Depth Versus Breadth: How Content Coverage in High School Science Courses Relates to Later Success in College Science Coursework

MARC S. SCHWARTZ
University of Texas, Arlington, TX 76019, USA

PHILIP M. SADLER, GERHARD SONNERT
Science Education Department, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

ROBERT H. TAI
Curriculum, Instruction, and Special Education Department, Curry School of Education, University of Virginia, Charlottesville, VA 22904, USA

Received 5 March 2008; revised 1 September 2008, 10 October 2008; accepted 17 October 2008

DOI 10.1002/sce.20328

ABSTRACT: This study relates the performance of college students in introductory science courses to the amount of content covered in their high school science courses. The sample includes 8310 students in introductory biology, chemistry, or physics courses in 55 randomly chosen U.S. colleges and universities. Students who reported covering at least 1 major topic in depth, for a month or longer, in high school were found to earn higher grades in college science than did students who reported no coverage in depth. Students reporting breadth in their high school course, covering all major topics, did not appear to have any advantage in chemistry or physics and a significant disadvantage in biology. Care was taken to account for significant covariates: socioeconomic variables, English and mathematics proficiency, and rigor of their preparatory high science course. Alternative operationalizations of depth and breadth variables result in very similar findings. We conclude that teachers should use their judgment to reduce coverage in high school science courses and aim for mastery by extending at least 1 topic in depth over an extended period of time. © 2008 Wiley Periodicals, Inc. *Sci Ed* 1–29, 2008

Correspondence to: Marc S. Schwartz; e-mail: schwarma@uta.edu
Contract grant sponsor: Interagency Eductional Research Initiative.
Contract grant number: NSF-REC 0115649.

INTRODUCTION

One of the most long-lived and contentious conflicts in science education concerns the optimal degree of content coverage in science courses (Anderson, 1995; Katz & Rath, 1992). Frequently, simplified into the opposing camps of "depth versus breadth," the distinction serves to characterize two separate and competing philosophies about which almost all educators hold passionate opinions. At one extreme is the emphasis on "full coverage" or "breadth," a view that students are best served by encountering a great number of topics relevant to a particular science discipline. Many teachers hold that they have an obligation, at a minimum, to cover the widest range of terms and concepts that students may encounter on a standardized test, or if they proceed further in science. The alternative view is typified by the terms "deep coverage," "understanding at many levels," or simply "depth." These proponents hold that there are certain fundamental concepts that are more important or beneficial to master than others and that spending focused time, at the expense of covering many other topics, is a far more productive strategy.

This issue is not new. Beliefs concerning the character and importance of scientific knowledge, as well as the nature of learning and learners, impact the positions educators have staked out concerning coverage. The original argument for "breadth" may well have been Aristotle's who argued for a broad curriculum (Shubert, 1986), at a time when a person could—and ideally should—learn everything there was to know about the natural world (Newmann, 1988). This view was rarely contested through the end of the 19th century. Authors often built on this premise, such as Jackman (1894), who asserted that broad curricula suited the cognitive ability of students.

However, the explosion of scientific knowledge during the 20th century forced the consideration of the alternative view. Advocates argued the importance of studying fewer topics in greater depth (Beittel et al., 1961; Hirsch, 2001; Newmann, 1988; Perkins, 1992; Schmidt, Hsing, & McKnight, 2005; Sizer, 1984; Westbury, 1973). Educators subscribing to this school of thought maintain that students should develop depth of understanding, rather than aim for maximum coverage, claiming that mastery of a few topics trumps the failure to master any. This view was bolstered during the last decade of the 20th century by the American Association for the Advancement of Sciences (AAAS) in their publication *Science for All Americans* (1989), which was later reinforced in their *Benchmarks for Science Literacy* (1993) and two influential books from the National Research Council (NRC), *How People Learn* (1999) and its more recent publication *Taking Science to School* (2007). Following this initiative, states like California, among others, explicitly affirmed this position in their framework for public education: "Content should . . . value depth over breadth of coverage" (Butts & Precott, 1990). Later in this important decade of educational reform in science, the NRC authored the *National Science Education Standards* (1997) in which depth of study is an embedded theme whose importance is still evident, although less explicit, in their most recent publication, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (Committee on Science, Engineering, and Public Policy, 2007). Yet these documents do little to resolve the problem of "science teachers struggling with the tension of pursuing science topics in depth, as required by the standards, versus the pressure to 'get through' the breadth of the provided curriculum material" (Van Driel, Beijaard, & Verloop, 2001, p. 147).

Contributing to this pedagogical tension is the way science textbooks are written. Publishers include chapters on all possible topics, to improve a textbook's chances of adoption, thus creating an "encyclopedic" curriculum. Individuals or committees who select the text for these books naturally have a range of views about what is important, and with enough variation in views, the collection of possible topics has led to science textbooks becoming massive tomes with little focus or relevance to students (Kesidou & Roseman, 2002; Statler, 1997). In Kesidou and Roseman's review of middle school science textbooks, the authors point out the failure of the most commonly used textbooks in achieving meaningful levels of in-depth understanding. They maintain that this problem extends to curricula supported by the National Science Foundation (NSF). However, this is not to say that authors and curriculum projects have not sought to reduce the number of topics and concepts in a textbook or curriculum (e.g., Project STAR, GEMS). Such efforts are supported by agencies such as the NSF but are far from dominating the curricula chosen in U.S. schools.

Whereas the preference for depth of study has become popular among many educators, others argue that students will be better served if the pendulum swings back toward breadth of study. Although state science standards are based on either the AAAS Benchmarks or NRC National Standards, states typically augment these national standards with many more topics (Schmidt et al., 2005).[1] The concomitant state tests are then based on these state standards, leaving teachers with the difficult task of deciding how much to cover. Many teachers fear that leaving out a topic will doom their students to miss relevant state examination questions, and that this will reflect on their own performance and that of their school.

As Kirst, Anhalt, and Marine (1977) noted, the focus on breadth instead of depth is one of political necessity. "In short, in order to mitigate political disputes many curricular decision makers use pragmatic methods of decision making that result in marginal changes. Conflict can be avoided by using vague language concerning standards and covering so many topics that no major interest group feels left out. Content priority is sacrificed to the political necessity of breadth in coverage" (p. 316). Additionally, there is the more subtle message shaped by the popular culture through television shows such as "Jeopardy" and "Are You Smarter Than a Fifth Grader?" Such contests reward participants lavishly for memorizing a wide range of unrelated facts about the world. Both situations put a premium on curricula that focus on a broad range of information.

