Active Learning in Science: An Experimental Study of the Efficacy of Two Contrasting Modes of Instruction

William W. Cobern,1* David Schuster,1 Betty Adams,1 Adriana Undreiu,2 Brandy Skjold,1 Brooks Applegate3

1The Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI 49008, USA
2University of Virginia’s College at Wise, Wise, VA 24293, USA
3Educational Leadership, Research and Technology, Western Michigan University, Kalamazoo, MI 49008, USA

This paper reports the findings of a research project 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.

*Corresponding author: William W. Cobern <bill.cobern@wmich.edu>
Active Learning in Science: An Experimental Study of the Efficacy of Two Contrasting Modes of Instruction

Abstract

The science education community has overwhelmingly adopted a guided-inquiry perspective on science teaching. Nevertheless there remains debate about instructional approaches across a spectrum from direct instruction through guided inquiry to open discovery. Questions about the efficacy of inquiry-based instruction linger. Proponents point to supportive studies, but critics counter that too little of this research is unconfounded, comparative or sufficiently controlled to allow for valid conclusions. In various states there is political pressure for a return to direct instruction. The lack of convincing comparative evidence for inquiry is thus of concern. To complicate matters, usage of the terms ‘inquiry’ and ‘direct’ tends to be varied and vague, rarely providing operational definitions or models of either. We further suggest that confusion enters the discussion when the ideas of ‘active’ or ‘hands-on’ learning, which have evidentiary support, are tacitly conflated with inquiry, and likewise ‘passive’ is tacitly conflated with direct. We therefore conducted a six-year randomized controlled study of the efficacy of carefully designed inquiry instruction and equally carefully designed direct instruction in science classroom settings at the middle school grades. The contrasting modes were operationally defined by explicit models, and both involved active learning though approached differently. We developed two-week instructional units in parallel in the two modes for two science topics, produced student materials and teacher guides, and provided teacher professional development. The units were implemented in a summer program with students randomly assigned to guided-inquiry or active-direct classrooms/teachers. We report on the contrasting instructional units, assessments, research design, classroom implementation, teacher fidelity, analyses and results.

Keywords: science education, inquiry, direct instruction, active learning, pedagogy, comparative experimental study, randomized controlled trial, middle school

Introduction

For over a hundred years, educational and political policy debates have raged over the relative merits of ‘inquiry-based’ and ‘direct’ approaches to science instruction, with strong opinions on both sides. Much past practice and instructional materials have reflected direct instruction, but in recent years, with the formulation of national and state science education standards (AAAS, 1990; National Research Council [NRC], 2002a) inquiry-based instruction has become the sine qua non of science teaching (Handelsman, 2004). Inquiry-based instruction in science is understood as reflecting the investigative approach and reliance on evidence that scientists use constructing new knowledge (NRC, 2002b). Students explore phenomena and from their investigations arrive at science concepts and principles, guided by the teacher. A related idea is that of ‘Scientific Teaching,’ which suggests that “the teaching of science should be faithful to the true nature of science by capturing the process of discovery in the classroom” (Handelsman et al., 2004; Handelsman, Miller, & Pfund, 2007). That is, guided-inquiry instruction models important aspects of scientific inquiry in developing science content. By contrast, in direct instruction the teacher explains the relevant science concepts directly to students upfront, as an established body of knowledge to be learned and understood. The role of subsequent student activities is usually practical verification of this theory and its application to problem-solving. However, there is ongoing confusion about terminology and a variety of interpretations of what exactly is meant by these two modes of instruction. There is also very little unconfounded comparative research that supports one mode over the other.
We wish to highlight crucial distinctions between concepts that are often conflated. Inquiry-based instruction, as understood above to reflect important aspects of scientific inquiry, is not synonymous with ‘active learning’ nor with ‘hands-on activities.’ Similarly, direct instruction is not synonymous with didactic presentation to passive receivers. That is, well-designed direct instruction should involve active learning and practical work, and poorly designed inquiry lessons could involve passive following of ‘scientific method’ stages. Implementation of both methods can and should involve student experiences with phenomena and active engagement with ideas, otherwise they would simply rate as poor science instruction.

Unfortunately, both of the conflations noted above are prevalent in instruction and research, making many results ambiguous or impossible to interpret. Thus the question we address in our work is not whether active, experiential learning of science is more effective than passive, non-experiential learning. That question has been unequivocally answered in favor of the former. Rather, we asked whether an inquiry or a direct approach to active, experientially-based learning is the more effective for science concept development, when both approaches are expertly designed and well taught.

Literature Review and Discussion

Inquiry-Based Instruction

What is generally meant by inquiry instruction? Inquiry instruction in science refers to an approach that reflects the investigative attitudes, empirical techniques and reasoning that scientists use to discover and construct knowledge. While learning science in school is not the same as doing real science, it can give a sense of how science produces new knowledge, if the approach models some of the features of scientific inquiry. The National Science Education Standards (NRC, 2000) describes five essential features of inquiry-based instruction and for each lists a spectrum of possible ‘levels’ of implementation, depending on the degree of teacher directedness. Hence, inquiry instruction may take on a variety of forms, and different forms may be appropriate for different aspects of learning. Note that while each of the features of inquiry should ideally be present somewhere in the instruction, the ‘level’ should be appropriate to the educational purpose at hand and is not a merit scale.

It can also be noted that in science itself there are different kinds of inquiry investigations with different purposes. One kind of science inquiry aims to discover or develop the science itself, while another kind applies the science to answer specific questions of interest. The same is true of science education. An inquiry-based approach to a topic has learners develop the fundamental concepts and principles of a domain. This discovery process is guided by the instructor within a coherent sequence of topic development. A rather different type of student inquiry task is an applied investigation. Here the basic concepts have already been learned, and the subsequent applied investigation involves science process skills. Unfortunately the distinction between these different activities with different purposes is often not recognized when talking about ‘inquiry-based’ instruction, nor when designing, researching or evaluating it, and hence interpretations of research findings may be problematic.

Historical use of inquiry instruction in science classrooms. The origins of the modern day concept of science teaching as inquiry lie with the 1960s NSF-funded curriculum projects (Anderson, 2003; DeBoer, 1991; Krajcik, Mamlok, & Hug, 2001; Rudolph, 2002; Schwab, 1962). The roots of inquiry learning run even deeper. 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, the gathering of data, rational argument, and the drawing of inferences from evidence and data. Thomas Huxley spoke of scientific training as “practicing the intellect in the completest form of induction...in drawing conclusions from particular facts made known by immediate observation of...
Laboratory activities became ubiquitous in science instruction on the basis that effective pedagogy must reflect the true 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). Thus it was reasoned that science instruction must be more than the clear explication of information, it must include the investigative processes that lead to concept development. In short, science instruction should be significantly experience-based and generate ideas via exploration, rather than presenting science as a final body of knowledge.

Since the late 1800s we have had a century of shifting curriculum tides and recurring controversies, with few certainties about how science should be taught at K–12 levels. However, in recent years there has developed a clear commitment to the pedagogy of inquiry. Indeed, under National Research Council and AAAS leadership, the United States 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, 2000, 2001). The science education community has overwhelmingly adopted an inquiry pedagogy perspective for science education, including the National Science Teachers Association (NSTA), the National Association for Research in Science Teaching (NARST), and the Association of Science Teacher Educators (ASTE).

**Why use inquiry-based instruction?** An inquiry-based instructional approach can be used to achieve a variety of purposes. It is a way to approach science conceptual development, while at the same time modeling aspects of scientific inquiry. There is also a possible motivational aspect; students may become curious and more interested. Furthermore if learners are guided to arrive at an idea themselves they might understand it better and retain it longer. Many educators feel that inquiry instruction rather than direct is most in keeping with the widely accepted constructivist theory of how people learn, i.e., that meaningful knowledge cannot simply be transmitted and absorbed but learners have to construct their own understanding. Thus Llewellyn (2007) states: “For many teachers, the principles of constructivism lay the foundation for understanding and implementing inquiry-based learning.” (p. 53) Note however that constructivism is a theory of learning rather than of instruction, and students must construct their own understanding regardless of the instructional inputs.

Despite the potential benefits, there are some problems with the use of inquiry instruction. For example, if inquiry is too open-ended, students have difficulties choosing correct variables to work with, forming testable research questions, linking hypotheses and data, and drawing correct conclusions from experiments (de Jong et al., 2005). Students can become lost and frustrated, and their confusion can lead to misconceptions (Brown & Campione, 1994). As a result, teachers may spend considerable time scaffolding students’ content and procedural skills (Aulls, 2002). Furthermore researchers and teachers alike have a wide range of conceptions about what actually constitutes inquiry-based instruction. Some judge ‘inquiry instruction’ more in terms of teacher behaviors than as a scientific-inquiry approach to learning science concepts. Poor inquiry practices may involve time-consuming student ‘process’ activities that develop little science. Detractors widely see inquiry as inefficient for developing the basics.

Varied interpretations and practices notwithstanding, inquiry is omni-present in the language of science education. As of 2002, Anderson asserted that research regarding the teaching of science had matured and thus “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). Implicit is the assumption that the case in favor of inquiry is firmly established. In 2008 Brady stated that, “We know how to teach science” implying that we know to teach science by inquiry. The only question left to ask appeared to be, “when will we do it?” (p. 607).
Similar views have been expressed in the editorial pages of SCIENCE (e.g., Allende, 2008). Settlage says about science methods textbooks that it would be “unlikely today that any methods book would have a chance in the marketplace if ‘inquiry’ was not prominent” (2007, p. 464). However, we believe that some assertions about the effectiveness of inquiry stretch the fabric of the research results.

**Direct Instruction**

**What is generally meant by direct instruction?** Direct instruction is the historic nemesis of inquiry, though it has been the approach most commonly used in the classroom. There is no single definition or model of direct instruction, though it is often taken to imply content-centered exposition, with the teacher providing the content directly. In a direct method, students are presented with the science concepts and principles up front rather than being guided to discover or generate it themselves. Common examples of direct instruction include the use of lecture, demonstration and/or audiovisual presentation, along with explanation, to provide information directly to students, and corresponding laboratory activities are highly structured. Note that in each case the instructor directs students’ actions and the science content is presented and explained as a known product, while lab activities are seen as verification of established knowledge. A criticism of much direct instruction is that it presents science as end-product knowledge, while neglecting the process aspects of science by which such knowledge is attained.

Compared to the case of inquiry science instruction, there have been fewer explicit attempts to describe desirable best practices for direct science instruction. There are general models for effective direct instruction, though not specifically for science, that can provide structures for planning direct lessons. The model developed by Hunter (2011) is based on observations of the practices of successful teachers. It includes the following: Review; Anticipatory set; Objectives/Purpose; Input; Modeling; Checking understanding; Guided practice and monitoring; Closure; Independent practice. Most educators, inquiry-oriented or not, would hardly disagree with many of Hunter’s elements of effective teaching. A well-known influential direct instruction model was developed in the 1960s by Engelmann and Becker, and used in an extensive federally-funded research and implementation program called Project Follow Through, in which DISTAR (Direct Instruction System for Teaching Arithmetic and Reading) gained prominence. Instruction in science was not involved though the principles would likely apply.

