M. (2013). Inquiry, process skills, and thinking in science. In L. Froschauer (Ed.), *A year of inquiry* (pp. 24–26). Arlington, VA: NSTA Press.
Peterson, P. E. (2011). Eighth-grade students learn more through direct instruction. *Education Next, 11*(3).
Randolph-Macon Woman's College. (2003) Direct instruction [Web Page]. [2003, December 17].
Renner, J. W., & Marek, E. A. (1990). An educational theory base for science teaching. *Journal of Research in Science Teaching, 27*(3), 241–246.
Robertson, B. (2006). Getting past "inquiry versus content". *Educational Leadership, 64*(4), 67–70.
Roehrig, G. H., & Luft, J. A. (2004). Constraints experienced by beginning secondary science teachers in implementing scientific inquiry lessons. *International Journal of Science Education, 26*(1), 3–24.
Rosenshine, B. (2008). *Five Meanings of Direct Instruction*. *Center on Innovation & Improvement*, Academic Development Institute.
Rudolph, J. L. (2002). *Scientists in the classroom: the cold war reconstruction of American science education*. New York: Palgave.
Rutherford, F. J. (1964). The role of inquiry in science teaching. *Journal of Research in Science Teaching, 2*, 80–84.
Schneider, B. (2004). Building a scientific community: the need for replication. *Teachers College Record, 106*(7), 1471–1483.
Schneider, S. (2011) Making Sense of SCIENCE: Efficacy Study of a Professional Development Series for Middle School Science Teachers. URL http://ies.ed.gov/funding/grantsearch/details.asp?ID=1186. Accessed 21 June 2012.
Schroeder, C. M., Scott, T. P., 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. *Journal of Research in Science Teaching, 44*(10), 1436–1460.
Schwab, J. J. (1962). The teaching of science as enquiry. *The Inglis Lecture*. Cambridge, MA: Harvard University Press.
Schwab, J. J. (1978). (reproduced essays) Westbury, I., & Wilkof, N., J. (eds.). *Science, Curriculum, and Liberal Education: Selected Essays*. Chicago, IL: University of Chicago Press.
Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.
Schwerdt, G., & Wuppermann, A. C. (2011). Sage on the stage—is lecturing really all that bad? *Education Next, 11*(3).
Secker, C. V. (2002). Effects of inquiry-based teacher practices on science excellence and equity. *Journal of Educational Research, 95*(3), 151–160.
Secker, C. V., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching, 36*(10), 1110–1126.
Settlage, J. (2007). *Moving past a belief in inquiry as a pedagogy. Inquiry in the classroom: realities and opportunities* (pp. 233–248). Charlotte, NC: Information Age Publishing.
Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56*(4), 411–436.
Shulman, L. S., & Keislar, E. R. (1966). *Learning by discovery: a critical appraisal*. Chicago: Rand McNally & Company.
Shymansky, J. A., Kyle, W. C. J., & Alport, J. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching, 20*(5), 387–404.
Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching, 27*(2), 127–144.
Stone, C. L. (1983). A meta-analysis of advance organizer studies. *Journal of Experimental Education, 51*(4), 194–199.
Sweller, J. (2009). What human cognitive architecture tells us about constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: success or failure?* (pp. 127–143). NY: Routledge.
Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: a reply to commentaries. *Educational Psychologist, 42*(2), 115–121.
Tai, R. H., & Sadler, P. M. (2009). Same science for all? Interactive association of structure in learning activities and academic attainment background on college science performance in the USA. *International Journal of Science Education, 31*(5), 675–696.

Tennyson, R. D., & Cocchiarella, M. J. (1986). An empirically based instructional design theory for teaching concepts. *Review of Educational Research, 56*(1), 40–71.

Thorley, N. R., & Stofflett, R. T. (1996). Representation of the conceptual change model in science teacher education. *Science Education, 80*(3), 317–339.

Timmerman, B. E., Strickland, D. C., & Carstensen, S. M. (2008). Curricular reform and inquiry teaching in biology: where are our efforts most fruitfully invested? *Integrative and Comparative Biology, 48*(2), 226–240.

Tretter, T. R., & Jones, M. G. (2003). Relationships between inquiry-based teaching and physical science standardized test scores. *School Science and Mathematics, 103*(7), 345–350.

Trowbridge, J. E., & Wandersee, J. H. (2005). Theory-driven graphic orgznizers. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Teaching science for understanding: a human constructivist view* (pp. 95–131). San Diego, CA: Academic Press.

U.S. Department of Education, N. C. f. E. S. (2012). The Condition of Education 2012. URL http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012045. Accessed 20 June 2012.

Udovic, D., Morris, D., Dickman, A., Postlethwait, J., & Wetherwax, P. (2002). Workshop biology: demonstrating the effectiveness of active learning in an introductory biology course. *Bioscience, 52*(3), 272–281.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring system: lessons learned. *International Journal of Artificial Intelligence and Education, 15*(3), 1–47.

Walberg, H. J. (1991). Improving school science in advanced and developing countries. *Review of Educational Research, 61*(1), 25–69.

West, L. H. T., & Fensham, P. J. (1974). Prior knowledge and the learning of science: a review of Ausubel's theory of this process. *Studies in Science Education, 1*, 61–81.

White, B. Y., & Frederickson, J. R. (1998). Inquiry, modeling, and metacognition: making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

Wittrock, M. C. (1964). *The learning by discovery hypothesis*. Washington, D.C: ERIC Clearinghouse.

Wright, J. (2013) Introducing Academic Strategies to Students: A Direct-Instruction Approach. URL http://www.interventioncentral.org/academic-interventions/general-academic/introducing-academic-strategies-students-direct-instruction-. Accessed 27 August 2013.

Wright, C. J., & Nuthall, G. (1970). Relationships between teacher behaviors and pupil achievement in three experimental elementary science lessons. *American Educational Research Journal, 7*(4), 477–491.

Yale University. (2012). The Center for Scientific Teaching. URL http://scientificteaching.yale.edu/about-us. Accessed 13 July 2012.

Yeany, R. H., & Miller, P. A. (1983). Effects of diagnostic/remedial instruction on science learning. *A meta analysis., 20*(1), 19–26.