Despite the long history and high profile of the breadth versus depth debate, the educational research literature provides precious little empirical evidence that supports either approach or one over the other. Educators and policy makers in the end rely on their own experience as learners or anecdotal evidence when deciding on the amount of coverage and level of detail students will encounter in the curriculum. The purpose of this study is to generate a modicum of systematic empirical evidence that bears upon an issue that hitherto has been primarily argued passionately at a philosophical or simply anecdotal level. We wish to contribute by examining the relationship of a high school teacher's choice to focus on depth or breadth of subject matter content coverage to students' subsequent performance in introductory college science courses on biology, chemistry, and physics.

LITERATURE REVIEW

Many prominent educators have advocated during the last century for a focus on in-depth learning of concepts to support richer understandings (Bloom, 1961; James, 1890; Novak & Gowin, 1984; Piaget, 1983; Sizer, 1984; Vygotsky, 1978; Whitehead, 1929). Particularly noticeable in the discourse is the utilization of value-laden terms such as "rich" or "deep" to characterize the kind of understanding that educators wish to develop and support in

---

[1] In addition to the author's observations (p. 558), we noted that despite the NRC Standard's attempt to eschew the teaching of atoms and molecules until the ninth grade, California's science standards introduce atoms and the periodic table at the third-grade level.

students. Rich understanding manifests itself as the ability of students to recognize patterns in data that fit larger, more inclusive theories, allowing students to easily find ideas or facts that support general conclusions and helping them recognize that understanding is context sensitive. For students to achieve this kind of relationship with knowledge, many of the authors cited earlier argue that an extended amount of time is a necessary, though not sufficient, precondition. As a culture, we have reached a point where this "in-depth" hypothesis has become nearly self-evident, considering the general strength of currently available cognitive models of the learning process.

Cognitive psychologists have pointed out that learning evolves through a coordination of ideas and experiences that requires focus and effort over time (Fischer & Bidell, 2006; Grannot & Parziale, 2002; NRC, 1999; Siegler, 1998). The models of cognitive development that these authors have developed and use in educational contexts stress the importance of building and rebuilding ideas in multiple contexts to achieve general principles that can be applied to new problems. Such an educational focus would of course require more time, a theme found in numerous books and position papers (AAAS, 1989; NRC, 1999; Schmidt et al., 1997; Wood, 2002). Most recently, the importance of developing curricula that allow students the time to focus in depth on subject material was clearly summarized in *Taking Science to School* (NRC, 2007), "Studies of instructional interventions carried out over weeks or months indicate that, with opportunities to practice or explicit instruction, even elementary and middle school children can master difficult concepts in science. However, to be successful, students need carefully structured experiences, scaffolded support from teachers, and opportunities for sustained engagement with the same set of ideas over extended periods of time (weeks, months, even years)" (p. 338).

Eylon and Linn (1988) reviewed the literature in cognitive science and science education prior to 1988 in their quest to highlight strategies for improving science education. They cite more than 200 references in their synthesis of four major research perspectives used in science education: concept learning, developmental, differential, and problem solving. Viewed collectively, the authors concluded, "... in-depth coverage of several science topics will benefit students far more than fleeting coverage of numerous science topics" (p. 251). However, they also recognize that "few empirical studies demonstrate the efficacy of suggested methods" (p. 252). This is the same concern voiced over 25 years earlier by Beittel et al. (1961), who noted that there were few studies available to support one position concerning coverage over the other. Although researchers such as Eylon and Linn have attempted to synthesize findings from a large number of research papers to provide a general framework for advocating depth over breadth, the few empirical studies in the literature do not collectively provide a clear direction for educators, as each case is unique due to differences in goals, methods, and/or context.

Empirical Studies

We begin with a few initial studies that clearly address the issue of depth versus breadth, although the empirical evidence to support the pedagogical strategy that logically follows from a theoretical position is less well developed. Onsoko (1989) found that as teachers grow in sophistication, they tend to emphasize "thinking and content exploration." These goals also lead "them to view breadth of coverage more negatively, to identify their source(s) of coverage pressure as externally imposed, to voice greater objections to this pressure, and to resolve the coverage dilemma in the direction of greater depth. . . " (p. 191). Although the teachers' views were often passionate, and there was general agreement among more experienced teachers, they had little opportunity to test their assumptions. More importantly, this study was not structured to resolve the issue.

Li, Klahr, and Siler (2006) focused on controlling for coverage to evaluate student performance on standards-based science tests. Their goal was to close the gap in standardized test scores between schools representing substantially different socioeconomic status (SES). Their treatment group was drawn from low SES schools, whereas their control group was from high SES schools. With the same test taken by all middle school students, they were able to narrow the gap in test performance, but only for items for which instruction time was increased. Although they "... push [ed] for mastery by narrowing [the] focus on skill or concept domains..." (p. 2), a large gap still persisted on the remainder of test items for which teachers left the related content uncovered. Ultimately, the authors could not claim that an increase in depth and reduction in breadth helped students achieve higher overall success in standards-based examinations.

In another attempt to understand the impact of depth over breadth, Goforth and Dunbar (2000) carried out an 8-year study at Baylor University in which instructors focused on only two areas in their introductory geology course rather than on the broad range of subjects that one might expect to find in such a course. Using postinstruction surveys to differentiate student views in these two types of courses, the authors concluded, "The student responses support the idea that exploration of a relatively narrow subject in depth offers many opportunities for discovery-based learning" (p. 202). Even though students claimed that the course covering two topics increased their interest in geology, no measure of preinstruction interest was noted. Furthermore, no comparative data were reported to evaluate changes in student understanding in a more traditionally taught course.

In a larger survey study, Sadler and Tai (2001) found that high school physics teachers' decision to focus in depth on fewer topics, rather than on broad coverage, was predictive of students' higher grades in their college physics course when controlling for student background factors. Also interesting was the finding that students who had not employed a textbook in their high school physics course achieved higher college grades than did those who had used a textbook.

Although the nature and use of textbooks generate important questions regarding how to use textbooks effectively, a few important studies on science curricular material focus on developing an understanding of specific topics over time, such as thermodynamics (Clark & Linn, 2003) or mechanics (White & Frederiksen, 1998). In general, these authors found that as middle school students spend more time exploring a science topic, they develop a deeper understanding. Linn and Hsi (2000) describe this deeper understanding as better "knowledge integration," which they claim survives into high school. Their students, who participated in a 12-week middle school curriculum, outperformed students in high school who had encountered traditional curricula. Clark and Linn (2003) also shortened their original 12-week curriculum by 2, 4, or 6 weeks and noted that the student's ability to develop knowledge integration decreased in a linear fashion.