**Confusions about direct instruction.** When science education specialists work with teachers, they talk about moving teachers away from behaviors considered forms of direct instruction and toward inquiry. The commitment to inquiry is both understandable and laudable given that inquiry instruction has a natural appeal to anyone who loves both science and the teaching of science, and because inquiry naturally provides an active learning environment. However, some of the support for inquiry is based on a distorted view of expository or direct instruction. For example, discussions about *Benchmarks for Science Literacy* and the *National Science Education Standards* when they were first introduced noted their “focus on inquiry-based science - rather than memorization” (Brady, 2008, p 607). This implies that 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 memorization (e.g., Hein, 2004).

This misrepresentation of teaching approaches was addressed by David Ausubel (1961, 1962) more than forty years ago. He argued that the true issue was meaningful learning vis-à-vis rote learning and that neither inquiry instruction nor direct instruction automatically leads to meaningful
learning. Instructional approaches can thus occupy various quadrants in Figure 1. Novak (1976, 1977) later adopted and elaborated Ausubelian views. Meaningful learning required that teaching of concepts be coherently organized and that students actively fit the concepts into their cognitive structures. Because Ausubel and Novak believed that learning from direct instruction could be either meaningful or rote, they developed the ideas of advance organizers and concept mapping as instructional tools for fostering meaningful learning, for any method. (Note that the Ausubelian literature uses the terms ‘discovery’ and ‘reception’ teaching where we use the current language of ‘inquiry’ and ‘direct’ teaching.)

Why use direct instruction? There is research support for direct instruction, though there has been considerable debate about its efficacy and the empirical research behind it. A meta-analysis by Adams and Engelmann (1996), finds a "mean effect size average per study of more than .75, which confirms that the overall effect is substantial.” VanLehn et al. (2005) found that explicit training in problem-solving was successful at helping students set up free-body diagrams, write equations, and solve them, and that this accelerated learning, also transferred to new areas (Chi & VanLehn, 2007). Also see Anastasiow, Sibley, Leonhardt, and Borich (1970), Klauer (1984), Walberg (1991), and Wright and Nuthall (1970). Schwartz and Bransford (1998) argue that ‘telling’ (direct exposition and explanation) can be very effective for learning once students are adequately prepared to actively process and accommodate the information. Moreover, direct instruction finds support in other areas of education such as developments from DISTAR, a 1960s project that was part of President Johnson's War on Poverty (Adams & Engelmann, 1996; American Federation of Teachers, 2003; Finn & Ravitch, 1996). According to these educators, effective instruction must be more teacher-led than student-directed. These views have an audience; in 2002, the recipient of the Council of Scientific Society Presidents annual award for Education Research was Siegfried Engelmann, who was cited for his research and development on direct instruction. It is also widely believed that direct instruction is more efficient than inquiry timewise.

Empirical Comparisons of Inquiry and Direct Instruction

Since both methods of instruction seem to have evidentiary support behind them, a number of studies have sought to explicitly compare the two in a controlled way, to determine if one is significantly better than the other in developing science content understanding and skills.

Recent studies in support of direct instruction are those of David Klahr (Chen & Klahr, 1999; Klahr, 2000, 2002), who found that children can more successfully learn the control of variables strategy with direct instruction, leading to better understanding of science principles and transfer to other topics (Klahr & Nigam, 2004). Klahr thus cautions that the “widespread belief that discovery learning is superior to Direct Instruction in early science education warrants careful empirical assessment.” Based on his research on the acquisition of science skills (such as controlling variables), he concludes that these “results challenge predictions derived from the presumed superiority of discovery approaches for deeper, longer lasting, and ‘more authentic’ understanding of scientific
reasoning processes, and suggest instead, a more nuanced examination of the most effective mixes and the most suitable matches between topic and pedagogy” (2002, p. 1, emphasis added).

Klahr’s assertions are not without support; see for example, Kirschner, Sweller and Clark (2006), ‘Why Minimal Guidance during Instruction Does Not Work,’ Mayer (2004), and Sweller, Kirschner and Clark (2007). However, note that Klahr’s research actually involves ‘open discovery’ approaches that are the most unstructured form of inquiry, and not the more guided approach to inquiry advocated by the NRC and AAAS. Also, his ‘direct’ mode arguably resembles aspects of strongly guided inquiry. Furthermore, Klahr’s work is about acquiring science process skills, not about science subject matter concept development. Thus, on both counts, Klahr’s findings do not speak to the core question regarding inquiry instruction. Similarly, Sweller (2009) summarizes research findings regarding problem-solving and the use of worked-examples concluding that the superior effectiveness of direct instruction is supported by both empirical findings and cognitive theory; but also implies that this conclusion legitimately extends to other learning outcomes such as concept development. Our view is that, whatever the findings of Klahr or Sweller in their respective areas of interest, we do not think that conclusions can be drawn about science concept learning unless there is research specifically focused on it. Nevertheless, both Klahr and Sweller draw attention to the precarious evidentiary support for inquiry instruction, and the fact that Inquiry Instruction versus Direct Instruction for science concept development has not been subjected to experimental controlled studies comparable to Klahr’s work or the work referred to by Sweller.

Part way through our present controlled comparative study and before teachers switched instructional modes, we published the concept and initial work done up till that point in a short paper entitled Experimental Comparison of Inquiry and Direct Instruction in Science (Cobern et al., 2010). At that early stage we only had two trials and small numbers and had not yet incorporated the essential control for teacher effects (mode switching). Although the preliminary findings were indicative of some of what we report now, the important issue of the roles of instructional mode versus teacher effect remained.

Problems with the Research and its Interpretation

There are several problems with some of the research on the educational effectiveness of different types of science instruction. Studies often conflate the idea of ‘active learning’ with ‘inquiry,’ leading to research where students are actively engaged in the inquiry lessons, but passive in the direct. As discussed previously, Ausubel (1961, 1962) and Novak (1976, 1977) argued in the 1960s and 1970s that effective conceptual learning meant meaningful learning as opposed to rote learning. Over time, however, the idea that direct instruction led to rote memorization stuck. Soon after came the dichotomy of passive learning versus active learning. Research has clearly demonstrated that learning is better when students are actively engaged cognitively, instead of simply memorizing (Bonwell & Eison, 1991; Bruner, 1961). However, direct instruction was again automatically linked to passive learning. We therefore suggest that confusion has entered the discussion on effective science instruction because the idea of active learning is typically conflated with inquiry, and passive with direct. We argue that effective direct instruction can and should involve both active engagement and hands-on science experiences.

While much time, effort and money was devoted to various curriculum and instruction projects, such as the 1960s NSF-funded curriculum projects, little of this was allotted to the scientific study of instructional effectiveness (see Welch, 1976; Welch & Walberg, 1972). Research evidence at the time and over the next decade regarding the effectiveness of inquiry as a strategy of science instruction was mixed and sometimes negative (Anastasiow et al., 1970; Ausubel, 1961 & 1962; Craig, 1956; Kersh,
Still later, the Ivins (1985) research findings failed to support the superiority of inquiry instruction for conceptual development. Although the research findings of Shymansky, Kyle, and Alport (1983) and Shymansky, Hedge, and Woodworth (1990) on the effectiveness of the NSF-funded curricula using meta-analysis techniques were supportive of inquiry, their studies were also open to some criticism, in part because many assumptions were necessary. Moreover, the studies were not specifically focused on inquiry instruction but were general curriculum program evaluations. More recent meta-analyses of inquiry science teaching have similar shortcomings (Springer, Stanne & Donovan, 1999). Recent research by Secker and Lissitz (1999), Secker (2002), Tretter and Jones (2003), Udovic, Morris, Dickman, Postlethwait, and Wetherwax (2002), and White and Fredrickson (1998), are more focused on inquiry and again the results are supportive of inquiry, but even these studies do not yield inferences sufficiently unconfounded that one can feel that the central issue about inquiry instruction has been adequately addressed. Moreover, the Minner, Levy, and Century (2010. p. 14) synthesis study noted that “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, particularly in the data quality items assessed...” Also see Mervis (2004).

Finally, while the research studies conducted on either inquiry or direct pedagogy do provide data on the educational outcomes of that method, most of them are not comparative controlled studies, which would test one clearly-defined method against another, in the same controlled circumstances.

The nature of many research studies on inquiry and direct instruction and the various design and interpretation flaws found in the science education research literature can result in great difficulty drawing valid conclusions from the data. The problems with some of the research can be summarized as follows.

- Only some studies are comparative, so control or alternative treatment is absent.
- Of the comparative studies, few pit inquiry against worthy or well-defined alternative instruction.
- There may be multiple confounding factors involved besides instructional mode. For example, some teachers may undergo professional development, while the others will not, or the new method will include features or enhancements that the alternative does not.
- Few studies use random assignment of subjects to treatment groups, control for confounding factors, or use quasi-experimental efforts to deal with differences between groups.
- If there is insufficient detailed specification of treatment and control instructional models then the study and its results may not be repeatable by other researchers.
- Tacit assumptions of teacher fidelity to instructional mode may be unwarranted; actual lesson implementation is too rarely compared to the intended instruction.
- In some reported studies, the instruction involves extended periods (a semester) or very short periods. The difficulty here is one cannot be sure what the ‘active agent’ is in a long series of lessons. In such a situation it is difficult to know what caused any overall learning differences. Conversely, just one or two school periods are not enough to develop a significant science sequence, nor for students to adapt to what is expected of them in an approach.
- Vague and ambiguous terminology, conflations of constructs, and lack of operational definitions of type of instruction.
- Investigating only one type of learning task and concluding that the results apply to all learning in the domain.
• Assessments of effectiveness which are carried out entirely by the researchers and/or instructional developers involved.

**The Present Study: Features of Method**

**Aim and Method**

Noting the widespread advocacy of inquiry-based instruction, the variety of practices claimed to be inquiry, and the continuing prevalence of various forms of direct instruction, we became more aware of the problematic evidentiary foundation for either instructional mode. Therefore we conducted a controlled experimental study to compare the efficacies of guided-inquiry and active-direct instruction for science conceptual development, focusing at the eighth grade level. Our research responded to calls to employ quantitative research designs with a premium on experimental controlled studies with random assignment of subjects (see American Educational Research Association [AERA], 2009). The project comprised several phases: project design, development of instructional units in contrasting modes, student learning materials, teacher guides, lab equipment, teacher preparation, classroom implementation, fidelity measures, data collection, analyses, findings, and interpretation. The research that we are now reporting contrasts two approaches to active learning, with Inquiry Instruction as one treatment versus Direct Instruction as the alternative treatment, the two treatments being defined by operational models and carefully designed in parallel. We addressed the research question via classroom field studies (see Anastasiow et al., 1970). If the arguments for either mode are correct, then that mode should lead to learning gains that are greater than those of the other mode, by an amount that is of practical classroom significance. This is in effect a ‘treatment-treatment-control’ study, where we have two alternative treatments and the ‘control’ (i.e. no instruction) is essentially provided by student pre-scores, and is the baseline from which to measure learning gains in the two alternative treatments.

**General Methodological Features**

To guide the research and protect against the numerous threats to validity that often plague educational research, we built four general methodological features into our research. These features were: specificity, fidelity, objectivity, and transparency.

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

**Fidelity.** We do not assume that instruction will be implemented as intended. Important for our purposes are fidelity to instructional mode and fidelity to curriculum/content, along with teaching quality generally. In our research we followed the rule of ‘prepare and verify,’ addressing fidelity issues in teacher preparation beforehand as well as evaluating it observationally at the trials.

