Experimental Comparison of Inquiry and Direct Instruction in Science[1]

William W. Cobern, David Schuster, Betty Adams, Brooks Applegate and Brandy Skjold
Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI 49008, USA

Adriana Undreiu
University of Virginia's College at Wise, Wise, VA 24293, USA

Cathleen C. Loving
Department of Teaching, Learning & Culture, Texas A&M University, College Station, TX 77843, USA

Janice D. Gobert
Department of Social Sciences & Policy Studies and Computer Science, Worcester Polytechnic Institute, Worcester, MA, USA

Abstract

There are continuing educational and political debates about 'inquiry' versus 'direct' teaching of science. Traditional science instruction has been largely direct, but recent national and state science education standards advocate inquiry throughout K-12 education. While inquiry-based instruction has the advantage of modeling aspects of the nature of real scientific inquiry, there is little unconfounded comparative research into the effectiveness and efficiency of the two instructional modes for developing science conceptual understanding. This research undertook a controlled experimental study comparing the efficacy of carefully designed inquiry instruction and equally carefully designed direct instruction in realistic science classroom situations at the middle school grades. The research design addressed common threats to validity. We report on the nature of the instructional units in each mode, research design, methods, classroom implementations, monitoring, assessments, analysis and project findings.

Introduction

For many decades there have been both educational and political policy debates over the merits of "inquiry-based" and "direct" approaches to teaching science, with strong opinions on both sides. In broad practice the pendulum has been mostly on the direct side, but in recent years, with the formulation of national and state science education standards, inquiry has become the *sine qua non* for science instruction (National Research Council, 2000b; American Association for the Advancement of Science, 1990; Alberts, 2008). By inquiry-based teaching of science we mean instruction that reflects the investigative approach, empirical techniques, and reliance on evidence that scientists use in making discoveries and constructing new knowledge. A related idea is that of "Scientific Teaching", Handelsman et.al. 2004, 2007) an aspect of which posits

---

[1] 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.

that ‘the teaching of science should be faithful to the true nature of science by capturing the process of discovery in the classroom.”

While learning science in school cannot be the same as doing real science, nor should it be, it can nevertheless give a sense of how science produces new knowledge, if the approach to science topics models some of the features of scientific inquiry. 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).

However note that constructivism is a theory of learning, not of instruction, and students must construct their own understanding regardless of the learning resources or their nature, be they laboratory activities, lectures, discussion, text, etc.

Convincing comparative evidence for the superiority of either direct or inquiry science instruction is hard to come by. Note that ‘experientially-based’ instruction and ‘active student engagement’ are certainly advantageous for effective science learning [Reference needed], but that such “hands-on” and “minds-on” (National Research Council, 2000a) aspects can occur in both inquiry and direct science instruction. Thus, the pertinent question we seek to address is not whether active, experiential learning of science is more effective than passive, non-experiential learning. Our research question is whether an inquiry approach or a direct approach to experientially-based instruction is more effective for science concept development, when both approaches are expertly designed and well executed.

Inquiry Science Teaching

Inquiry teaching of science aims to reflect the nature of scientific inquiry in the classroom, and it can be approached in many ways, as noted by Alberts (2008). He lists several goals of science education: to know, use, and interpret scientific explanations of the natural world; to generate and evaluate scientific evidence and explanations; to understand the nature and development of scientific knowledge; and to participate productively in scientific practices and discourse (Alberts, 2009). Brady (2008, p. 607) notes that most science educators claim that we ‘know how to teach science,’ meaning to teach science by inquiry, and that the only question remaining appears to be ‘when will we do it?’ Handelsman et. al (2004) state that “many scientists are still unaware of the data and analyses that demonstrate the effectiveness of active learning techniques.” (p. 521). To the extent that the authors include a pedagogy of inquiry there is a clear claim here that such a pedagogy has been soundly tested and found effective. However ‘active learning’ can mean many things, and active learning and inquiry are sometimes unwittingly conflated, but they are not the same thing.

Note that proponents of both inquiry-based and direct instruction are convinced of their respective positions regarding pedagogical efficacy. But are these claims justified by evidence? Advocacy of either approach, inquiry or direct, needs support from comparative efficacy studies.

A recent meta-analysis conducted by the Educational Development Center (2007) of research studies into inquiry instruction does not find them to yield sufficiently unconfounded inferences to adequately address the central issue. Indeed, the EDC reports that research rigor had in fact declined from 1984 to 2002. Various confounding threats to the validity of inquiry research since the 1960s include:

- There are few controlled studies which compare inquiry against worthy alternative instruction. Controls are often absent or represented by poor or nebulous “traditional” teaching.

- Few studies use randomized assignment of subjects to treatment groups or quasi-experimental efforts to control for differences between subject groups.
- Evaluations not independent of the developers and researchers can be problematic.
- Replication may not be feasible due to insufficiently detailed specification of treatments.
- Rarely is implementation compared to the intended instruction, so that tacit assumptions of fidelity may be unwarranted.

We elaborate later on these threats to validity, but turn first to the issue of using a significant control, or in our case, alternative treatment. The fair test opponent in our research is a specified model of good direct instruction.