The results from these three studies (with curricula lasting 3–4 months) do allow the authors some latitude in judging the impact of time on student performance. However, it is important to note that their studies were not constructed to control for confounding variables such as teaching experience, teaching style, student backgrounds, SES, different environments, or to some extent, even time itself.

Outside the domain of natural science, Kennedy, Rodrigue, and Davis (2000) compared two undergraduate psychology courses to explore the impact of depth versus breadth. One class received the complete survey of the field through 16 chapters, whereas their experimental group covered only 8 chapters. The students in the class focusing in depth on fewer topics were more satisfied with the course. They also performed just as well as the comparison group on examinations. In this case, depth appears to be as effective as breadth.

In a similar vein, Beittel et al. (1961) investigated the impact of depth versus breadth in art education. Their study included a control group as well as two experimental groups of students, one of whom followed a "depth" curriculum, whereas the other followed a "breadth" curriculum. They found that the "depth" approach "produced the greatest gain in individual student progress over a one-year period" (p. 86). However, the authors also noted that students resisted the "depth" approach, preferring the "breadth" approach to which they had become accustomed in prior art classes.

All the studies cited above result from a search of seven major reference databases (ERIC, Education Abstracts, Academic Search Premier, Dissertation Abstracts, Social Science Citation Index, Canada Institute for Scientific and Technical Information, and Google Scholar) and the help of colleagues interested in the issue. In addition, we consulted the Leibniz Institute for Science Education (University of Kiel) listing of more than 7000 journal articles focusing on empirical studies and theoretical issues in the learning and teaching of science (Duit, 2006). There was no shortage of papers advocating philosophical positions, but only a small number of articles that broached the problem through empirical research. These few studies that examine a small number of classrooms, single subject areas, or the efforts of a few teachers are insufficient to convincingly debate the merits of depth versus breadth of study. Hence, despite the impassioned and articulate views of many authors who philosophically argue for one position over the other, the record of evidence to support either is meager. Nonetheless, this lack of evidence has not impeded policy makers and advisory bodies from developing positions on the subject (Kirst et al., 1977; Newmann, 1988).

A noteworthy aspect of the literature on depth versus breadth and the passionate tenor of arguments is the way the debate is often framed through provocative slogans. Sizer's (1984) pronouncement that "less is more" attracted public attention to the tension in pedagogies by effectively summarizing what was at stake—students exposed to "less" coverage would have a greater chance of developing a "more" thorough understanding of the topics encountered. Likewise, we might assume that "an overstuffed but undernourishing curriculum" (Raizen, 1997, p. 12) (where breadth is the focus) is analogous to the brain surviving on junk food. Another slogan that has effectively captured the public's imagination, as well as the author's opinion, concerns curricula that are a "mile wide and an inch deep" (National Commission on Excellence in Education, 1983; Schmidt et al., 1997). In the same spirit, the discussion on breadth versus depth also appears to be a mile wide, and the empirical basis for any reasoned decision is closer to an inch deep. And ironically, although educational philosophers propose one strategy, teachers tend to prefer the other (i.e., breadth) because it is a pragmatic solution to the educational problems they face and a solution that students seem to prefer (Katz & Rath, 1992; Newmann, 1988).

Li et al. (2006) characterized this situation as a misalignment between curricula, standards, and standardized tests (e.g., "high stakes" state tests, Standardized Aptitude Tests [SATs], ACTs, and Advanced Placement [AP] exams) that ultimately drive teachers to focus on breadth at the expense of depth. VanSledright (1997) argues that while curricula that focus on "depth" appear to be promoting a desirable position, breadth often offers curricular coherence for teachers who wish to provide students an overview of an entire field of study. This position is often reinforced by state testing that influences teachers to cover more topics to ensure their students' success (National Center for Educational Statistics, 2006; Onosko, 1989). Nonetheless, Sadler and Tai (2001) point out that, although high scores on high-stakes tests (e.g., state graduation tests, SAT, ACT) may be the short-term focus of many teachers and students, other measures may be more important, such as the adequate preparation for success in later coursework. For those students interested in science-related fields, introductory college science courses serve a “gatekeeping” function, barring or granting access to high-income, high-status professions through science, engineering, health, and technology majors.

The Strength and Weakness of the Dichotomy

Although the “depth versus breadth” dichotomy helps capture the imagination of the public and policy makers in describing and thinking about program goals in education, Wineburg (1997) points out that the dichotomy also impoverishes the discussion by masking and overgeneralizing valued competencies such as being able to differentiate, evaluate, integrate, and qualify concepts, ideas and facts in a discipline. The terms do not distinguish between substantive knowledge and epistemological knowledge (i.e., “knowledge that” versus “knowing how”). Wineburg claims that, while the “mile wide, inch deep” view has helped dramatize the breadth versus depth issue, the slogan also suggests that only one of the two educational positions makes sense.

The role of dichotomies in public debate is not unusual. Hirsch (2001) concluded that the United States has a history of associating educational interventions as either liberal or conservative, which leads to “premature polarization.” Examples that he offers include whole language versus phonics or using calculators versus memorizing the multiplication tables. Such dichotomies have the effect of ending conversations too quickly before developing a well-reasoned position. The suggested polarity in the debate limits one’s ability to understand the strengths in each position. For example, Hirsh offers two educational principles that appear to be in conflict: “The best way to learn a subject is to learn its general principles and to study an ample number of diverse examples that illustrate those principles” and “broad knowledge is the best entree to deep knowledge. . . . The suggested conflict only exists because we have grown accustomed to the ‘sloganized’ polarity between deep understanding and the rote learning of mere facts” (p. 23). Thus readers might assume that to learn a subject they must decide between learning how to apply a few general principles and trying to grasp the many elements that make up a domain. Although either of Hirsh’s principles is a provocative statement that makes intuitive sense, we note that he offers no evidence to support either claim.

In response to black-and-white views, some authors take a more nuanced position on the debate. Murtagh (2001) claims that either extreme is unproductive. His courses provide students the opportunity to experience a balance between the two perspectives where students focus “on one specialty within computer science that uses this specialty as a vehicle to present an overview of the techniques and principles underlying all work in computer science” (p. 37). Wright (2000) also advocates a balanced approach, however, for different reasons. He differentiates the problem by target audience and claims that balance between the two positions is important and necessary to maintain the enthusiasm students display for science during their elementary years. Thus, students in secondary schools should have the choice between courses that develop literacy through broad overviews and those that develop aspiring scientists through in-depth studies. In both cases, a new position emerges out of the apparent dichotomy; however, this position, as all others, is based upon personal experience and anecdotal evidence, not empirical research.