**Objectivity.** Our research design embedded several areas of ‘blindness’ in order to minimize possible bias. a) Teachers were blind to the tests. b) Assessment items were independently reviewed for content and construct validity, for fit with learning objectives, and for any possible bias favoring one instructional mode over the other. Instrument validity and reliability were also established
independently of the researchers. c) External evaluators coded and analyzed student assessment data without knowledge of teacher or mode assignments, and independently of the researchers. d) Teaching fidelity was monitored and evaluated by independent observers.

Transparency. Transparency provides other researchers with enough detail to know exactly what the work comprised and thus enable possible replication attempts or checks on method or results. We have made our research as transparent as possible by providing links to our webpages at http://www.wmich.edu/way2go/, where project information is posted along with the complete set of learning objectives, student materials, teacher guides, and assessments.

The Structural Design of the Study

The structural design of our study is shown in Table 1. During the first year, instructional units were developed in each mode, and student and teacher materials produced. This was accompanied by development of assessments aligned to the instruction and learning objectives. This was followed by a pilot year where teachers prepared to teach the units in assigned modes, and then implemented them in classrooms in a pilot summer program. Four years of main trials to compare mode efficacy followed. The basic design involves two instructional units A and B, taught each of the four years, by four main teachers, with two teaching by inquiry and two by direct. A fifth teacher and class was included to handle enrollment overflow and/or possible teacher loss. Since no absence occurred, the fifth data set was also included in the study. All five teachers switched modes after the first two trials.

Table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Unit</th>
<th>Project activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Dynamics (force &amp; motion)</td>
<td>unit and assessment development</td>
</tr>
<tr>
<td></td>
<td>B Light (climate &amp; seasons)</td>
<td>unit and assessment development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher 1</td>
<td>Teacher 2</td>
</tr>
<tr>
<td></td>
<td>A Direct</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td>B Direct</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td>Teacher 3</td>
<td>Teacher 4</td>
</tr>
<tr>
<td></td>
<td>A Direct</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td>B Direct</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td>Teacher 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A Direct</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td>B Direct</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Inquiry</td>
</tr>
</tbody>
</table>

Specification of Instructional Models

Single-word descriptors for instructional type, such as ‘inquiry,’ ‘direct,’ ‘traditional,’ ‘conventional,’ etc., are ambiguous and open to interpretation. To be clear and specific about what we mean by inquiry and direct science instruction we needed to characterize them operationally with explicit models. Below we describe the nature of our inquiry and direct instruction in terms of clear
models with specified features.

Model of guided-inquiry instruction. Our inquiry lessons were based on the Karplus Learning Cycle (1977), which involves three main phases, Exploration, Concept Formation, and Application (Figure 2). The exploration and concept formation phases reflect a scientific inquiry approach to a topic. These activities help students develop the relevant scientific concepts and ‘discover’ the relationships and laws, guided and scaffolded by the instructor. These phases are inquiry-based, and inductive (though not purely so). The Application phase, where students develop the ability to apply the science and solve problems, is mostly deductive. Note that the Karplus cycle is also at the heart of the BSCS ‘5-E’ learning cycle (Bybee et al., 2006).

Model of active-direct instruction. By contrast, in our direct instruction model the teacher first presents, explains and illustrates the concepts and principles up front, as known finished products. This is accompanied by discussion. Practical activities follow, but their main function is to test and replicate known theory using specified steps. Thereafter students apply the science to problems. We call this instructional model ‘direct-active’ because following instructor exposition it includes student practical activities. Components of this direct model are: Presentation and explanation, Testing/Replication, Application. This sequence may be viewed as a learning cycle (Figure 3) analogous to that for guided inquiry, but the term has not usually been applied in the direct case. Note that stage 2 is most often called ‘verification’ in instruction but we prefer to cast it as testing/replication, i.e., in terms better reflecting an aspect of scientific practice.

Discussion of models. For any topic, inquiry starts with learner exploration of a phenomenon, which reflects ‘science-in-the making.’ By contrast, in direct instruction the concepts and principles are presented directly to the student as ‘ready-made-science.’ Each aims at good understanding of the science content, but inquiry also aims to reflect how scientific endeavors develop that content. Some surmise that this aspect may help enhance concept learning, compared to direct instruction. Direct instruction on the other hand is widely regarded as being more efficient in time and/or coverage. Note that questions and discussion occur in both models, as do lab activities. For inquiry, lab activities are mostly investigatory, while direct, lab activities are mostly verification of theory. There is an ‘application’ phase in both cycles; where students apply the concepts to new problems. However, in the direct approach, the instructor will ‘go over’ a final-product problem solution. By contrast, an inquiry-based approach to problem-solving instruction first encourages students to consider the nature of the problem and ways of approaching it.

It will be noted that our active-direct model comprises not only instructor presentation, but also discussion, practical work and problems. It is not merely transmission and passive reception of knowledge, something one might call ‘direct passive.’ Note that the distinction between the models used is at root epistemological, concerning the nature of knowledge and of learning, as enacted in teaching and learning practice.
The Composite Nature of Lessons and the Distinction between Inquiry and Direct

All lessons are composites, made up of various different elements over a teaching session or topic unit, whether the instruction is inquiry, direct or anything else, and each element of the composite contributes to the ultimate effects of the lesson. A consequence is that in reality no lesson or lesson series is instructionally ‘pure,’ i.e., homogeneous in all aspects with regard to intended method. No lesson can be purely inquiry or purely direct throughout and still be considered generally effective instruction. Trying to treat every aspect inductively would not work, nor would unbroken didactic presentation.

Given this common sense but sometimes unacknowledged reality about lessons, the relevant question becomes: what then are the essential features distinguishing inquiry and direct methods? We believe that the essential distinction resides in “how students come to the concept.” Do they come to it through exploration and concept invention, i.e. by a process of minds-on inquiry, or are they told it up front, i.e. receiving it as a finished product presented and explained by the instructor? Once the ‘active agent’ which distinguishes modes is made clear, then other lesson ingredients not relevant to an inquiry/direct distinction can be common to both modes. A simple example might be a teacher demonstrating and explaining how to use equipment, if this is not the intended focus of an objective for concept learning. For both research and teaching purposes, the active agent is thus identified and treated appropriately. The assessment then focuses on the active agent, the specific concepts which have been treated differently.

Any research that hopes to minimize threats to validity must then deliberately and specifically construct lessons using constituent parts in such a way that only the ‘active agent’ is changed across the lesson types. Existing “stock” lessons will not work for this purpose. For research purposes it is important to be able to identify which distinguishing features of the instructional situation are at work in producing an effect. Yet many existing studies have designs which fail in this respect. A recent high-profile paper by Deslauriers, Schelew, and Wieman (2011) on improved learning in physics classes, was critiqued as follows: “the study was not controlled enough to tell which of the changes in teaching might have accounted for the difference in students’ scores” (Carey, 2011).

Design and Development of Instructional Units in Inquiry and Direct Modes

For our instructional units we chose two science topics with substantial conceptual demand, where students are known to have difficulties and where ‘alternative conceptions’ are common.

The two units developed were:

A. *It’s Dynamic!* *The relation between force and motion.* A dynamics unit involving concepts of force, net force, motion, mass, and their interrelationship in Newton’s first and second laws of motion. Restricted to straight-line situations for our project purposes.

B. *It’s Illuminating!* *Sunlight, geometry, climate and seasons.* A unit on temperature variations on Earth due to radiant energy from the sun and the earth-sun geometry involved. This comprises a foundation of basic science (light energy dependencies on angle, distance and time), followed by application to temperatures on Earth, viz. temperature variation by location/latitude (climate) and by time of year (seasons), and treated from both ground-based and space-based perspectives.

The design of the units in contrasting modes was guided by the ideas described above. The intended learning outcomes were consistent with the National Science Education Standards. Instructional units were designed as two-week modules involving an hour of classroom time each day. The units were produced in inquiry and direct versions, carefully designed in parallel to ensure
equivalence in science content, equipment/materials, and approximate teaching time, while differing in pedagogical and epistemic approach and hence in route to the science concepts as experienced by learners. For each unit, in each instructional mode, we produced booklets of student learning materials/worksheets. Detailed teacher guides were written, comprising the student materials accompanied by suggested teaching narratives for each lesson, differing markedly between modes obviously. We viewed the detailed narratives as supporting teacher professional development by clarifying the meaning of inquiry or direct instruction in specific topic contexts, i.e. supporting topic-specific pedagogical content knowledge for both content and instructional mode. The narratives were not intended to be used in the classroom as teaching ‘scripts’, since although this might provide apparent fidelity to mode, over-scripted lessons would be wooden and artificial, whereas to be effective teachers should ‘make them their own’ and feel natural in dynamic learning/teaching situations.

The complete instructional units, including domain- and topic-specific learning objectives, student materials/worksheets, teacher guides, equipment lists, embedded formative assessments, and summative assessment instruments can be accessed online.

**The Nature and Development of the Assessments**

With the assessment, as with the instruction, we aim to be specific as to its nature and characteristics, and to make examples available. Assessments need to be of quality since they provide the basis of our student performance data. Furthermore having the assessments well aligned with both learning objectives and instruction is critical for the validity of the study.

In designing assessments our premise was that mastery of science fundamentals is best indicated by the ability to *use* one’s understanding to answer questions and solve problems. Items thus embody our criterion for conceptual understanding, the ability to *apply* the concepts in new situations (Anderson & Krathwohl, 2001). All items are at Bloom’s taxonomy levels 2 and 3 (comprehension and application). Items involve conceptual understanding of the underlying ideas rather than being formula-based. The assessment instruments were sets of up to 23 conceptual multiple-choice questions, each with four choice options, together with a three-level indicator of confidence. Although objective selected-response questions have limitations, the fairly demanding conceptual nature of the questions and of the response options should be clear from the examples provided.

Two examples of assessment items for each topic are given in Figure 3-A below, to give a concrete idea of the nature of the science assessments as well as the type and level of conceptual understanding that is the desired outcome of either mode of instruction.
Items and instruments were field-tested and refined during the pilot year. To ensure balance, objectivity and content validity, independent external evaluators and experts in subject matter and assessment checked the items. Based on experience analyzing data and results from the first two trials, for research purposes we modified the assessments for the next trials by eliminating items for which pre-scores had been very high, indicating these items were either too easy or familiar to students from previous instruction, so they would reduce the power of the study to detect mode differences for current instruction. The complete assessment instruments for each unit, along with detailed learning objectives, may be viewed as Supplementary Materials S5 and S6 in the online version of this paper.

**Figure 3-A. Examples of assessment items from the two units**

Items and instruments were field-tested and refined during the pilot year. To ensure balance, objectivity and content validity, independent external evaluators and experts in subject matter and assessment checked the items. Based on experience analyzing data and results from the first two trials, for research purposes we modified the assessments for the next trials by eliminating items for which pre-scores had been very high, indicating these items were either too easy or familiar to students from previous instruction, so they would reduce the power of the study to detect mode differences for current instruction. The complete assessment instruments for each unit, along with detailed learning objectives, may be viewed as Supplementary Materials S5 and S6 in the online version of this paper.
The assessment was the same for both instructional modes, and the same tests were administered pre-and post-instruction. The teachers implementing the lessons were blind to the assessment questions, to minimize the possibility of ‘teaching to the test.’