Direct Instruction

Science education specialists invariably talk about moving teachers away from direct instruction toward inquiry. This commitment is understandable and laudable for anyone who loves both science and the teaching of science. A less commendable argument for inquiry, however, is to contrast it with straw man caricatures of alternative modes of instruction, in particular, direct instruction cast as exposition, memorization, and cookbook laboratory work. David Ausubel (1961) addressed such misrepresentation more than 40 years ago. He argued that the true issue was rote learning vis-à-vis meaningful learning, and that neither inquiry instruction nor direct instruction *automatically* lead to meaningful learning.

Since considerable educational and political support exists for various forms of direct instruction, countering the move toward inquiry in the standards, the question remains: is some expert form of direct instruction more effective than inquiry instruction? David Klahr (2002) thinks so, claiming his research findings “challenge predictions derived from the presumed superiority of discovery approaches for deeper, longer lasting, and ‘more authentic’ understanding of scientific reasoning processes…” Such views are not isolated; in 2006 Kirschner, Sweller and Clark published a paper entitled “Why Minimal Guidance During Instruction Does Not Work,” with Sweller furthering the critique in 2009. However, upon inspection, one notes that Klahr’s research comparing teaching approaches involves open “discovery,” the most unstructured form of inquiry, rather than the guided approach advocated by the NRC and AAAS, and his “direct” mode arguably involves aspects of guided inquiry. Moreover, the study is about acquiring science process skills like the control of variables strategy, rather than about science content learning. Thus, on neither count do Klahr’s findings speak to the core question regarding inquiry instruction for concept development.

Nevertheless, Klahr and Sweller draw attention to the precarious evidentiary support for inquiry instruction. Convincing evidence is similarly lacking for direct instruction, which has simply had the advantage of acceptance as the status quo. Hence, inquiry instruction vis-à-vis direct instruction has not been subjected to experimental controlled studies comparable to that of Klahr or the work referred to by Sweller.

Research Framework

Given the widespread advocacy of inquiry instruction by many educators, and of direct instruction by other stakeholders, together with the problematic evidentiary foundation, we conducted a controlled experimental study to compare experientially-based inquiry instruction and experientially-based direct instruction for science conceptual development at 8th grade level. To minimize threats to validity, we built four features onto our research: Specificity, Fidelity, Objectivity, and Transparency.

Specificity

Rather than rely on single-word descriptors like ‘direct’, ‘inquiry’, ‘traditional’ ‘hands-on’ etc, we specify exactly what we mean by each mode of instruction, by characterizing them operationally with explicit *models*. These include unit structure, components, sequencing and approach, making clear what is different and what is common between modes. Our Guided Inquiry instruction model is based on the Karplus Learning Cycle (Lawson et. al., 1989), designed to reflect important features of scientific inquiry. It involves three main phases, Exploration, Concept Formation, and Application. The Karplus cycle is also at the heart of the BSCS ‘5E’ learning cycle (Bybee et. al., 2006). The exploration and concept formation activities help students toward “inventing” the relevant scientific concepts and “discovering” the relationships and laws, guided or scaffolded by the instructor. By contrast, in our direct instruction model the teacher presents and explains the concepts, relationships and laws directly to the students, as finished products to be learned and understood. We call our direct model “direct active” since it includes hands-on practical work, though of a confirmatory nature and with prescribed steps. The direct-active cycle also has phases, namely Presentation, Explanation, Confirmation, and Application.

For our instructional units we chose two science topics with substantial conceptual demand and known to give difficulties. The units are:

- *It’s Dynamic!* Dynamics unit. The concepts of force, motion and mass, and their interrelationship in Newton’s first and second laws of motion.
- *It’s Illuminating!* Light, climate and seasons unit. A foundation of basic science (light energy dependency on angle, distance and time), followed by application to temperatures on Earth, viz. variation by location (latitude) and time of year (seasons), treated from both ground-based and space-based perspectives.

The inquiry and direct versions of each unit were designed in parallel, to ensure equivalence in science content and approximate teaching time while differing in epistemological and pedagogical approach.

All lessons are composites, comprising a number of sequenced elements, and no lesson can be 100% inquiry or 100% direct throughout and still remain generally effective instruction. Trying to treat every detail inductively would not work, nor would unbroken didactic presentation. What then is the essential aspect distinguishing inquiry and direct modes? We believe it lies in “how students come to a concept.” That is, do students develop the concepts and principles from exploration, or are they told? This represents the ‘active agent’ distinguishing inquiry from direct. Thus for example, both modes can include experiential activities but the sequencing and epistemological nature of such activities is different in the two modes. Lesson ingredients which are not relevant to an inquiry/direct distinction can reflect generic features of good instruction generally and thus be common to both modes. Demonstrations or explanations of how equipment works and how to use it would be an obvious example, if this is not the intended focus of an inquiry objective. It would detract from the real inquiry objective of a lesson if students had to take time to find out such things for themselves, while direct instruction students did not.

The complete units, including learning objectives, lesson materials and assessments are available online at www.wmich.edu/way2go/.

For our assessment, as for curriculum, we aim for specificity as to its nature and rationale. Having the assessment closely aligned with learning objectives and content is critical for the validity of the study. The assessment instruments are sets of 24 conceptual multiple-choice questions, each with four choice options, together with a three-level indicator of confidence. The questions focus on main principles and embody our criterion for conceptual understanding, viz. the ability to *apply* the concept (Anderson & Krathwohl, 2001). They involve application of principles to new situations rather than recall of factual knowledge. Examples of items for each unit are provided in the Appendix, and also serve to give an idea of the kind of conceptual understanding we are aiming at. Note that although objective selected-response questions have limitations, the conceptual nature of questions and response options should be clear from the examples, and the items were field-tested and validated.