Although the issue of depth versus breadth has been recognized by a number of fields as important, the availability of studies to help educators support their positions is rare. Our study uses two quantitative variables to provide some clarity concerning the relative merits of curricula that focus on breadth or depth. Instead of treating depth and depth as choices along a continuum, we looked at each as categories defined by the time a teacher invests or dedicates to topics within a discipline.

METHODS

The Sample

Our study is based on data collected from college students enrolled in introductory science courses. This data set was created as part of a larger study, “Factors Influencing College Science Success” (FICSS, NSF-REC 0115649). The project used surveys to collect data from students enrolled in 55 four-year colleges and universities out of an original sampling of 67 institutions across all 50 states. We applied a stratified random sampling approach that grouped the colleges and universities according to size of undergraduate enrollment. We chose this approach based on the relative number of institutions versus the relative number of students enrolled. Large state universities accounted for only 5% of 4-year colleges and universities but enrolled nearly 50% of undergraduate students. As a result, our stratified random sampling approach ensured that large institutions would be included in the study sample. For enrolled undergraduates, the three categories we used were small (<3000), medium (between 3000 and 10,000), and large (>10,000). Early on in the research project, we decided to focus on college courses, applying the most common course format: large lectures accompanied by smaller recitation sections and separate laboratory periods. All of the schools selected in the random sampling followed this format.

Factors Influencing College Science Success included a series of surveys involving about 18,000 undergraduate students from across the sequence of introductory college science courses for science and engineering majors. For the purposes of this analysis, we focused only on students enrolled in the first course of each sequence during the fall semester to attain a high level of uniformity in our participants’ college science course experience. Hence, we began with 11,064 students across the disciplines of biology, chemistry, and physics. In addition, we omitted graduate students and nontraditional students, limiting our sample to 10,772 undergraduates for whom high school coursework was apparently a more recent experience. We also omitted students who spent their high school years outside of the United States, since the typical biology–chemistry–physics sequence is not followed in many other countries, reducing the number to 10,320. Some of those had missing values on the breadth and depth variables, mainly because they did not take a high school course in the respective discipline. Our final sample of those for whom we had information on the breadth and depth variables (as well as on the dependent variable, i.e., college grade) comprised 8,310 undergraduates.

In large-scale surveys, missing responses are not uncommon. In dealing with missing responses, list-wise deletion is the simplest and most commonly applied option. For the extended models with additional control variables, N was reduced to 6999 overall, when using list-wise deletions of subjects with one or more missing values on the added variables. The variable with the most missing values was SAT-Math, where 7% of the data was missing. In light of the low percentages of missing values, list-wise deletion was deemed adequate.

The Survey

The purpose of FICSS was to collect evidence that would support or refute hypotheses held by science educators, using a nationally representative sample of college science students. Most studies concerning the impact of curricula (e.g., emphasis on particular content), classroom pedagogy (e.g., laboratory, project work), technology (computer usage), advanced coursework (e.g., advanced placement), gender differences (e.g., in college performance), and structural changes (e.g., *Physics First*, block scheduling, class size) do not allow generalization to the national population. This is because researchers often use samples drawn from a small number of classrooms or single institutions. Moreover, many studies do not employ methods that account for alternative hypotheses related to student background (e.g., parental education, community SES, students' mathematics or English proficiency). To do so, such data must be collected along with the primary independent variables that researchers wish to study.

The FICSS survey instrument was constructed to collect data concerning high school biology, chemistry, and physics experiences of those students enrolled in the introductory course in the same college subject. Items were constructed after an examination of relevant research literature and interviews with professors and high school science teachers for relevant predictors (Schwartz, Hazari, & Sadler, 2008). Of particular help in constructing nonambiguous questions was prior survey research experience surveying 2000 college physics students from 20 different colleges and universities (Sadler & Tai, 2001). Our survey went through a series of developmental steps: student focus groups to critique question formats, two pilot surveys that included 304 college science students, a review by an advisory panel of college professors and high school teachers, and a separate reliability study in which 113 college chemistry students took the survey twice, 2 weeks apart, to compare their responses.

As self-reporting is a central feature of the FICSS survey, it is important to recognize that accuracy and reliability depend on context, relevance, and survey clarity (Bradburn, 2000; Niemi & Smith, 2003; Pace, Barahona, & Kaplan, 1985). In addition, Kuncel, Credé, and Thomas (2005) point out that self-reporting is reasonably accurate in samples where the surveys address issues relevant to the respondents. Thus by surveying introductory college science students in their first college science course, we felt that their prior high school experience would be relatively fresh and poignant. By the same logic we excluded from our sample graduate and special students whose high school experiences were much more distant. By having professors administer the survey in class, we increased its perceived importance.

The respondents were also a good match for the task of filling out the survey as college students are highly accurate in self-reports of course taking, grades earned, and standardized test scores (Anaya, 1999; Baird, 1976; Sawyer, Laing, & Houston, 1988). We focused on increasing accuracy and recall through careful design, contextual cues, and participant relevance. Students were asked to complete four-sheet survey booklets while in their college classes. The cover page asked for students' names and ID numbers, and later college instructors entered students' final course grades and then removed student identifiers. Students reported on demographic variables (parents' education, profession, etc.). Home ZIP codes allowed us to match the surveys with existing data on median income, home value, and average educational background of each home locale. To view the questions, download a sample survey at: http://www.cfa.harvard.edu/smg/ficss/research/survey.html.

Selection of Variables

Dependent and Control Variables. The dependent variable in this study is the grade reported by each professor for each student. We choose this measure as an outcome variable and did not develop our own measure to test student knowledge in biology, chemistry, or physics, for the following reason. Athough introductory college courses are similar in content, it is the professor who decides which topics to emphasize and the weight that he or she gives to examinations, homework, project work, class participation, and laboratory work. Moreover, it is the grade awarded in these introductory courses that serves to encourage or bar students from further study in each discipline. In this way, the grade that a student receives can be considered a more authentic measure and one with more real-life relevance than the score obtained on a test constructed by an outside party. We recognize that the academic grade in introductory college courses is not only the standard for measuring the fruition of precollege preparation but also an accessible and universally measured variable. We seek not to interpret our finding as a simple representation of the amount of science a student has learned. Rather, our analysis predicts how well a student performs in comparison with others in their college science course, and this may involve memorization of facts, understanding of concepts, and acquisition of laboratory and other skills. We wish to measure the significant differences between college science students who earn high grades compared with those with low grades. The numerical or letter grade reported by professors was converted to a number on a 100-point scale (e.g., A = 95, A− = 91, B+ = 88).