**Teachers, Mode Assignments and Preparation**

**Teacher recruitment.** We recruited experienced middle school teachers who were generally familiar with the topics and with both modes of instruction. Experienced teachers were able to concentrate on fidelity to unit and instructional mode rather than worrying how to handle classrooms and students.

**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 would have introduced a confounding factor for some teachers but not others, involving switching natural style and thus affecting instructional quality, at least initially. Note however that in the second two years of trials the teachers switched modes, with sufficient time to make the switch and practice it. Teachers and researchers worked together in multiple sessions in the pilot year to ensure that taught lessons were as close as possible to the intended curriculum and instructional mode. Toward that same end, we produced extensive teacher guides to go along with each lesson in each mode.

**Teacher mode switch.** In comparing modes, we needed to control for possible teacher effects, therefore teachers switched modes for trials 3 and 4. The crossover also provided information on the nature and magnitude of teacher effects, although a limitation is the small number of teachers (five) in the current study. An alternative design, having each teacher teach in both modes the same day, was attractive in theory for study design, but we rejected it due to the reasonable concern that this would increase the challenge of maintaining fidelity to method. This decision was later validated by the opinions of our teachers when after two years they switched modes. They were more comfortable handling the mode switch after having longer experience with the units.

**Setting, Subjects and Implementation**

We organized an annual summer science program called “Way2Go” for middle school students about to enter 8th grade. School districts sent out advance program announcements to parents and student participation was a family decision. The subjects in this research were incoming 8th graders from several urban, suburban, and rural school districts in the Midwestern USA. Over a hundred participated each year in classroom trials during five summers, from 2006 to 2010 including the pilot year. We used an 8-day format, Mondays through Thursdays, 9 AM till noon, both topics (two lessons) each day, over the last two weeks of June. Running our own program enabled us to randomly assign students to treatment groups, which is difficult to do in a regular school setting. It also enabled us to teach as we planned and to control lesson time more easily. Our instructors were veteran middle school science teachers and their personal judgement was that the composition of students attending the trials was not noticeably different from that in their regular middle school classes, with regard to academic ability, interest, and behavior.

A voluntary summer program, however, despite having certain advantages, does have limitations and drawbacks. Students are there without the external motivator or incentive of grades. Moreover, in a voluntary summer program it is unrealistic to assign homework and reading assignments. Hence, learning gains achieved are solely dependent upon in-class student engagement with the lesson activities and discussions. While interesting activities comprising both hands-on and minds-on features were built into the lesson designs and pedagogy, observers noted that minds-on engagement was often less evident than hands-on for some of the students in this summer program.
We elected to work with 8th grade middle school students because the middle school years are transitional between elementary school and high school, and as such are critical to success at high school. The middle school years also begin the formal study of science. We worked with the chosen physical science topics because they are widely found in middle school curricula and the content forms modules amenable to our two-week format. We choose to develop two instructional units, different in nature, because it should not be assumed that instructional pedagogies and outcomes are not influenced by content topic areas.

**Teaching Fidelity**

Since one cannot simply assume that lessons are implemented as intended, we needed to judge fidelity to mode and fidelity to topic. Fidelity to inquiry or direct mode is clearly a critical feature for being able to compare their relative effects. As described earlier, we used the rule of 'prepare and verify,' preparing and practicing during the pilot year and verifying by observation during teaching. Teacher fidelity was monitored in three ways. First, by independent observers: the project contracted with 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 & Barley, 1999). Second, teachers posted journal notes each day on how the teaching went, how the students responded, and where they may have deviated from intended lessons. Third, all lessons were videotaped and could be reviewed. We reviewed videos both as a way to monitor fidelity and for teaching development purposes. The teachers had access to the videos.

Reasonable fidelity expectations for teachers must take into account the flexibility inherent in good teaching. Hypothetically one could obtain complete ‘fidelity’ by having teachers read and enact scripted lessons in each mode. Apparent fidelity might be high, but teaching would be wooden and boring and effectiveness low. Good teaching involves interacting with students and shaping things dynamically as the lesson proceeds, with a considerable degree of personalization. The reality is that all classrooms have variability, due to students, teachers, and events; this is the natural classroom variation that exists in all real teaching situations. Our operational interpretation of sufficient fidelity was that experienced 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. Observers visited two instructional days per unit for each teacher, 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; thus, in subsequent sessions they had the appropriate teacher notebooks and could score teachers specifically on fidelity to both mode and intended lessons. Qualitative and quantitative fidelity findings are included below with the study results.

**Results and Discussion**

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

**Teaching Fidelity and Quality**

As noted above, lessons were observed by independent evaluators from SAMPI. Each teacher was observed four times each trial year, twice by one observer and twice by another. Observers documented the instruction, identified what mode they were observing, and scored fidelity. Final documentation of instruction and fidelity scores were arrived at by consensus during a group meeting
of the observers after the last observations were completed. The fidelity scores are a mean of the four observer scores.

For the first two trials in 2007 and 2008, all five teachers met the fidelity standard. In the 2009 trial, the first where teachers switched approaches, one direct instruction and one inquiry instruction teacher fell below fidelity standards, and hence the data from their classrooms were not included in the mode comparison analyses below. Preparation for the 2010 trial included having teachers review the videos of their teaching. Working in the group of five teachers, these two teachers were able to identify their difficulties. As a result, in the subsequent 2010 trial, all five teachers met the fidelity standard.

Our teacher fidelity-to-mode median rating of 6 on the 7 point scale is arguably adequate for our research purposes, while remaining realistic with respect to inevitable variation in actual science classrooms. Fidelity scores were somewhat higher for direct instruction than inquiry, which is not unexpected since direct is easier to ‘execute’ than guided inquiry. The independent evaluation results gave us confidence that the difference between instructional modes was sufficiently clear in both the construction and implementation of lessons.

**Pre- and Post-Test Data and Score Distributions**

As noted earlier, quantitative data on student conceptual understanding of each topic unit was obtained by administering sets of 21-23 multiple choice items. Items tested student ability to apply the science concepts to new situations. There was a range of difficulty; some items were relatively easy or resembled cases seen in instruction, while some were challenging and novel. Students indicated their degree of confidence in each answer for each item. 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 data spread we acquired, Figure 4 shows the distribution of pre- and post-test score data for the Light unit over all five teachers and all four years. Normal curves are also fitted, but note that post-test scores may not be normally distributed due to the score ceiling; the curves are simply to depict the width and mean for each distribution, and the shift of mean (gain).

Average pre-test scores on the
multiple choice assessment instruments were around 50%, with standard deviations around 20%. Standard deviations on post-tests were similar to those on the pre-tests. In every category we analyzed there were statistically significant though modest gains. Several pre-tests with scores of 100% were discarded (no gain possible), and there were 15 students with perfect post-test scores in the Light unit. Rare anomalous cases where a post-score was below 60% of the pre-score were omitted as outliers likely arising from a student’s motivational challenges in the summer setting. ‘Pure guessing’ on test items with four choice options would average around 25%, which would contribute to score spread. However, guessing throughout was rare, and our three-level confidence indicator for each item supports this. Individual item pre-scores varied considerably between items, and some were so high it was likely that students already knew that particular aspect from previous instruction. Such items gave students little room for gain from the research unit lessons; therefore, for research purposes, halfway through the trials we chose to disregard certain Dynamics items. The adjustment resulted in lower pre-test averages and tighter distributions. On the other hand, items with low pre-scores may reflect common student conceptions, or simply involve a less familiar situation.

Gains and Normalized Gains

From the pre- and post-test data we calculated performance gains (post-score minus pre-score). These were expressed as both raw and normalized gain scores. The latter is defined as the ratio of actual gain to maximum possible gain from a given pre-score. It has become fairly common practice to look at this measure of pre/post improvement, as a way to take into account different pre-test starting scores (initial familiarity offers less potential gain). The defining equation for normalized gain is: \( g = \frac{\text{post-score} - \text{pre-score}}{100\% - \text{pre-score}} \). Normalized gains are thus ratios between 0 and 1, with 1 being the maximum achievable. To minimize distortions that can occur if a pre-score is fairly high and the gain is negative, we used the concept of normalized change as needed in subsequent calculations (Marx & Cummings, 2007). Normalized change is the gain or loss over the maximum possible gain or loss respectively, expressed as a percentage.

Effect sizes (Cohen’s d) for overall raw percentage gain over four trials/years were .71 for the Dynamics unit and .81 for the Light unit. Correlations (Pearson) between pre-scores and post-scores were typically positive and significant. Raw gains showed consistent (negative) correlations with pre-test scores, but normalized gains did not, indicating that normalization was effective, and that pre-scores would not have served as a valid predictor of ability.

Mean normalized gains were just over 0.2 for the Dynamics unit and over 0.3 for the Light unit, for both instructional treatments. These are of the same order as typical normalized gains from pre- to post-testing on the well-known Force Concept Inventory (FCI), which ranged from about 0.2 for traditional courses to about 0.35 for courses involving more active engagement (Hake, 1998). Note that our assessment items are conceptually demanding, like those on the FCI, involving not just knowledge recall but application of the concepts to cases. Project data shows that students get higher gains on factual knowledge, but that is not the learning objective for our research.

Gains were modest for reasons that could be identified from the classroom observations, performance data, unit and test characteristics, and project conditions (summer program). Note in general that if average gains are relatively modest (less than a standard deviation) then differences in such gains between instructional modes will necessarily be even more modest, and may not reach statistical significance in a ‘real world’ context such as this, given the score spreads obtained and classroom variations observed.
Student scores on the pre-tests indicate 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 around 20-25 students.

**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 five teachers, the two instructional modes, and the four trial years. To address our main research question comparing modes of instruction, we aggregated the data into Direct and Inquiry instructional modes. 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 .05. Given that randomization was at the student level, the student was taken as the unit of analysis, which allows analysis with respect to student characteristics. The performance data collected also allows us in principle to analyze at the detailed level of individual assessment items, though we have not yet carried out such extensive analysis.

Within both the Dynamics unit and the Light unit, data from trials 1 and 2 (teachers using the same mode both years) are shown aggregated, as are the data from trials 3 and 4 (teachers using the other mode both years). We justified this on the basis that within teacher/class variance between these pairs of trials was no more than expected by chance. Therefore, all four charts and tables below are grouped by pairs of trial years within each unit topic. Each figure shows results for the five teachers and for the two instructional modes, Direct and Inquiry (summarized in the larger central columns). Class average pre-scores, post-scores, and raw gains are displayed graphically as bar charts with error bars (95% confidence interval), and also in tabular form below (along with normalized change) as numerical means with standard deviations.

**Results of Similar Trials for the Dynamics Unit**

Results for the Dynamics (force and motion) unit are shown in the charts and tables in Figure 5 (trials 1 and 2) and Figure 6 (trials 3 and 4, after teachers switched between Direct and Inquiry). All results and conclusions regarding instructional mode are based on the classes/teachers who met the fidelity standard, therefore 2009 data for Ann and Sam are not included.
Figure 5. Dynamics results - trials 1 and 2.
In the first pair of trials of the Dynamics unit (Figure 5), there were no statistically significant differences between teachers or between modes on raw gain or normalized gain/change. The second two trials in Dynamics (Figure 6) yielded statistically significant differences for normalized change between Liz and Ann \( (t(59) = 2.311, p = .024; \text{effect size } d = .68) \), between Liz and Joe \( (t(91) = 2.375, p = .020; d = .50) \), and between Liz and Tom both teaching in Direct mode \( (t(91) = 2.081, p = .040; d = .44) \). There was a smaller but statistically significant difference between Inquiry and Direct on normalized change \( (t(175) = 2.010, p = .046; d = .32) \), but not on raw gain.