The assessment was the same for both instructional modes, and was administered pre-and post-instruction by independent project evaluators. The teachers implementing the lessons were blind to the assessment questions, to minimize the possibility of inadvertent or deliberate "teaching to the test."

Fidelity

Teachers were allocated to the two treatment modes according to their preference, i.e. the way they felt 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 an extension to the project in 2009/10 we are giving teachers time to make the transition and practice it.

One cannot simply assume that lessons are implemented as intended. In classrooms, two important aspects of fidelity are fidelity to mode and fidelity to curriculum. Fidelity to inquiry or direct mode is clearly a critical feature for comparing their relative effects. We used a "prepare and verify" approach, practicing for fidelity during teacher preparation and evaluating it during teaching. Teacher fidelity was monitored in three ways. First, independent observers were provided by the Science and Mathematics Program Improvement (SAMPI) group (2009), specializing in evaluating science instruction. Observers were initially "blind" to teacher mode assignments, but because fidelity to mode was reasonably good, they soon identified the direct and inquiry teachers; thus in subsequent trials they had the appropriate teacher notebooks and could score teachers specifically on fidelity to the intended lessons. Second, teachers posted journal notes on each day's teaching, and third, all lessons were videotaped and could be reviewed.

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 scripted lessons in each mode. Apparent fidelity might then be high, but teaching would be wooden and effectiveness low. Good teaching involves interacting with students and shaping things dynamically as the lesson proceeds, with some 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 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. Our teacher fidelity-to-mode median rating of 86% 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.

Objectivity

Our research design embedded several areas of "blindness" to minimize bias. Teachers were blind to the assessment, and SAMPI coded and analyzed teacher performance data without knowledge of group. Independent observers initially blind to teacher assignments visited two lessons per unit for each teacher and two observers saw each teacher. The observers documented the instruction and then scored fidelity to method. Final evaluation of both instruction and fidelity was arrived at by consensus in observer meetings after the last observations. The 2008 and 2009 observation reports are available online at www.wmich.edu/way2go/.

Transparency

Transparency of research would allow others to know exactly what the project did and thus facilitate attempts to evaluate or replicate it. Yet many research reports give no more than a brief description of various aspects without access to important detail. We have made our research as transparent as possible by placing all critical information on our project web site. This contains the complete sets of unit descriptions, learning objectives, student materials, teacher guides, and assessments. With this accessible, anyone can see how the research attempts to implement theoretical ideas regarding advocated practices of inquiry and direct pedagogy, and potentially contribute constructive comment.

Subjects and Setting

In the 2007 and 2008 summer trials the subjects were 180 incoming 8th grade students from several Midwest school districts, urban, suburban and rural. Over two weeks in June, classes met in the morning for four days a week, covering one lesson of each unit each morning. Districts sent out advance program announcements to parents and hence participation was a family decision. Our instructors were veteran middle school science teachers. Their subjective judgment was that the students were not noticeably different overall from those in their regular school courses. We ran a special summer program to enable random assignment of students to treatment and control groups, which is difficult to do in a regular school setting, and this also enabled us to control time-on-task. A voluntary summer program, however, has limitations. There are no grade incentives, and to include homework or reading assignments would be unrealistic. Learning is thus essentially dependent upon in-class student engagement. While both ‘hands-on’ and ‘minds-on’ features were built strongly into the lesson designs and pedagogy, observers noted that minds-on engagement was less evident than hands-on for some of the students in this summer program.

Results and discussion

Student performance data can be analyzed and results compared in a number of ways across various factors. Data were grouped by topic, teacher/classroom, program year and instructional mode. We could thus make various comparisons, and aggregate categories of data if results were similar within statistics. Besides the main research question about inquiry vis-à-vis direct instruction, analyses provided information about randomization, pre-and post-tests, and performance gains, and we summarize these aspects first.

Randomization of students

Student scores on the pre-tests indicate that randomization of students across classrooms was effective, i.e. any variation in pre-scores between classes was consistent with that expected by chance for class sizes of around 20 students.

Time comparison

Overall, direct lessons took about 10 minutes less per nominal 1-hour session than inquiry lessons; although this varied by lesson, and variations between teachers were at least as great as variations between modes.

Assessments: pre- and post-tests

Average pre-test scores on the multiple choice assessment instruments were around 50%, with standard deviations around 20%. Individual item prescores varied considerably between items. Items with higher prescores indicated that students already knew something about aspects of these topics, which thus limits possible improvement. For items with low prescores, popular distracter choices reflected the common preconceptions, as expected. The 'pure guessing' rate for items with four choice options would be 25% if students actually chose at random and would also contribute to score spread. However, guessing throughout is rare, and our three–level confidence indicator for each item supports this. Pre- and post-test questions were identical. In every sub-category we analyzed, there were statistically significant though modest gains. Standard deviations on post-tests were similar to those on the pre-tests. Figure 1 shows representative pre- and post-test score data, for the case of the Light/Climate unit in inquiry mode. Post scores may not be normally distributed due to the score ceiling; the curve is used simply to show the mean, width and shift of the distribution compared to the pretest.