To control for differences associated with the course and with the institution, hierarchical linear models were used, as described below. In our multivariate models, we further controlled for the effects of background factors that have been found to affect course performance (Tai, Sadler, & Mintzes, 2006). Each of the three college disciplines (biology, chemistry, and physics) was treated as a separate data set, offering the opportunity to use three separate samples to test hypotheses. If similar results and patterns were found for each, this would support an argument for generalizability of our findings to all sciences. For each subject, we accounted for the college year in which students were taking their introductory science course. Parental education (averaged between two parents) served, as it does in many studies, as a proxy for students' SES. This variable ranged from "Did not finish high school" to "Graduate school." The average educational level and income in a students' home ZIP code was averaged as a measure of the community's SES. To distinguish between institutional types of high schools in our sample, we accounted for students' attendance in public high school (which constituted 80% of the sample) versus attendance in other types of schools (e.g., private, parochial, home-schooled).

Because mathematics course-taking and performance are predictors of student success in science, we include several measures from this area in our models (Sadler & Tai, 2007). We use the students' SAT quantitative scores, when reported, and students' ACT scores are converted using concordance tables (Dorans, 1999; Dorans, Lyu, Pommerich, & Houson, 1997). We include whether students took either calculus or AP calculus (AB or BC) in high school, as well as their grade in their most advanced mathematics course. As a measure of English proficiency, we use students' grade in their last English course. We also take into account whether students had an Honors or AP course in the same subject as their introductory college course and their last grade in that last science course.

Independent Variables. As mentioned, we are intent on a quantitative test of the breadth versus depth debate, as arguments concerning pedagogical strategies have so far remained almost exclusively at a philosophical level. The definitions of breadth and depth, as they are typically used, are often ambiguous. To attempt an empirical analysis of the phenomenon, we thought it is advisable to reduce the conceptual complexity in favor of simple quantitative indicators that are based on the amount of time spent in high school classrooms on topics that are generally taught. Our teacher/professor interviews and review of textbooks allowed us to characterize biology and chemistry with eight key topics. For physics, the total was 6. The choices for the amount of time spent were none at all—a few weeks—a month—a semester—it was a recurring topic.

In high school biology, cell biology was the topic most commonly reported as studied for the long periods of time and the one most rarely left uncovered (Figure 1). History of biology was the topic most frequently reported as a recurring and the topic most frequently left out.

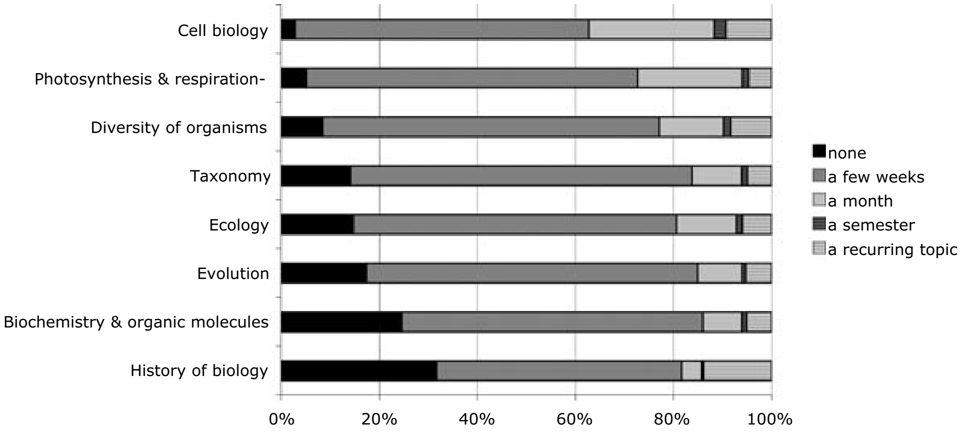

Figure 1. Frequency of coverage of topics in high school biology.

In high school chemistry, the periodic table is the topic most commonly reported as recurring or studied for a semester (Figure 2). Biochemistry is reported as not covered at all by the majority of students. The topic of nuclear reactions is reported as not covered by 43% of students.

In high school physics, mechanics is the topic reported as most emphasized and least likely to be left out. More than a third of students report no coverage of the history of physics, but for many (16%) it is a recurring topic (Figure 3).

In the case of all three high school sciences, the majority of students report covering each topic for a period of a few weeks to a month (with the exception of biochemistry in chemistry). Although this observation is an argument for uniformity in teachers' curriculum decisions, there is a wide variation at the ends of the coverage spectrum. Some topics were often avoided (e.g., 55% reported no biochemistry). Some topics were often reported as covered extensively or as a recurring topic throughout the year (e.g., 35% occurrence for the periodic table).

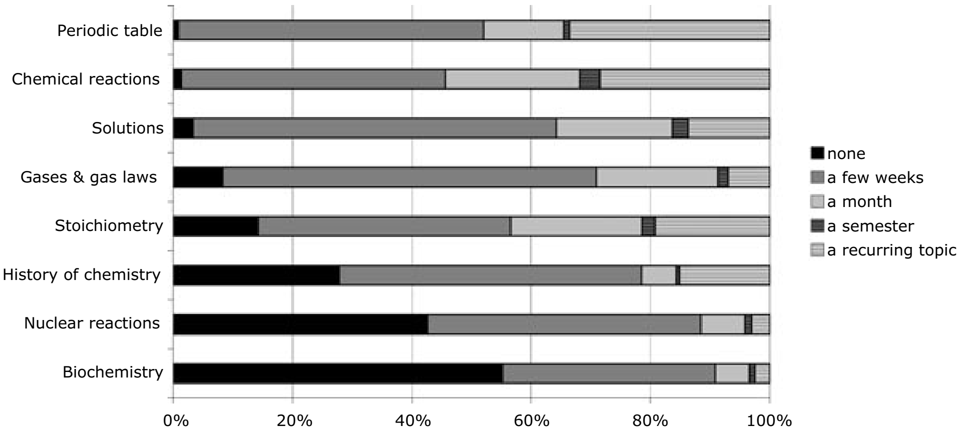

Figure 2. Frequency of coverage of topics in high school chemistry.