**Results of Similar Trials for the Light Unit**

Results for the Light (climate and seasons) unit are shown in the charts and tables in Figure 7 (trials 1 and 2) and Figure 8 (trials 3 and 4, after teachers switched between Direct and Inquiry). All results and conclusions regarding instructional mode are based on the classes/teachers who met the fidelity standard, therefore 2009 data for Ann and Sam are not included.
### Figure 7. Light results – trials 1 and 2.

<table>
<thead>
<tr>
<th>Means</th>
<th>Ann (N=40)</th>
<th>Joe (N=32)</th>
<th>DIRECT (N=72)</th>
<th>INQUIRY (N=107)</th>
<th>Liz (N=36)</th>
<th>Sam (N=36)</th>
<th>Tom (N=35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-score</td>
<td>59.4 (17.5)</td>
<td>50.1 (16.5)</td>
<td>55.3 (17.8)</td>
<td>53.8 (20.1)</td>
<td>55.2 (18.3)</td>
<td>53.0 (23.0)</td>
<td>53.2 (19.3)</td>
</tr>
<tr>
<td>post-score</td>
<td>70.5 (16.2)</td>
<td>65.5 (19.9)</td>
<td>68.2 (17.4)</td>
<td>68.3 (21.9)</td>
<td>67.6 (22.2)</td>
<td>65.5 (22.2)</td>
<td>71.8 (20.5)</td>
</tr>
<tr>
<td>% gain</td>
<td>11.0 (13.8)</td>
<td>16.3 (13.2)</td>
<td>12.9 (13.0)</td>
<td>14.4 (16.1)</td>
<td>12.4 (15.3)</td>
<td>12.5 (15.9)</td>
<td>18.6 (16.6)</td>
</tr>
<tr>
<td>normalized change</td>
<td>29.9 (29.3)</td>
<td>32.5 (29.3)</td>
<td>31.1 (29.9)</td>
<td>35.7 (35.2)</td>
<td>30.0 (35.4)</td>
<td>33.6 (36.7)</td>
<td>43.6 (32.9)</td>
</tr>
</tbody>
</table>
In the first pair of trials of the Light unit (Figure 7), the only statistically significant difference found was between teachers Ann and Tom on raw gain ($t(73) = 2.132$, $p = .036$; $d = .61$), though not on normalized change ($t(73) = 1.857$, $p = .067$). For the second pair of Light trials (Figure 8), 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 ($t(65) = 2.683$, $p = .009$; $d = .73$), Joe ($t(92) = 3.030$, $p = .003$; $d = .63$), and another Direct teacher, Sam ($t(67) = 2.692$, $p = .009$; $d = .71$). We found a smaller but statistically significant difference between Inquiry and Direct on normalized change ($t(156.6*) = 2.692$, $p = .008$; $d = .40$), but not on raw gain. (*equal variances not assumed, Levene’s test)

**Figure 8.** Light results – trials 3 and 4.
Results Across Unit Topics

Differences between results for the Dynamics and Light units on pre-score, post-score, normalized change ($t(781) = 7.359, p < .001; d = .52$), and raw gain (mean difference 3.1, standard error of difference 1.1) ($t(805) = 2.976, p = .003; d = .21$) were all statistically significant, but not greatly. There was no reason to expect them to be the same, since the units were on different content with different assessments, but it indicates the relative consistency within the student sample pool regarding the challenges of these science topics. Dynamics proved more difficult, particularly after replacement of assessment items which had the highest pre-scores or were least discriminating.

(*equal variances not assumed, Levene’s test)

Results Across Teachers

From Figures 5 through 8 one can get a sense of the variation in results not only by mode but also by teacher. Observationally, in the classrooms, ‘natural teacher variations’ in personal teaching styles and practices were clearly evident, as one might expect. As can be noted above, differences in gains between modes were often smaller than differences in gains between teachers, even within mode. Sometimes both the highest and lowest gain scores per trial were within the same mode (see Figure 8, Direct mode). Another potentially important observation is that across teachers, neither mode proved consistently more effective. For a graphical depiction of this, Figures 9 and 10 convey broad overviews of Dynamics and Light results by teacher.

The bars in Figure 9 show the mean overall raw gains in the Dynamics unit for each teacher. The average gain score across all teachers is represented by a dashed line traced across the graph. Within each teacher’s column, you can see their average gain scores marked for each of the four trials/years, with an indication of whether it was taught in Direct mode (solid circle) or in Inquiry mode (empty square), along with 95% confidence interval error bars. The solid ovals represent the average result for that teacher’s two Direct trials, while empty rectangles represent the average result for their two Inquiry trials. The two classes in 2009 which did not meet the threshold for fidelity-to-mode are indicated by an X.

Over all four years combined, within the Direct

Figure 9. Dynamics results – across teachers
mode for Dynamics, Tom’s gain scores were higher by a statistically significant amount than Joe’s ($t(78) = 1.998, p = .049; d = .46$) and also than Liz’s ($t(91) = 2.344, p = .021; d = .49$). Within the Inquiry mode for Dynamics, over all four years, Joe’s gain scores were higher by a statistically significant amount than Liz’s ($t(70.1^*) = 2.344, p = .021; d = .49$), and Ann’s were also higher than Tom’s ($t(50) = 2.215, p = .031; d = .68$). (The 2009 data points for Ann and Sam did not contribute to comparisons within intended instructional modes.)

The bars in Figure 10 show the mean overall raw gains for each teacher in the Light unit. Again, the average gain score across all teachers is represented by a dashed line traced across the graph. Within each teacher’s column, you can see their average gain scores marked for each of four trials/years, with indication of whether it was taught in Direct mode (solid circle) or in Inquiry mode (empty square). Solid ovals represent the average result for a teacher’s two Direct trials, while empty rectangles represent the average result for a teacher’s two Inquiry trials. Again, the two classes in 2009 which did not meet fidelity thresholds are marked by an X.

Over all four years combined, within the Direct mode for Light, there were statistically significant differences in gain scores between teachers Sam and Liz ($t(67) = 2.599, p = .011; d = .68$), and Sam and Ann ($t(60) = 2.982, p = .004; d = .80$). Within the Inquiry mode for Light, over all four years, Liz’s gain scores were lower by a statistically significant amount than Joe’s ($t(61.0^*) = 2.045, p = .045; d = .48$), and Ann’s ($t(54) = 2.179, p = .034; d = .62$). Ann’s Inquiry Light gains were also significantly higher than Sam’s ($t(54) = 2.102, p = .040; d = .60$). (The 2009 data points for Ann and Sam did not contribute to comparisons within intended instructional modes.)

Looking across teachers over all four years, with Direct and Inquiry instructional modes combined, and the Light and Dynamics units also combined, there are statistically significant differences in percentage gains between Liz and Ann ($t(318) = 2.541, p = .012; d = .29$), Liz and Joe ($t(320) = 3.144, p = .002; d = .35$), and Liz and Tom ($t(317.4^*) = 2.726, p = .007; d = .30$). (*equal variances not assumed, Levene’s test)
Results across Inquiry versus Direct instructional modes

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

Comparison within units. Looking at only the Dynamics unit over all four trial years, the differences between Inquiry and Direct modes in raw gain and normalized change were not statistically significant. Over all four trials of the Dynamics unit, Tom’s Direct mode scores were higher to a statistically significant degree than his Inquiry scores ($t(82) = 2.238$, $p = .028; d = .50$). Within the Light unit, the difference in raw gain between modes was not statistically significant, however, the difference in normalized change was somewhat ($t(361) = 2.143$, $p = .033$) (mean difference 7.2, standard error of difference 3.4, effect size $d = .23$).

Comparison overall. Combining the results of both the Dynamics and Light units over all four trial years (Table 2), using only teachers/classes meeting our fidelity-to-method threshold, the raw gain difference between Inquiry and Direct modes was not statistically significant (mean difference 1.1, standard error of difference 1.1, effect size $d = .07$). The difference in normalized change between Inquiry and Direct was statistically significant ($t(715) = 2.167$, $p = .031$), but with a fairly small effect size (mean difference 4.9, standard error of difference 2.3, effect size $d = .16$).

Table 2
Comparison of Inquiry versus Direct methods overall (Dynamics and Light units combined)

<table>
<thead>
<tr>
<th>mean gain/increase (SD)</th>
<th>Ann</th>
<th>Joe</th>
<th>Liz</th>
<th>Sam</th>
<th>Tom</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>11.4</td>
<td>12.7</td>
<td>11.0</td>
<td>17.8</td>
<td>16.9</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td>(12.9)</td>
<td>(13.7)</td>
<td>(14.6)</td>
<td>(15.9)</td>
<td>(17.0)</td>
<td>(15.1)</td>
</tr>
<tr>
<td>N=79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inquiry</td>
<td>21.2</td>
<td>18.2</td>
<td>10.9</td>
<td>11.6</td>
<td>13.8</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>(17.5)</td>
<td>(15.5)</td>
<td>(12.5)</td>
<td>(14.7)</td>
<td>(16.7)</td>
<td>(15.6)</td>
</tr>
<tr>
<td>N=36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>14.4</td>
<td>16.0</td>
<td>11.0</td>
<td>13.9</td>
<td>15.6</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>(15.1)</td>
<td>(15.0)</td>
<td>(13.7)</td>
<td>(15.4)</td>
<td>(16.9)</td>
<td>(15.4)</td>
</tr>
<tr>
<td>N=115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Over all four trials, Dynamics and Light combined, Joe’s Inquiry mode scores were higher than his Direct mode scores to a statistically significant degree ($t(157) = 2.297$, $p = .023; d = .37$), as were Ann’s (excluding her 2009 data) ($t(53.0^*) = 3.020$, $p = .004; d = .69$).

Over all four trials, combining both units and all teachers meeting fidelity standards, there was a statistically significant difference in normalized change between Inquiry and Direct, but the effect size of .16 was quite small relative to several differences between teachers within modes, and between modes within teachers (above). The small difference was not of practical significance, given the teacher variation. Over all four years/trials, combining both units, and all teachers (meeting fidelity standards), there was not a statistically significant difference between Inquiry and Direct modes of instruction with regard to gain/increase in percentage correct from pre-score to post-score on...
conceptual assessments (effect size $d = .07$). (For a graphical depiction, Figures S1 and S2 convey broad overviews of Dynamics and Light results by mode/method.)

### Results Across Different Trial Years

This factor can be viewed as a replication in four successive years with different student subjects, and with instructors switching modes after two trials. Results were similar across the four trials, with average performance data seeming to increase slightly with year, but this was not statistically significant, and there were no statistically significant differences in average normalized gain/change between years. Thus this replication data could be aggregated in 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) ($t(44)=2.370, p = .022; d = .72$). (For a graphical depiction of overall results, Figures S3 and S4 convey broad overviews of Dynamics and Light results by year.)