INSERT FIGURE 1 HERE

Performance gains and normalization

Performance gains between pre- and post-tests were expressed as both raw and normalized gain scores, the latter being the ratio of gain to maximum possible gain given the pre-score. For example if pre-score is 40% and post-score 55%, the raw gain is 15%, and the normalized gain is (55-40) / (100- 55) = 15/45 = 0.33 or 33%. To avoid distortions that can occur if a gain is negative, we used the concept of normalized *change* in calculations (Marx & Cummings, 2007). Normalized change is the gain or loss over the maximum possible gain or loss respectively, expressed as a percentage.

Mean normalized gains were of the order of 20% for the Dynamics unit and 30% for the Light unit, for both treatments. The effect sizes (Cohen's d) for raw gain were .69 for the Light unit and .54 for the Dynamics unit. Effect sizes for normalized gain were 0.99 for Dynamics and 1.4 for Light. Raw gains showed consistent (negative) correlation with pretest scores, but normalized gains did not, indicating that normalization was working in this regard.

These gains can be put in perspective by noting that in a large survey of courses which used the well-known Force Concept Inventory (FCI) for pre and post testing in mechanics, typical normalized gains varied from ~ 20% for traditional courses to ~ 35% for courses involving more active engagement (Hake, 1998). The normalized gains in our study are of the same order. Note that our assessment items, like those on the FCI, are conceptually demanding, involving *application* of the concepts to cases, not just knowledge recall. Project data shows that students get higher gains on factual knowledge, but that is not the objective here.

If average gain is relatively modest (say less than a standard deviation) then *differences* in such gains between instructional mode will intrinsically be smaller and may not reach statistical significance in this realistic context, given the score spreads obtained and classroom variations observed. We found this to be the case in this summer program field study, for the most part. Gains were modest for reasons that could be identified from the classroom observations, performance data, and unit and test characteristics. Some reasons are intrinsic while others provide insights for refinements to the units, instruction and implementation.

Comparative results

In the sections below we present and discuss results for science content understanding as assessed by the pre- and post-instruction tests, analyzed by topic, program year, teacher and instructional mode.

Results for the Light/Climate and Dynamics units are displayed as bar charts in Figures 2 and 3, with standard error bars overlaid, and accompanied by tabulated values for scores, gains, and standard deviations. The two charts display results by topic, teacher, and instructional mode, while aggregating data from two program years.

INSERT FIGURE 2 HERE

INSERT FIGURE 3 HERE

This project data allowed us to make comparisons for different program years, teachers, topics and instructional modes. Each of these aspects is discussed below.

Different program years. This particular factor can be viewed as a replication in successive years with different students. Results were similar across the two trials, with small mean differences, which were not statistically significant. Thus this replication data could be aggregated in studying other factors.

Different teachers. Student gains for different teachers within the same instructional mode were compared, and no statistically significant differences in normalized gain were found among either the Inquiry teachers or the Direct teachers. Note however the limits on power to detect such differences due to numbers and the broad score distributions. Observationally, “natural teacher variations” in personal teaching style and practice were clearly evident as one might expect.

Different topics. There is no a priori reason to expect that different topic units (Dynamics and Light/Climate) should necessarily have the same pre –scores, post scores or gains, except inasmuch as they are at similar conceptual levels. Rather, we are interested in comparing inquiry and direct gains *within* each topic unit. On the other hand one might possibly expect a *correlation* between student gains in two topics taught by the same teacher. We found that raw percent gain did not correlate between the Light and Dynamics unit (Pearson’s, r=.077, *p*=.308), but normalized gain/change showed a statistically significant correlation (Pearson’s, r=.238, *p*=.001).

Different instructional modes. Regarding the central research question, i.e. comparing inquiry and direct instruction, the center portions of the bar charts in Figures 2 and 3 represent these results. Findings for the summer program over two years are as follows.

For the Light unit, over two trials, the difference between direct and inquiry groups on normalized gain/change was not statistically significant (*t*(178)=.755, *p*=.451) (mean diff. 3.8, std. error diff. 5.1, effect size Cohen’s d=.12). Note that on raw gain there was a small but statistically significant difference between one direct teacher (Ann) and one Inquiry teacher (Tom) (*t*(73)=2.132, *p*=.036), but that upon normalization the difference was not statistically significant (*t*(73)=1.857, *p*=.067). Similarly, for the Dynamics unit, differences between direct and inquiry groups on normalized gain/change were not statistically significant. (*t*(178)=.717, *p*=.474) (mean diff. 3.1, std. error diff. 4.4, effect size Cohen's d=.11).

Thus with teacher data aggregated within each mode and over both years, we find that normalized gain differences between modes of instruction were small and not statistically significant, given standard deviations reflecting the wide range in scores that students obtain in pre-and post-tests.

Note that, given such score spreads, it is clear that both the gains and gain differences would need to be larger than those observed in order to show statistical or practical classroom significance. A single larger-scale study would of course provide larger N size, but at the cost of precision, since practically it becomes more difficult to prepare, control and monitor instructional and classroom situations, thus increasing variation. 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.

Insights gained

While the main research goal was to obtain the comparative results described above, a valuable additional outcome was that new insights were gained during the project into the multiple aspects involved in such an endeavor, including instructional design, assessment, materials, teacher development, classroom implementation, student learning, and research design and methodology. These have implications for both research and instruction and suggest issues for further research.

Discussion of the findings from various perspectives

It is perhaps not surprising that students should acquire comparable conceptual understanding of science subject matter via inquiry and direct modes, for a number of reasons representing various perspectives.