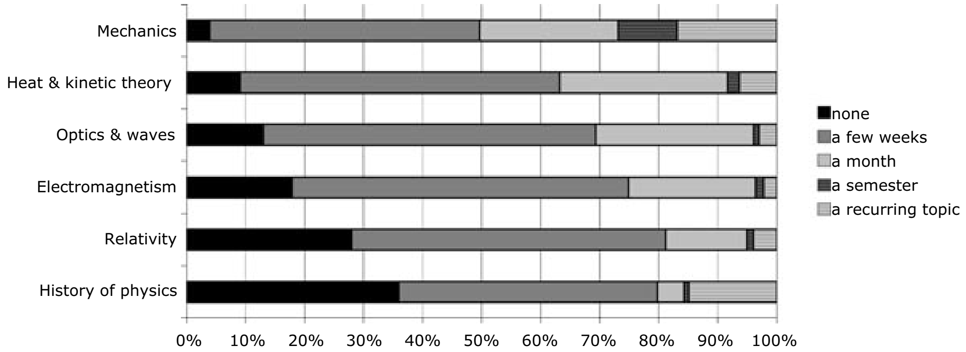

Figure 3. Frequency of coverage of topics in high school physics.

On the basis of these data, for the purpose of our analysis, we defined two dummy variables for breadth and depth in the following way.[2] Breadth was set to 1 if the student reported that at least some amount of time was spent on every one of the mentioned topics; the dummy variable was 0 if the student reported that no time was spent on one or more of the mentioned topics. Depth was set to 1 if the student reported that 1 month or longer was spent on at least one of the mentioned topics (i.e., exposure for a month, a semester, or in a recurrent fashion). It was 0 otherwise. This characterization differs from the common assumption that depth and breadth are opposites. We find that they are not. The correlations between these two variables calculated for each subject are biology, .01; chemistry, .02; physics, .05; and .03 overall. These resulting correlation figures indicate that these measures—as defined in this analysis—appear to be unrelated. The fact that there are no unpopulated cells within Table 1 is also an indication that breadth and depth may be unrelated to one another, since 23% of students report courses with both breadth and depth and 21% report neither.

Although these are averages across the three sciences, another interesting pattern emerges by looking at the commonality of not covering topics and of spending focused time on topics. Roughly two thirds of all students report that their high school course emphasized one or more topics for a month or longer, or that it was a recurring theme. This means that a large fraction of students were exposed to a curriculum in which at least one topic area was emphasized in great depth. In addition, the students surveyed were roughly split, concerning breadth of coverage with the smaller portion spending at least a few weeks on every topic listed, and the remainder reporting that at least one listed topic was not covered at all. This breakdown of depth and breadth reported by subjects concerning their high school science course is shown in Figure 4.

ANALYSIS

Multiple linear regression (and the related technique of hierarchical linear modeling [HLM] used here) is the ideal tool for the proposed analysis because it can calculate the predictive values of variables while holding others constant, isolating the effect of conditions that may covary. The inferential analysis uses baseline and extended models for the prediction of students' grades in their college science courses. The baseline models examine only depth and breadth variables. The extended models include relevant demographic and control variables that account for differences in student background and preparation.

---

[2] Alternative characterizations of these definitions are explored later in the paper.

TABLE 1
Frequencies of Reported Depth and Breadth Across Three Subject Areas Depth of Coverage: None or At Least One Topic Recurring or Covered for a Month or More

	Biology			Chemistry			Physics			All		
Breadth	None	≥1 Topic	Total	None	≥1 Topic	Total	None	≥1 Topic	Total	None	≥1 Topic	Total
≥1 topic not covered	25%	31%	56%	19%	52%	71%	19%	39%	58%	21%	41%	63%
All topics covered	20%	24%	44%	8%	21%	29%	16%	27%	42%	14%	23%	37%
Total	45%	55%	100%	27%	73%	100%	34%	66%	100%	36%	64%	100%

Each is a three-level hierarchical linear model, accounting for the structure that students are within courses, that these courses are within institutions, and that grading stringency may differ by professor and college.

The method of HLM used in this paper resembles the more widely known method of ordinary regression (Raudenbush & Bryk, 2002). From a technical point of view, HLM is more appropriate than the ordinary regression in our study because the particular data structure with which we are dealing here violates a basic assumption of ordinary regression—that of the statistical independence of each observation in the sample. The observations (i.e., students) in our sample occurred in three hierarchical levels: students within classes (of a professor) within institutions. HLM accounts for this data structure and divides the total variance into three variance components, one at each level: variance associated with students, with classes, and with institutions. This approach produces estimates of standard errors that are more conservative than in ordinary least squares (OLS). Although OLS and HML produce similar values for the coefficients of independent variables, HLM ensures greater reliability within groups and can control for error created by differences between groups at each level.

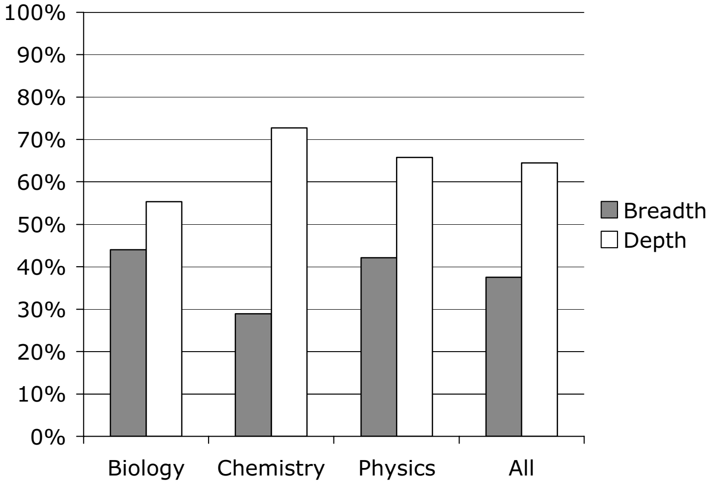

Figure 4. Percentages of students taking high school courses where depth or breadth was present. (Note: Breadth is signified when all major topics are covered. Depth is signified when at least one topic is covered for longer than a month.)

TABLE 2
Three-Level Baseline Hierarchical Linear Models Predicting Grade

	Biology		Chemistry		Physics	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	78.9***		78.8***		80.6***	
Breadth	−0.99**	0.38	−0.37	0.41	−0.56	0.53
Depth	2.28***	0.38	2.21***	0.42	2.88***	0.56
N		3280		3382		1648
Pseudo-R^2		0.139		0.118		0.200

*$p < .05$. **$p < .01$. ***$p < .001$.

To ascertain whether a three-level hierarchical model was appropriate for our data set, we fitted a three-level unconditional means model. For the whole data set, this model partitioned the total variance into the three levels, as follows: —within students, 85.1% ($p < .0001$); between classes, 8.3% ($p < .0001$); and between institutions, 6.7% ($p = .0012$) (Tables 2 and 3). Whereas all three levels contained significant amounts of variance, the bulk of the variance clearly resided at the student level.[3] In this article, we focus exclusively on student-level variables and do not consider any variables at the class or institution levels. Our three-level hierarchical models merely serve as a method to control for variance residing at these levels.