### Lesson Time Comparisons Between Modes and Between Teachers

Overall, direct lessons took about 10 minutes less per nominal 1-hour session than inquiry lessons, although this difference varied according to the particular lesson involved and the particular teacher. Time variations between teachers were at least as great as time variations between instructional modes.

### Conclusions

The results from our experimental study comparing specified models of guided-inquiry and active-direct instruction, implemented in realistic classroom environments in a two-week program, are that inquiry and direct modes led to comparable science conceptual understanding in roughly equal instructional times. While students made statistically significant learning gains in both units, the wide range of scores on pre-and post-tests (reflected in large standard deviations of the data) suggest that both the gains and gain differences would need to be larger than those observed in order to show statistical or practical classroom significance. Question pre-scores would need to be lower, with less spread. A larger-scale study would provide larger N-size, but at the cost of precision, since in practical terms it becomes much more difficult to prepare, control and monitor all the instructional and classroom situations and factors, thus increasing variation further. Following Cronbach (1975), a number of separate local studies in various environments would be more informative, to see whether and how the findings generalize to other situations and to refine and study the effect of various parameters. In any case, gain differences between instructional modes were thus far not found to be of statistical or practical significance compared to the observed natural variation of students, teachers and classrooms. It is of interest to consider possible reasons why this might be so, viewed from a number of perspectives.

First, we used soundly-designed units based on acceptable models of good instruction for both modes. Second, we are not comparing active with passive learning. Active-learning activities occur in both modes, though their characters differ. Third, we tried to make the lessons interesting regardless of mode. Fourth, and importantly, there is an application phase in both models, during which students apply the concepts they have just learned to a variety of cases and problems, which further enhances learning and conceptual understanding. The application phase in instruction should, to some extent, tend to even out differences in initial concept learning. The fact that all lessons are composites, with some aspects common, is not an issue for the research results since only the active agent aspect which differs between methods is assessed.
Even though the *approach* to learning a new concept is very different in the two modes, learning processes are not neat and linear. Essentially, students need to construct their own conceptual schemata however the instruction is organized. Thus, whether students initially ‘find out’ or are ‘told’ about science concepts and laws, various knowledge elements and connections need to be revisited several times while making sense of the concepts and making them their own. Furthermore, even if curriculum and instruction is true to intended mode, students’ various (tacit) epistemologies about science and about learning will affect how that instruction is perceived and processed. It will naturally take a while for students to adjust their conceptions of what science is all about, and their own approaches to learning, in response to the nature of new instruction. Differences in concept learning due to inquiry or direct instruction might not be as evident initially. Therefore, for a combination of reasons, differences in initial concept learning via one mode or the other, even if significant, may be evened out considerably by the various factors in play as learning proceeds naturally.

On the other hand, more than science content knowledge is always conveyed in lessons. Inquiry-based instruction may promote appreciation of scientific inquiry, at least those aspects of it directly involved in the lessons students experience. Most science educators feel that inquiry instruction, by its very nature, provides crucial added value, in having students ‘do’ science for themselves. Affective factors also play a role in learning, even if the prime focus is on science concepts. It may be that interest is sparked more by an inquiry approach, promoting positive attitudes toward science, which could lead to better performance. In direct instruction, lesson structure and purpose may be clearer and preferred by some learners, another affective factor. Direct instruction may be easier from the *teaching* point of view, particularly for less experienced teachers or those not yet confident with the content. There is also merit to the time argument in favor of direct, but our study shows that the time differential is not as great as usually claimed, if both modes include experiential and application aspects, and if inquiry is focused and well guided. Finally, if students develop a concept themselves rather than ‘receive’ it directly, transfer of knowledge to new situations may conceivably be enhanced and longer-term retention improved, since they should be able to reconstruct that concept. However we reiterate that our study focused explicitly on science conceptual understanding, measured by the ability to apply the concepts to conceptual problems. Thus the interesting issues noted above become, of course, further research questions.

Returning again to our main research question, and given the composite nature of all lessons and the realities of implementation in classrooms, we see that some common claims for the superiority of either direct and inquiry instruction in regard to concept acquisition may be viewed as somewhat overstated. Our study shows that good direct and inquiry instruction led to similar understanding of science concepts and principles in comparable times. It may well be that under more tightly controlled and rehearsed conditions one could better distinguish the performance differential due to mode of instruction, which would be of significant theoretical interest; but this study gives a practical indication of what is likely to happen in the field under less clinical conditions. Thus, our findings show that the promotion of one mode of instruction over the other, where both are based on sound models of expert instruction, should not be based on content acquisition considerations alone.

Inquiry-based instruction clearly offers significant potential advantages for science education, by modeling scientific inquiry during concept learning. Such concomitant benefits would need to be studied in research specifically designed to assess understanding of scientific inquiry, preferably by devising case-based items aligned with the instruction, just as we assessed concept understanding. However, as far as science conceptual development is concerned, our conclusion is that expertly designed instructional units, sound active-engagement lessons, and good teaching are as important as whether lessons are cast as inquiry or direct.
Acknowledgement

Funded by 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 National Science Foundation.

Supplementary Documents

The following supplementary materials are included as an appendix to this article:

- Further charts of results: S1, S2, S3 and S4
- Learning objectives and complete assessment instruments: S5 and S6
REFERENCES


http://template.aea267.iowapages.org/lessonplan/index.html


and Methods (Science - version B: for multiple classroom observations to improve programming). Kalamazoo, MI: Science and Mathematics Program Improvement (SAMPI).


APPENDIX
Supplementary Documents

- Further charts of results: S1, S2, S3 and S4
- Learning objectives and complete assessment instruments: S5 and S6
Figure S1. Dynamics results – across modes/methods.
Figure S2. Light results – across modes/methods.
Figure S3. Dynamics results – across years.
Figure S4. Light results – across years.
Supplementary document S5:

It’s Dynamic! The relation between motion and force

Unit description, objectives and assessments

Unit description

This unit focuses on the relation between motion and force. It deals with two central ideas, viz. 1. If there is no net force on an object then its motion does not change, i.e. it keeps going steadily in a straight line or remains at rest, and 2. A net force on an object causes it to change its motion, i.e. to speed up, slow down or change direction.

These two important ideas are not intrinsically difficult, especially in simple cases, but students often have conceptual difficulties with them, since the ideas may sometimes seem at odds with the observed behavior of objects in a world with friction.

Although our main goal is the relation between motion and force, in order to do this we also need to consider the concept of force, and of net force. However we do not intend a detailed study of various types of forces. Due to time constraints we limit the unit to straight-line motions and do not include changes of direction.

Thus the unit starts with force and net force, goes on to motion with no net force (‘unforced’ motion), and then to the effect of force on motion (‘forced’ motion). The main sections are as follows:

Forces

- Force is viewed simply as a ‘push or pull’ at this level.
- Forces can act by contact or at a distance.
- How forces combine: the idea of resultant or net force.

Motion with no force (Newton’s first law)

- If there is no net force on an object it continues moving at steady speed in a straight line, or stays at rest.

How do forces affect motion? (Newton’s second law)

- If there is a net force on an object it changes its motion (speeds up or slows down).
- A greater force causes a more rapid change in velocity.
- The effect is less if the object’s mass is greater.

Real situations and friction, consolidation and application

- Friction and air resistance play a role in many real situations.
- Consolidation of unit, applying concepts to various real situations and problems, which may include resistive forces.

Broad standards statements relevant to this unit

Standards documents use broad terms for stating benchmarks and objectives. Thus, a main broad objective for this unit is that middle school students will “Be able to explain the motion of objects due to forces acting.” Equivalent formulations are found in a number of standards documents, e.g. “Describe how things around us move and explain why things move as they do,” and “Relate motion of objects to forces.” Or slightly more specifically: “Relate changes in speed or direction to forces” or “Describe how forces (pushes and pulls) speed up, slow down or change the direction of an object.”

All of the above and similar statements about force and motion can be found in the documents listed below, with page or section numbers specified.

Detailed unit-specific objectives

Detailed objectives specific to this instructional unit are given below. Objectives are written both to guide teaching and with the construction of assessment in mind.

1.0 Forces

1.1 Be able to state that a force is a push or pull by one object on another at that moment.
1.2 State that some forces act by contact, others at a distance, and give examples.
1.3 Be able to measure forces in units of Newton, using spring scales.
1.4 Be able to add forces in the same direction and subtract forces in opposite directions, to get the net force; and to state that if the forces are equal and opposite they give no net force.
1.5 Be able to represent forces on an object in force diagrams.

2.0 Motion where there is no net force

2.1 Given a situation where no force (or no net force) acts on an object, be able to state or predict that it will continue its motion in a straight line or remain at rest.
2.2 Vice versa, given a situation of an object moving steadily in a straight line, infer that there is no net force on the object.
2.3 For a moving object, given distance travelled and time taken, be able to calculate speed, including units.
2.4 On a diagram showing the positions of an object at equal time intervals, be able to explain whether the object is traveling at a steady speed or speeding up or slowing down.

3.0 Motion where a net force acts

3.1 Be able to describe the effect of a net force on motion, i.e. that the object will speed up or slow down, depending on the force direction. Vice versa, given situations where an object is speeding up or slowing down, be able to infer that there is a net force and identify its direction.
3.2 Be able to state that a larger force causes speed to change more rapidly, and apply in particular cases.
3.3 Be able to state that a larger mass changes its motion less (rapidly) in response to a force, and apply in particular cases.

4.0 Real situations, friction, and unit consolidation

4.1 Be able to state that friction and air resistance acting on moving objects affect their motion, and apply to examples.
4.2 Be able to explain that if frictional forces are exactly equal to the applied force on an object, it will move at constant speed (and vice versa), and apply to examples.
4.3 Be able to identify the forces acting in various situations which include friction.
4.4 Be able to apply concepts relating force and motion to real world contexts.

The objectives and assessments are well aligned.
Student Post-Assessment

IT’S DYNAMIC!

Please read each item carefully. Then circle the correct answer to the question (A, B, C, or D).

For each answer, also circle whether you are—
A. Very sure, B. Somewhat sure, or C. My best guess.

1a. Amy pushes a wagon containing some packages along a level floor. The wagon has such good wheels that friction is very small.

If Amy pushes continuously all the time with a constant pushing force, and friction can be ignored, then the wagon will . . .

A. not move.
B. move with a constant speed.
C. speed up steadily.
D. slow down.

1b. How sure are you of your answer? A. Very sure  B. Somewhat sure  C. My best guess

2a. Amy suddenly stops pushing when the wagon is moving quite fast.

What will happen next? (Remember that friction can be ignored). The wagon will . . .

A. stop immediately.
B. continue moving at the speed it was going.
C. speed up.
D. slow down quickly, then stop.

2b. How sure are you of your answer? A. Very sure  B. Somewhat sure  C. My best guess
3a. Amy’s sister joins her and both push together on the loaded wagon, each pushing just as hard as Amy did before.

![Diagram of two people pushing a wagon]

Compared to when Amy was pushing alone, the wagon will now . . .

A. travel at the same steady speed.
B. travel at a faster steady speed.
C. speed up just as before.
D. speed up quicker than before.

3b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

4a. Amy now unloads the packages from the wagon and then pushes it with the same constant force as she used before.