There are certain curriculum-related features of science teaching and learning that are relevant. Firstly, we used soundly designed units based on acceptable models of good instruction for both modes. Secondly, the lessons have certain 'generic' instructional features in common, in areas where an inquiry/direct distinction is not relevant, even if the 'active agent' component differs markedly between modes. Thirdly, hands-on experiential aspects occur in both modes, albeit approached very differently. Fourthly, there is an 'application' phase in both learning sequences, to further enhance understanding in problem-based fashion as students get practice in applying the concepts just learned, so to some extent the application phase should tend to even out differences in initial concept learning.

From a cognitive perspective, even though the approach to a new concept is very different in the two modes, learning processes are not linear but 'hyperlinked'; concept learning involves a learner putting together pieces of both new and existing knowledge. Students need to construct their own conceptual schemata no matter how the instructional sequence 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 while making sense of these concepts and laws, during assimilation and accommodation in Piaget's terms. Furthermore, even if curriculum and instruction are both true to mode, students' current (tacit) epistemologies about both science and learning will affect how instruction is received. It will 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 instruction. Therefore differences in concept learning due to inquiry or direct instruction might not be as evident initially.

Thus, for a combination of reasons, differences in initial concept learning via one mode or the other, even if significant, may be evened out considerably thereafter as learning proceeds naturally.

On the other hand, more than concept understanding is always conveyed in lessons in the two modes. Inquiry-based instruction may better promote student appreciation of scientific inquiry; in fact it would be surprising if it did not. However our study focused explicitly on science concept understanding, measured by ability to apply the concept in conceptual problems. Affective factors also play a role in learning, even if the focus is on science concepts; it may be that interest is sparked more naturally by inquiry, thus promoting positive attitudes toward science, which could lead to better performance. In addition, if students develop a concept themselves, under guidance, rather than 'receive' it directly, transfer of knowledge to new situations may be enhanced and longer-term retention may be improved. In our study many of the assessment items did involve new situations, i.e. most problems were 'unseen'. However a summer program does not easily allow longer-term follow-up.

All of the issues mentioned above are of course further research questions.

Conclusions

The results from our experimental study comparing specified models of inquiry and direct instruction, implemented in realistic classroom environments in a two-week program, are that inquiry and direct methods led to comparable science conceptual understanding in roughly equal instructional times. Gain differences between instructional modes were not statistically significant within the observed natural variation of students, teachers and classrooms.

The project studied two science topic units, and although they were very different in nature we cannot necessarily generalize our findings to all topics; the nature of a topic might conceivably play a role in what kind of instruction is effective. However, note that each unit was fairly comprehensive, involving many concepts and different types of task, as would be the case with other topics too.

Mastery of science content in the alternative modes was our central research question, but the inquiry-versus-direct debate is not just about content: it is also about the nature of science and about efficiency. Most science educators feel that inquiry instruction, by its very nature, provides crucial added value, in having students 'do' science for themselves. This gives a 'feel' for science and hence some appreciation of the nature of scientific inquiry. For direct instruction, given our finding that it does not lead to a better grasp of the basics, it is not as clear what other grounds there are on which to argue superiority. One is that direct instruction is easier from the *teaching* point of view, particularly for less experienced teachers or those not confident with the content. Another is that the cut-and-dried structure of direct instruction may be less demanding for weaker students, at least initially. There is also merit to the time argument, 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. True, direct is certainly more efficient than unguided 'open discovery', but no one is really advocating the latter. On both content and time grounds, therefore, the 'efficiency' of direct instruction is not markedly greater, and any time saving is likely to be outweighed by loss of other benefits. Direct instruction does risk sending the message that science is simply a body of knowledge to be learned.

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 effects of 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. Thus, 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 alone.

Inquiry-based instruction potentially offers significant advantages for science education, by modeling scientific inquiry during concept learning; these concomitant benefits would have to be studied in research designed for that purpose. There is a need for such research since some claims for inquiry methods regarding understanding of nature of science are not sufficiently supported by evidence. However, as far as science concept understanding is concerned, our conclusion is that expertly designed instructional units, sound active-engagement lessons, and good teaching are as important as whether a lesson is cast as inquiry or direct.

REFERENCES

Alberts, B. (2008). Considering Science Education. Science, 319(5870), 1589.

Alberts, B. (2009). Redefining Science Education. Science, 323(5913), 437.

American Association for the Advancement of Science (AAAS). (1990). Science for all Americans: Project 2061. New York: Oxford University Press.

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

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

Brady, T. E. (2008). Science Education: CASSANDRA'S PROPHECY. Phi Delta Kappan, 89(8), 605-607.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes. N. (2006). The BSCS 5E Instructional Model: Origins, Effectiveness and Applications. Colorado Springs, CO: BSCS.

Cronbach, L. J. (1975). Beyond the Two Disciplines of Scientific Psychology. American Psychologist, 30(2), 116-127.

Education Development Center, I. (2007). Inquiry-based Science Instruction and Students' Science Content Knowledge: A Research Synthesis. Paper presented at the annual meeting of the National Association for Research in Science Teaching New Orleans.

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

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

Handelsman, J., Miller, S., & Pfund, C. (2007). Scientific Teaching. NY: W. H. Freeman and Company, 1.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist, 41(1), 75-86.