Whereas, in ordinary regression, the R^2 statistic (the proportion of outcome variation explained by the independent variables) is a convenient and intuitive measure of how well a particular model accounts for the dependent variable, the situation is more complicated with HLM because here the total outcome variation is partitioned into several components. It is still a matter of debate among statisticians what measure should take the place of the R^2 statistic in HLM. We used a pseudo-R^2 statistic that is based on the total outcome variation (across all levels) and thus is closely analogous to the R^2 statistic in ordinary regression (Singer & Willett, 2003, pp. 102–103). One way of computing the R^2 statistic in ordinary regression is to square the sample correlation between the observed and the predicted values of the dependent variable. We use this method to calculate the pseudo-R^2 statistic for our HLM models.

Baseline Model

The baseline model is a three-level hierarchical linear model (students within courses within institutions) that predicts college biology, chemistry, or physics course performance from the high school breadth and depth variables for the same subject course (if taken).

At this level of analysis, depth has a consistently positive and significant relationship to college performance whereas breadth has a significant negative relationship in biology only.

---

[3] For the individual disciplines, a more complex picture emerged from the unconditional means models. In biology and chemistry, the variance components associated with the course level were below the significance level, whereas in physics the institutional level was nonsignificant. For analyses of individual disciplines, two-level models would also be viable, but they would have to be different for physics and the other two disciplines. Because this would make the analyses more cumbersome and potentially confusing, we decided to use a single model type—a three-level hierarchical model—throughout the main body of this article. We did estimate a set of two-level models for each discipline to parallel the three-level models reported in Tables 2 and 3. As was to be expected, given the relative dearth of variability at the course and institutional levels, the results were very similar. In particular, in each pair of parallel models, the variables of interest (breadth and depth) behaved identically in terms of reaching or failing to reach the significance level.

TABLE 3
Student-Level Variables of the Three-Level Extended Hierarchical Linear Models Predicting College Science Grade in Biology, Chemistry, and Physics

		Biology			Chemistry			Physics		
Category	Variable Name	Est	SE	*p*	Est	SE	*p*	Est	SE	*p*
	Intercept	39.03	2.05	<.0001	39.21	2.02	<.0001	39.60	3.04	<.0001
College year	Sophomore year	0.25	0.48	.6072	−0.17	0.46	.7180	−0.22	0.87	.8010
	Junior year	0.38	0.62	.5373	−0.16	0.64	.7989	1.31	0.94	.1634
	Senior year	1.72	0.83	.0385	1.05	1.00	.2930	0.09	1.10	.9323
Background factors	Parents' education	0.82	0.20	<.0001	0.66	0.18	.0003	0.27	0.28	.3323
	Public high school	−1.18	0.51	.0202	−0.91	0.47	.0526	−0.18	0.72	.7976
	Community SES level	0.13	0.11	.2309	0.36	0.11	.0014	0.20	0.15	.1786
	Last HS English grade	2.59	0.35	<.0001	1.17	0.32	.0002	1.26	0.47	.0075
Math	SAT quantitative score	0.02	0.00	<.0001	0.02	0.00	<.0001	0.02	0.00	<.0001
	Regular calculus taken	1.66	0.57	.0038	2.21	0.52	<.0001	1.76	0.73	.0158
	AP calculus taken	2.65	0.46	<.0001	3.04	0.46	<.0001	2.74	0.68	<.0001
	Last HS math grade	1.95	0.27	<.0001	2.89	0.26	<.0001	3.20	0.39	<.0001
Science	Second year	2.57	0.59	<.0001	3.64	0.59	<.0001	2.39	1.13	.0350
	Last HS science grade	2.04	0.27	<.0001	2.11	0.27	<.0001	2.31	0.44	<.0001
Coverage	Breadth	−1.15	0.37	.0020	−0.74	0.39	.0557	−0.31	0.52	.5512
	Depth	1.22	0.37	.0011	0.89	0.40	.0251	1.54	0.56	.0061
	N	2760			2902			1337		
	Pseudo-R^2	.315			0.335			0.385		

Extended Model

The extended three-level hierarchical linear model includes numerous control variables in addition to breadth and depth.

The additional variables in the extended model were chosen to control for relevant differences in student background and high school achievement that could reasonably contribute to performance in a college science course. Each variable is significant for at least one of the subject areas examined. Year in college, particularly senior year, is a significant predictor of final grade. The maturity and experience of students is imagined to grow over their college education, and seniors represent the most experienced of undergraduates. Background factors listed are significant for biology, chemistry, and physics. The impact of highly educated parents, private school education with its smaller class sizes, community affluence, and facility in English are all viewed as substantive advantages for students. Likewise, measures of mathematics performance and course-taking are significant for each subject. Mathematics is the "language of science," and facility with quantitative problem solving and interpretation of graphs bestows an advantage on students. Taking a second year of a subject in high school (typically AP) and the grade in the last science course are both significant variables for each subject. Students exposed to a subject for 2 years in high school have a greater mastery of the subject and perform better in the college subject.

The depth variable remains significant for all subjects and the breadth variable remains significant in biology after the addition of background variables in the extended model. Compared with the baseline model containing only those two predictors of interest at the student level, the magnitude of the depth and breadth coefficients decreases with the inclusion of control variables by roughly half while the variance accounted for by the model doubles. List-wise deletion reduced the biology sample by 16%, the chemistry sample by 14%, and the physics sample by 19%. As expected from the initial findings, the positive depth effect was most pronounced in physics. The positive effect also remained in biology and chemistry. In consideration of breadth, the effect was negative in all three domains and significant in biology (not chemistry or physics). There is no support for the view that breadth is beneficial in chemistry or physics, and it appears to be counterproductive in biology (Figure 5).

Nonetheless, the general pattern was similar across all three disciplines. This was formally examined by inserting interactions between breadth and depth and the disciplines into an extended model that included the students of all three disciplines. None was significant.