![Diagram of Amy pushing an empty wagon]

The empty wagon will now . . .

A. move just the same as when it was loaded.
B. move at a steady speed, but faster than before.
C. move at a steady speed, but slower than before.
D. speed up, more rapidly than before.

4b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

5a. You are pushing a sled through the snow by applying a horizontal force with a magnitude of 70 newtons. There is a force of friction between the sled and the snow with a magnitude of 20 newtons.

What is the resulting (net) force on the sled?

A. 20 newtons
B. 50 newtons
C. 70 newtons
D. 90 newtons

5b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess
6a. We see an object moving straight past our window at a steady speed. What can we say about the possible forces acting on the object?

A. There are no forces at all acting on the object.
B. There may be forces on the object, but if so they are equal and opposite (balanced).
C. There is a net (unbalanced) force on the object in the direction it is moving.
D. There is only one force on the object, in the direction it is moving.

6b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

7a. A stone falls to the ground. Any air resistance can be ignored. As the stone falls, what is true of its speed and the force, if any, acting on it?

A. Its speed is constant and there is a constant force on it.
B. Its speed is constant and the force on it is zero.
C. Its speed increases and the force on it is constant.
D. Its speed increases and the force on it is increasing.

7b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

8a. Tom is pushing on a crate with a constant force of 90 newtons, across a level floor. The floor creates a friction force on the box of 40 newtons. What is the size of the net (unbalanced) force on the box causing it to speed up?

A. 50 newtons
B. 65 newtons
C. 90 newtons
D. 130 newtons

8b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess
9a. After the box in question 8a has gained some speed, Tom reduces his applied force to 40 newtons (which is the same as the friction force). What will be the motion of the box when he is pushing with 40 newtons?

A. It will stop immediately.
B. It will still speed up, but not as quickly as before.
C. It will move at a steady speed.
D. It will slow down.

9b. How sure are you of your answer? A. Very sure  B. Somewhat sure   C. My best guess

10a. A car is traveling at a steady 70 mph along a straight level highway, not speeding up or slowing down.

What is true of the forward and backward force(s) acting on the car in this situation?

A. There are no forward or backward forces acting on the car.
B. There is a forward force and a backward force, which are exactly equal and opposite (balanced).
C. There is a forward force and a backward force, but the forward force is bigger.
D. There is only a forward force on the car.

10b. How sure are you of your answer? A. Very sure  B. Somewhat sure   C. My best guess

11a. A passenger is in a car traveling at a steady speed along a straight level road. What is true of the forward and backward force(s), if any, on the seated passenger? (Note: this asks about forces on the passenger, not on the car).

A. There are no forward or backward forces on the passenger.
B. There is a forward force on the passenger, but no backward force.
C. There are forward and backward forces on the passenger, but the forward force is bigger.
D. There are forward and backward forces on the passenger, exactly equal and opposite.

11b. How sure are you of your answer? A. Very sure  B. Somewhat sure   C. My best guess
12a. A rocket is out in space where there is no friction at all to affect its motion. You can also ignore any effect of gravity. At first the rocket engine is operating continuously as shown in the top diagram, and the rocket is speeding up.

Then the engine suddenly shuts off.

After the engine shuts off, what is true of the forces acting on the rocket and how it moves?

A. The force remains as before, and the rocket keeps speeding up.
B. The force remains as before, and the rocket travels at a constant speed.
C. The force becomes zero, and the rocket travels at constant speed.
D. The force becomes zero, and the rocket stops.

12b. How sure are you of your answer? A. Very sure     B. Somewhat sure     C. My best guess

13a. A heavy box full of toys weighing 100 newtons is on rough ground. Andy starts pulling on the box to the right with a force of 60 newtons, but then Betty and Cathy start pulling to the left with forces of 50 newtons and 30 newtons. To everyone's surprise, when all three are pulling as shown, the box does not move one way or the other.

We can conclude that the size of the friction force on the box is . . .

A. 20 newtons.
B. 40 newtons.
C. 100 newtons.
D. 140 newtons.

13b. How sure are you of your answer? A. Very sure     B. Somewhat sure     C. My best guess

14a. The direction of the friction force in question 13a above is . . .

A. down.
B. up.
C. to the left.
D. to the right.

14b. How sure are you of your answer? A. Very sure     B. Somewhat sure     C. My best guess
15a. A round boulder is placed loose in the back of a pickup truck. The driver starts the truck and begins speeding up. As the truck starts moving forward, the driver hears a noise, and looking back notices that the boulder has hit the back (tailgate) of the pickup. The driver wonders about the reason for this.

Sally and Sam give different explanations, as follows.

1. Sally says that the boulder tended to ‘stay put’ as the truck started moving, so it did not move forward with the truck. It only appeared to move backward in the truck because the truck moved forward.

2. Sam says that since the boulder hit the back of the truck, there must have been a backward force acting on it, causing it to do so.

What is the best comment about these two statements?

A. Only Sally’s statement (#1) is correct.
B. Only Sam’s statement (#2) is correct.
C. Both Sally and Sam have made correct statements.
D. Neither Sally’s nor Sam’s statements are correct.

15b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

16a. 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.

16b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess
17a. A ball is thrown upward as shown, and as it moves upwards its speed gets less. Air resistance is small and can be ignored. When the ball is on the way up as shown (at position 1), what forces, if any, are acting on it?

A. No forces.
B. Only an upward force.
C. Only a downward force.
D. An upward and a downward force, the upward force being larger.

17b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess

18a. A stone is being swung around in a circle at the end of a string. The diagram shows a top view of the path of the stone. The string suddenly breaks when the moving stone is in the position shown.

Which numbered path best shows the motion of the stone after the string breaks?

A. Path 1  
B. Path 2  
C. Path 3  
D. Path 4

18b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess
19a. 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, which diminishes 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

19b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

20a. A box is at rest on the floor. You apply a horizontal force to it by pushing from the side, but it does not move. What can be said about the friction force, if any, acting on the box while you are pushing?

A. There is no friction force.
B. The friction force is less than the force you apply.
C. The friction force is equal to the force you apply.
D. The friction force is greater than the force you apply.

20b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

21a. There are plates, cups and other dishes placed on a tablecloth lying on a dining table. You hold the edge of the tablecloth with both hands and pull it toward you with a quick jerk. The tablecloth comes out and the dishes stay pretty much where they were! The reason the dishes don’t get pulled along with the tablecloth and fall off the table is that while the cloth is being pulled . . .

A. there is no friction between tablecloth and dishes, so the dishes will experience no horizontal force and remain where they are.
B. a force due to friction with the tablecloth does act on the dishes, but it is insignificant and cannot affect the dishes.
C. a force due to friction with the tablecloth does act on the dishes, but acts for a very short time if the cloth is pulled out quickly, so the dishes have little time to move.
D. no force is being applied directly by your hands to the dishes, so there is no reason for the dishes to move.

21b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess
22a. Our hero notices that a wheeled cart is rolling by itself toward a baby on the level sidewalk. She rushes in and pushes backward on the cart as hard as she can. What will happen?

A. She will have no effect if the cart is a lot bigger than she is.
B. The cart will stop immediately, since there is nothing pushing it forward.
C. The cart keeps going forward even though she is pushing backward, but it will slow down.
D. The cart will instantly go backward instead of forward.

22b. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

23a. Students are investigating the motion of a loaded cart and find that it speeds up when they push it.

They suspect that how quickly the cart speeds up might depend on two things, namely: the size of the force they apply to the cart, and the mass of the cart. They decide to investigate this, but two groups use different methods, as follows.

<table>
<thead>
<tr>
<th>First Method (Group 1)</th>
<th>Second Method (Group 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 first keeps the mass of the cart constant, and applies three different forces, to find the effect of force size. Next they keep the force constant and apply it to the cart loaded with three different masses, to find the effect of mass.</td>
<td>Group 2 varies both the force and the mass, from one trial to the next, to get six different combinations of force and mass. Then they look at their data for all six trials, seeking possible patterns that show the effect of both force size and of mass.</td>
</tr>
</tbody>
</table>

Which of these two methods is better for finding out how speeding up depends on force size and on mass?

A. The first method is better.
B. The second method is better.
C. Both methods are equally good.
D. Neither method is good.

23b. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess
It’s Illuminating! Sunlight and Earth’s Temperature Variations

Unit description, objectives and assessments

Unit description
This unit includes both basic science and its applications to the real world. The science is about light energy received on a surface and how it varies with angle, distance and time, while the applications are the effects of sunlight on temperatures on earth, explaining why temperatures vary with location (latitude) and with time of year (seasons). The basic and applied components of the unit are as follows:

1. **Basic science**
   - Light carries energy, and can heat a surface.
   - Energy received depends on the angle of light to a surface, distance from the source, and duration.

2. **Applications**
   - Temperatures on earth vary with location (latitude). (Due to different angles of incoming sunlight at different latitudes).
   - Temperatures vary with time of year. (Due to the tilt of the earth’s axis, which gives rise to different angles and durations of sunlight in summer and winter).

*Ground-based and space-based viewpoints*

We approach each application from two points of view. First from the *ground-based* viewpoint, observing the path of the sun across the sky, and how its angle (and duration) varies with location and with time of year. Second, we view the earth-sun system as if ‘from space,’ involving the geometry of a spherical earth spinning on a tilted axis while orbiting the sun. Note that from the first point of view we accept the observed behavior of the sun at different locations and times of the year, but do not know why it behaves this way. Here the second point of view provides an explanation, in terms of the geometry of the sun-earth system.

*Broad domain objectives for this unit*

Standards documents use a broad brush for stating the main objectives for a domain. Broadly stated, our domain objectives for this unit are as follows.