Klahr, D. (2002). Paths of Learning and their consequences: Discovery Learning versus Direct Instruction in elementary school science teaching. Seminar presentation at the Pittsburgh Scientific Reasoning Supergroup 2002.

Lawson, A. E., Abraham, M. R., & Renner, J. W. (1989). A theory of instruction: Using the learning cycle to teach science concepts and thinking skills. NARST.

Lawson, A. E., Abraham, M. R., & Renner, J. W. (1989). A theory of instruction: Using the learning cycle to teach science concepts and thinking skills National Association for Research in Science Teaching.

Llewellyn, D. (2007). Inquiry Within: Implementing Inquiry-based Science Standards in Grades 3-8, Second Edition, Corwin Press.

Marx, J., & Cummings, K. (2007). Normalized change. American Journal of Physics, 75(1), 87-91.

National Research Council. (2000a). How people learn. Washington, DC: National Academy Press.

National Research Council. (2000b). Inquiry and the National Science Education Standards: A Guide for Teaching and Learning. Washington, DC: National Academy Press.

Science and Mathematics Program Improvement. [Web Page]. URL http://www.wmich.edu/sampi/ [2009, August 10].

Sweller, J. (2009). What human cognitive architecture tells us about constructivism. S. Tobias, & T. M. Duffy (editors), Constructivist Instruction: Success or Failure? (pp. 127-143). Routledge.

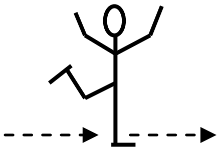
APPENDIX: ASSESSMENT EXAMPLES FOR THE TWO UNITS

Examples of objective assessment items for each topic are given below, to give an idea of the conceptual nature of the science assessment as well as the kind of understanding that is the desired outcome of either mode of instruction. The full set of items may be viewed at www.wmich.edu/way2go/ along with detailed learning objectives.

1. Example items from the Dynamics unit

These items test conceptual understanding of Newton's first and second laws in particular contexts; the response options include common alternative conceptions.

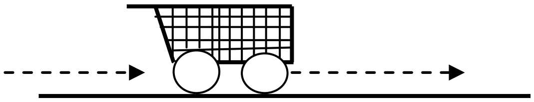
1a. Ice skater. An ice skater is skating on an indoor ice rink. She first gets up to a fast speed, then stands on one skate and keeps going steadily across the ice without any apparent effort.
The reason for her keeping moving like this without effort is that . . .



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

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

1b. Shopping cart. A shopping cart is given an initial push to get it moving on a carpeted floor. The cart then travels along the carpet on its own for a while, gradually slowing down.



Dashed line indicates the direction the cart is moving.

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

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

2. Example items from the Light, Climate and Seasons unit

These are two of the more challenging conceptual questions for this unit, involving understanding of how various factors can be a basis for temperature variations.

2a. Reasons for seasons. In Kalamazoo, Michigan, it is **colder** in January than it is in July. People suggest various possible reasons for this difference, as follows:

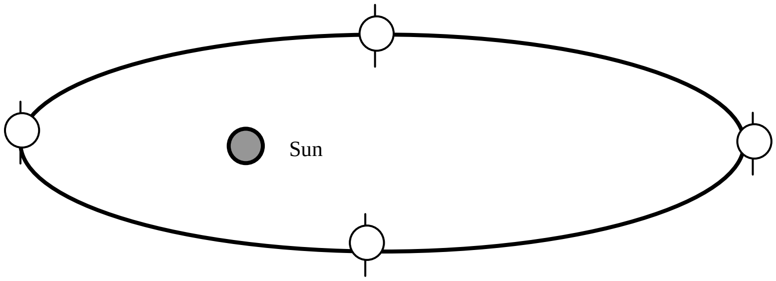
i. The sun's rays strike the ground at more of a slant in January.
ii. Daytime is shorter in January.
iii. The Earth is farther away from the Sun in January.

Which 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.

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

2b. Imaginary situation. Imagine that instead of being nearly circular, the earth's orbit was a long ellipse, with the sun nearer to one end of it than the other as shown. Then as the earth orbited, its distance from the sun would vary a lot. Also imagine that the earth's axis has *no tilt.* This is all shown in the diagram.