*Students will be able to state that light carries energy, and say qualitatively how the energy received on a surface depends on angle, distance from the source, and duration. Using these principles and geometry, be able to explain why temperatures on earth depend on location (latitude) and on time of year (seasons), from both a ground-based perspective and a view from space.*

*Detailed unit-specific objectives*

Detailed objectives specific to this unit are given below. They are written to guide teaching, learning and assessment. Page or section numbers given are from standards documents.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Standards documents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic science</strong>&lt;br<em>Learners should be able to do the following:</em></td>
<td></td>
</tr>
<tr>
<td>1. State that light carries energy, and describe simple ways to show this.</td>
<td>STAT 133, U 13</td>
</tr>
<tr>
<td>2. State how light energy received on a surface depends qualitatively on angle, on distance and on duration. Design and describe simple ways to demonstrate each dependency.</td>
<td>AAAS 66, STAT 133, NAEP 37 E.11, Table 19 p45</td>
</tr>
<tr>
<td>3. Explain why the intensity of surface illumination depends on angle, and compare cases. (By considering area illuminated by a beam).</td>
<td>AAAS 66, STAT 133</td>
</tr>
<tr>
<td>4. Explain why intensity depends on distance for a diverging light beam. (By considering area illuminated by a beam).</td>
<td></td>
</tr>
</tbody>
</table>
### Application to the earth

*Learners should be able to do the following:*

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>Explain that virtually all energy on the earth’s surface comes from the sun and that sunlight warms the earth’s surface.</td>
<td>NAEP E8.11, Gd 4 &amp; 8, STAT U 13, NSES 159, 161, 189, AAAS 83, 85,</td>
</tr>
<tr>
<td>6.</td>
<td>Describe how earth’s temperature varies with location (latitude).</td>
<td>STAT 126, 133</td>
</tr>
<tr>
<td>7.</td>
<td>Describe and compare the apparent (observed) motion of the sun in the sky from different locations on earth.</td>
<td>AAAS 68, 336, STAT 29, U 18, NAEP 33, NSES 134, 159,</td>
</tr>
<tr>
<td>8.</td>
<td>Explain why average temperatures vary with latitude on earth. Do this from both ground-based and space-based viewpoints. (In terms of different angles of sunlight to the ground).</td>
<td>STAT 126, U 19</td>
</tr>
<tr>
<td>9.</td>
<td>Describe how earth’s temperature varies with time of year (seasons).</td>
<td>STAT 126, 133</td>
</tr>
<tr>
<td>10.</td>
<td>Compare the apparent (observed) motion of the sun in the sky at different times of year (midsummer and midwinter).</td>
<td>STAT U 18, 20, 29, NAEP 33, NSES K-4 134, 159.</td>
</tr>
<tr>
<td>11.</td>
<td>Explain why temperatures on earth vary with time of year (seasons). Do this from both a ground-based and space-based viewpoints.</td>
<td>AAAS 66, 68, 69, STAT 29, 39, 131, 134, NSES 161</td>
</tr>
<tr>
<td>12.</td>
<td>Explain that distance from the sun is <em>not</em> a factor in causing earth’s seasons.</td>
<td>MS8 17, AAAS 217, 270</td>
</tr>
<tr>
<td>13.</td>
<td>Show how basic principles of science and math apply in real world situations.</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Use science and geometry to predict the effects on earth if certain factors were different (e.g. if earth’s axis were not tilted or if orbital distance varied significantly).</td>
<td>STAT 15, AAAS 217, 270, MS8 17</td>
</tr>
</tbody>
</table>

### Related mathematical objectives

*Learners should be able to do the following:*

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15.</td>
<td>Determine area by counting small squares within a given outline.</td>
<td>AAAS 223</td>
</tr>
<tr>
<td>16.</td>
<td>Represent physical systems and behavior by geometrical diagrams and sketches.</td>
<td>AAAS 268</td>
</tr>
<tr>
<td>17.</td>
<td>Relate a change in one quantity to a change in another.</td>
<td>AAAS 217</td>
</tr>
</tbody>
</table>

### Standards references

- **AAAS** | AAAS Benchmarks for Science Literacy.
- **NSES** | National Science Education Standards.
- **NSES-I** | Inquiry and the National Science Education Standards.
- **STAT** | State Essential Goals and Objectives for Science Education 1991.
- **STAT-U** | Update to STAT in 2000.
- **NAEP** | National Assessment of Educational Progress for 2009.

The objectives and assessments are well aligned.
1a. In Figure 1 below, a beam of light strikes nearly straight on to a surface and lights up an area on the surface.

Figure 1. Striking nearly straight on

Figure 2. Striking more slanted

In Figure 2 above, the direction of the light beam is more slanted to the surface. What will happen to the area lighted by the beam?

A. It will be smaller.
B. It will be the same.
C. It will be larger.
D. One cannot say from this information alone.

1b. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

1c. When the beam of light strikes at more of a slant (as in Figure 2 above) what will happen to the intensity (brightness) of the lighted surface area?

A. It will be less.
B. It will be the same.
C. It will be greater.
D. It is not possible to say.

1d. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess
2a. Some students notice that a sheet of metal gets warm when it is placed near a lighted household light bulb. They wonder what affects how much energy the metal sheet receives from the light, and suggest three possibilities:
   I. The distance of the sheet from the light bulb.
   II. The angle of the sheet to the light.
   III. The length of time the sheet is in the light.
Which of these are factors affecting the energy the sheet receives?
   A. I only
   B. I and II only
   C. I and III only
   D. I, II and III

2b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

3a. The diagram to the right shows the earth with the equator and the rotation axis marked. Temperatures on earth are hotter near the equator and colder toward the poles. The reason for the differences in temperature is that . . .
   A. the sun’s light beams are more straight on to the ground at the equator and more slanted toward the poles.
   B. regions near the equator are much closer to the sun than regions near the poles.
   C. the length of daylight is always greater near the equator than the poles.
   D. the earth’s axis is tilted.

3b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess

4a. Tom turns on a flashlight in his bedroom and shines it on the wall 2 feet away to produce a small circle of light. He then shines the same flashlight on his ceiling 6 feet away to produce a larger circle of light.
Does more light reach the ceiling or the wall?
   A. More light reaches the ceiling.
   B. More light reaches the wall.
   C. **Equal amounts of light reach the ceiling and the wall.**
   D. The amounts cannot be compared.

4b. How sure are you of your answer?  A. Very sure     B. Somewhat sure     C. My best guess
5a. At a particular city at noon one day, the sun’s rays come in *directly downward* onto the ground as shown in Figure 1 below.

![Figure 1](image1.png)

Figure 2 below shows a different view, of the earth as a globe, with light rays coming from the sun.

![Figure 2](image2.png)

Three locations are marked on the earth as A, B and C. Where on the globe must the city be located in order to have the sun’s rays come in to the ground as shown in Figure 1?

A. Location A only.
B. **Location C only.**
C. Locations A and C only.
D. All locations A, B and C.

5b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess

5c. Looking again at the locations in Figure 2 above, which location would have the highest average temperature?

A. Location A only.
B. **Location C only.**
C. Locations A and C only.
D. Locations A, B and C would have the same temperature.

5d. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess

6a. Light from a flashlight strikes a surface, as shown in the picture below.

![Figure 3](image3.png)

The intensity of light (brightness) on the surface depends . . .

A. only on the distance from the light source.
B. only on the angle of the surface.
C. **both on the distance and the angle.**
D. neither on the distance nor the angle.

6b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess
7a. The diagram below shows the earth in relation to the sun at a particular time of the year. The earth’s axis, equator and the locations of Chicago in North America and Santiago in South America are shown.

![Diagram of Earth's axis and location of Chicago and Santiago]

In the diagram above, what season is it in Chicago?

A. Summer  
B. Fall  
C. Winter  
D. Spring

7b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess

7c. In the diagram above, what season is it in Santiago?

A. Summer  
B. Fall  
C. Winter  
D. Spring

7d. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess

8a. The earth travels around the sun in a nearly circular orbit, while spinning on an axis which is tilted. Students wonder what differences there might be if the earth’s rotational axis was NOT tilted, but was perpendicular to the orbit. They come up with the following ideas.

If the earth’s axis was **NOT** tilted, then possibly. . .  
I. temperature would not vary with latitude.  
II. there would be no seasons.  
III. there would be no night and day.

Which one or more of these suggested effects are correct?

A. I only  
B. II only  
C. I and II only  
D. I, II and III

8b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess
9a. The diagram below shows the earth in relationship to the light coming from the sun at a particular time of the year.

At this particular time of the year, what season(s) will be occurring in the northern and southern hemispheres?

A. Summer in the northern hemisphere and Summer in the southern hemisphere  
B. Summer in the northern hemisphere and Winter in the southern hemisphere  
C. Winter in the northern hemisphere and Winter in the southern hemisphere  
D. Winter in the northern hemisphere and Summer in the southern hemisphere

9b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess

10a. Imagine that instead of being nearly circular, the earth’s orbit was very oval shaped, with the sun toward one end as shown in the picture below. Then as the earth orbits around the sun, its distance from the sun would vary. Also imagine that the earth’s axis has no tilt (as in the picture below).

With this oval orbit and no tilt, which one of the following statements would be true?

A. Temperatures on the 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.

10b. How sure are you of your answer?  
A. Very sure  
B. Somewhat sure  
C. My best guess
11a. A plastic globe and a flashlight can be used as a “model” to represent the earth-sun system. We can move the globe and flashlight around to show how sunlight strikes the earth and explain how the seasons occur. However, the distances and sizes in this model are NOT the same as in the real earth-sun situation. Should a scientist use a model like this, or not?
   
   A. No, because models should include ALL aspects of the real thing as accurately as possible.
   B. Yes, models can be a way of showing the important aspects of the real thing and may ignore other aspects.
   C. No, a model of this kind should not be used at all because it can be misleading.
   D. No, because models must be made to an accurate scale.

11b. How sure are you of your answer?  A. Very sure B. Somewhat sure C. My best guess

12a. In Boston, it is colder in January than in July. The reason is that in January . . .
   
   A. the earth is further from the sun.
   B. the earth is nearer to the sun.
   C. the sun is lower in the sky during the day.
   D. the sun is higher in the sky during the day.

12b. How sure are you of your answer?  A. Very sure B. Somewhat sure C. My best guess

13a. The sun has two planets, X and Y. One planet is larger and further away from the sun, as shown. Both planets receive light from the sun.

How will the temperatures on the two planets compare, with each other?
   
   A. Planet X will be a lot warmer.
   B. Planet Y will be a lot warmer.
   C. Planets X and Y will have similar temperatures.
   D. Their temperatures have nothing to do with the sun, so we cannot say.

13b. How sure are you of your answer?  A. Very sure B. Somewhat sure C. My best guess
14a. In Chicago it is colder in January than in July. People suggest various possible reasons for this difference, as follows:

I. The sun’s rays strike the ground at more of a slant in January.
II. Daytime is shorter in January.
III. The earth is farther away from the sun in January.

Which one or more of the above are correct reasons for the temperature difference?

A. I and II only
B. III only
C. II and III only
D. I, II and III

14b. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

15a. The diagram shows a perspective view of the earth orbiting the sun, with four positions marked with numbers. Arrows show the direction of movement. The north-south (NS) line through the earth shows the “tilt” of the earth’s axis.

In the drawing above, the time taken by the earth to travel from position 1 to position 3 is . . .

A. 1 month. B. 3 months. C. 6 months. D. 1 year.

15b. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

15c. In the drawing above, when the earth is at position 1, then in the northern hemisphere it will be . . .


15d. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

15e. In the drawing above, when the earth is at position 1, then in the southern hemisphere it will be . . .


15f. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess
16a. The light beam from a flashlight strikes a card and illuminates it. Simon and Sarah both think that the brightness on the card might depend on two things: 1) the *distance* of the card from the flashlight and 2) the *angle* that they hold the card.

![Diagram of light beam]

They do experiments to find out. They each make six measurements, but use different methods.

<table>
<thead>
<tr>
<th>Simon’s Method</th>
<th>Sarah’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simon tests six different combinations of distance and angle, changing both the</td>
<td>Sarah first fixes the card at one particular angle and tries three different</td>
</tr>
<tr>
<td>distance and the angle for each measurement. He then looks for patterns in the</td>
<td>distances to find the effect of distance. Then she fixes the card at one</td>
</tr>
<tr>
<td>data that would show the effects of distance and angle on brightness.</td>
<td>particular distance and tries three different angles to find the effect of</td>
</tr>
<tr>
<td></td>
<td>angle.</td>
</tr>
</tbody>
</table>

Which of these methods is better for finding out how brightness depends on distance and on angle?

A. Simon’s method is better.
B. **Sarah’s method is better.**
C. Both methods are equally good.
D. Neither method is good.

16b. How sure are you of your answer?  
A. Very sure  B. Somewhat sure  C. My best guess