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

A. Temperatures on earth would not vary with time of year. i.e. there would be no seasons.

B. Temperatures would vary during the year, with seasons being opposite in the northern and southern hemispheres

C. Temperatures would vary during the year, with seasons being the same in the northern and southern hemispheres.

D. It is not possible to say without more information.

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

FIGURE 1

Fig. 1. Prescore/Postscore distribution

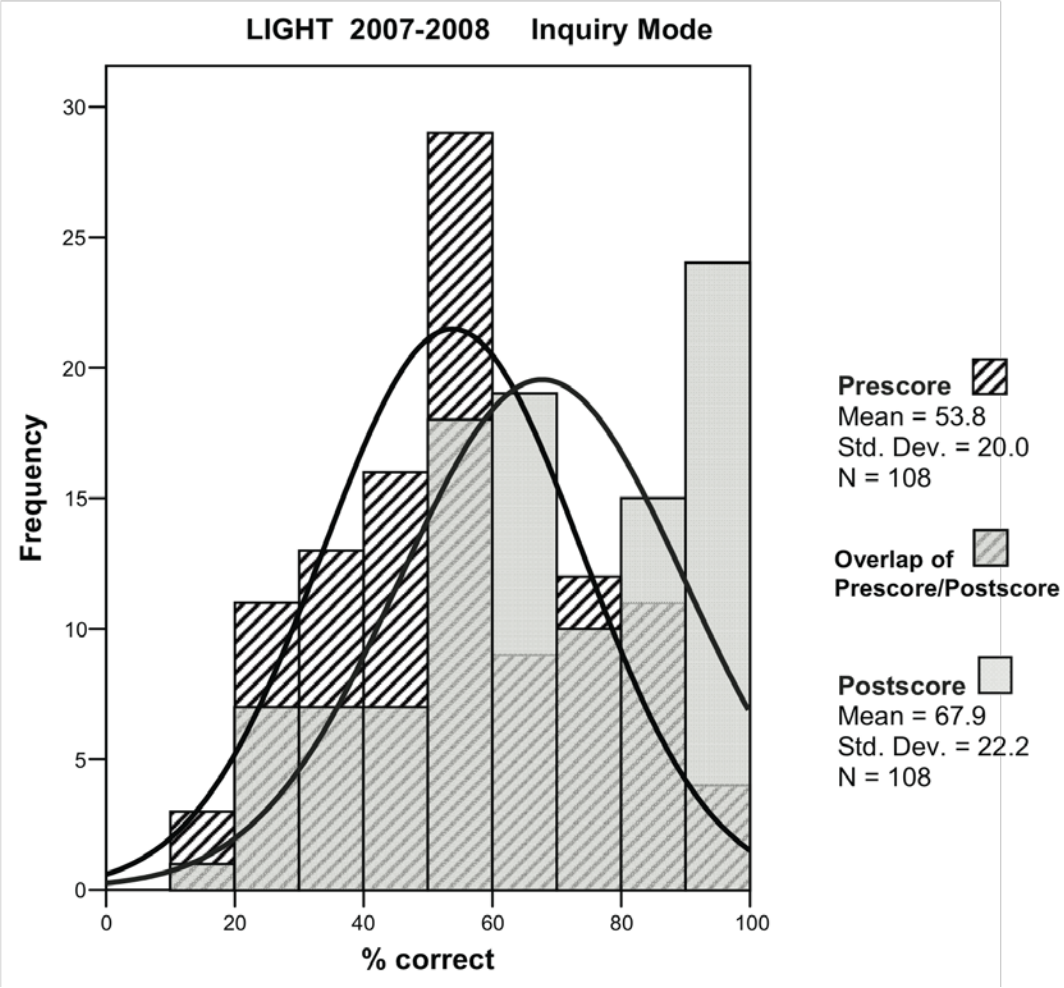
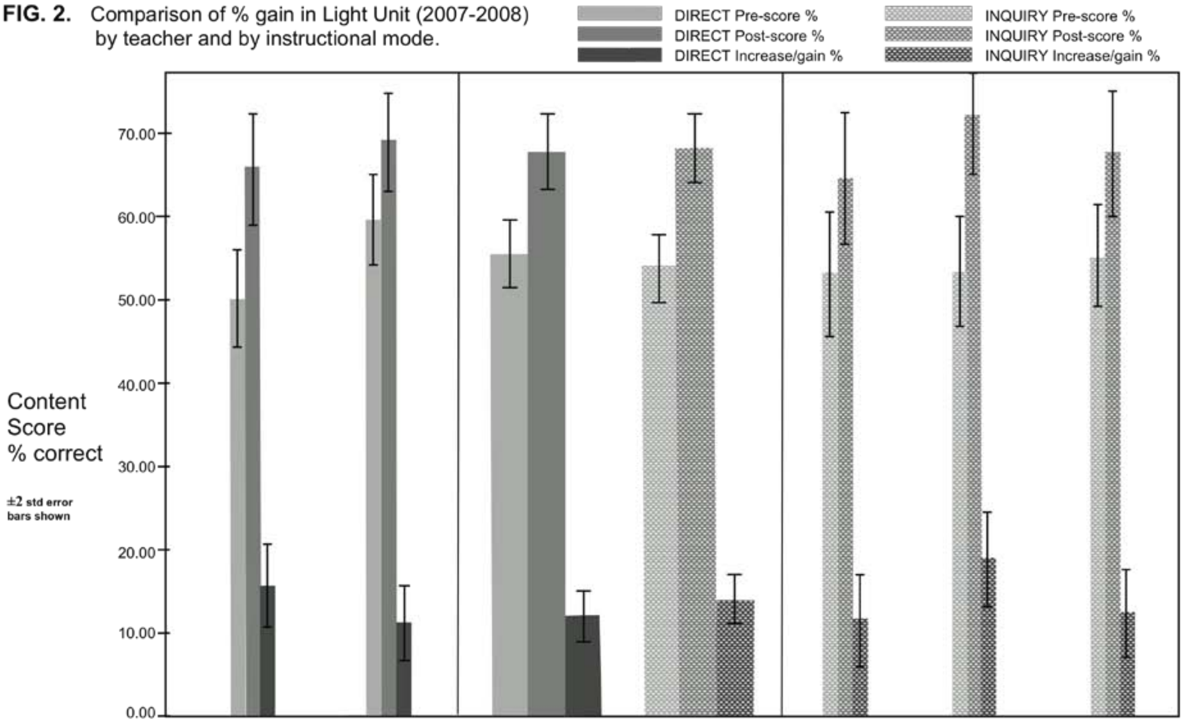


FIGURE 2

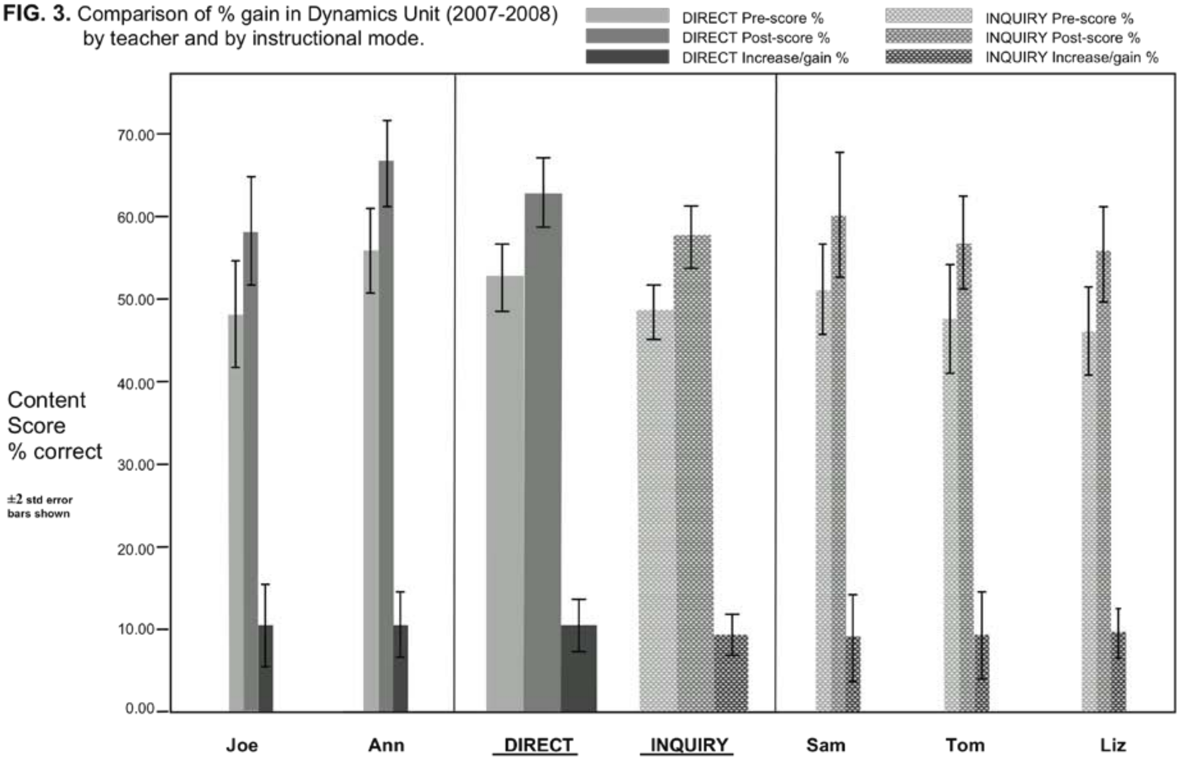
FIG. 2. Comparison of % gain in Light Unit (2007-2008) by teacher and by instructional mode.



Means (SD)	**Joe** (N=32)	**Ann** (N=40)	**DIRECT** (N=72)	**INQUIRY** (N=108)	**Sam** (N=37)	**Tom** (N=35)	**Liz** (N=36)
Pre-score	**50.1** (16.5)	**59.4** (17.9)	**55.3** (17.8)	**53.8** (20.0)	**52.9** (22.7)	**53.2** (19.3)	**55.2** (18.3)
Post-score	**65.5** (18.6)	**70.5** (16.3)	**68.2** (17.4)	**67.9** (22.2)	**64.5** (23.6)	**71.8** (20.5)	**67.6** (22.2)
% Increase	**15.3** (13.2)	**11.0** (13.8)	**12.9** (13.6)	**14.1** (16.4)	**11.5** (16.6)	**18.6** (16.8)	**12.4** (15.3)
Normalized % change	**32.5** (28.3)	**29.9** (30.8)	**31.1** (29.5)	**34.9** (35.9)	**31.5** (38.4)	**43.6** (32.9)	**30.0** (35.4)

FIGURE 3

FIG. 3. Comparison of % gain in Dynamics Unit (2007-2008) by teacher and by instructional mode.



Means (SD)	Joe (N=32)	Ann (N=40)	DIRECT (N=72)	INQUIRY (N=108)	Sam (N=37)	Tom (N=36)	Liz (N=35)
Pre-score	**48.1** (18.0)	**54.6** (14.9)	**51.7** (16.6)	**48.2** (17.3)	**51.2** (16.7)	**47.4** (19.3)	**46.0** (15.8)
Post-score	**58.2** (18.5)	**65.5** (16.2)	**62.2** (17.5)	**57.4** (19.4)	**60.1** (23.2)	**56.6** (17.4)	**55.4** (17.1)
% Increase	**10.1** (13.9)	**10.8** (13.2)	**10.5** (13.4)	**9.2** (13.6)	**8.9** (15.5)	**9.3** (15.5)	**9.4** (8.8)
Normalized % change	**19.6** (28.8)	**24.1** (27.8)	**22.1** (28.2)	**19.0** (29.2)	**21.8** (37.9)	**16.3** (27.6)	**18.7** (19.1)