The effects of the depth and breadth variables are small, on the order of a single point on a 100-point scale (Figure 5). Calculating effect sizes for the breadth and depth variables is a simple matter of dividing these variables by the standard deviation of student grades for each subject (Table 4). The resulting effect sizes are small, with the magnitude of significant parameter estimates on the order of one tenth of a standard deviation in course grade. However, a useful way of interpreting the magnitude of the effect of breadth and depth is to compare these effects with the magnitude of another variable in the model—an additional year of high school instruction in the scientific discipline. This allows the interpretation of teaching with or without depth or breadth to be compared to the relationship between taking 1 or 2 years of the science subject in high school (students with no exposure to the subject in high school are not included in the model, because they could not answer questions about depth or breadth). The difference between taking 1 or 2 years of the science subject in high school is 2.57 in the college biology course, 3.64 for chemistry, and 2.39 for physics (see Table 3). This represents roughly one quarter of a letter grade in college science, which is much smaller than most teachers assume (Sadler & Tai, 2007). By dividing the depth and breadth coefficients in Table 3 by the appropriate 2nd-year coefficients, the result is the

TABLE 4
Magnitude of Depth and Breadth Variables

		Biology	Chemistry	Physics
College course grade	Mean	79.95	80.29	81.47
	SD	11.38	11.36	11.55
Effect size in units of SD	Breadth present	−0.10**	−0.07	−0.03
of college grade	Depth present	0.11**	0.08*	0.13**
Effect in units of a year of high	Breadth present	−45%**	−20%	−13%
school study in subject	Depth present	48%**	25%*	64%**

impact of coverage in units of a year of high school study in the subject. The results offer an alternative explanation of how teacher decisions concerning coverage play out. Students who experience breadth of coverage in high school biology perform in college biology as if they had experienced half a year less preparation than students without breadth of coverage, whereas those who are exposed to in-depth coverage perform as if they had had half a year more preparation than the students without depth of coverage. In chemistry, depth appears to be equivalent to one quarter of a year more of high school preparation. In physics, the effect is closer to two thirds of a year more preparation.

As discussed earlier, depth and breadth appear unrelated, and students report that their teachers employ coverage in differing combinations. For each high school science subject, the most popular matching of depth and breadth is Depth Present, Breadth Absent as shown in Figure 6a. The least popular combination is Depth Absent, Breadth Present. High school colleagues suggest these characterizations of each category in reference to how the textbook is followed:

- Depth Absent, Breadth Present: Covering each textbook chapter in about a week to finish the text in a year, without spending focused time on any topic.

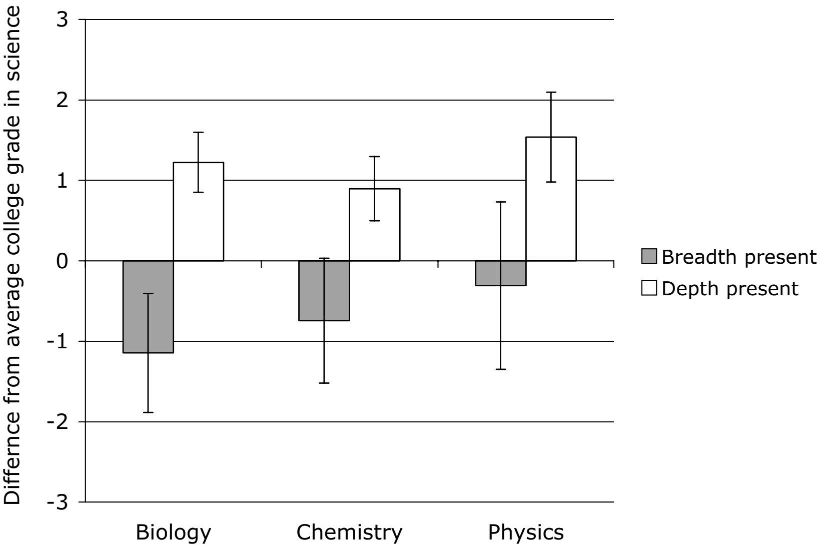

Figure 5. Comparison to average college grade of students reporting the presence of depth and breadth in their high school science courses (values on the *y* axis are the number of points on a 100-point scale).

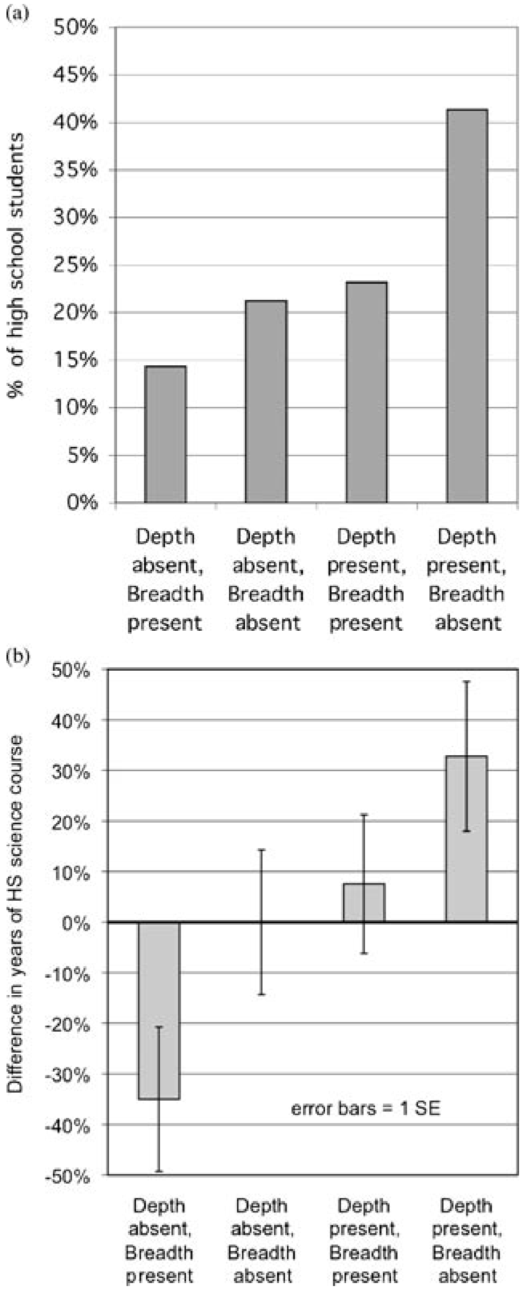

Figure 6. (a) Comparison of frequency of coverage groups with subjects reporting different combinations of depth and breadth in their high school science courses. The baseline chosen is both absent. (b) Comparison of performance of coverage groups with subjects reporting different combinations of depth and breadth in their high school science courses. The baseline chosen is both absent.

- Depth Absent, Breadth Absent: Covering a reduced number of textbook chapters without spending extended time on any topic.
- Depth Present, Breadth Present: Covering each textbook chapter in a short period of time to finish the text in a year and spending additional focused time on at least one particular textbook chapter.

- Depth Present, Breadth Absent: Spending focused time on at least one particular textbook topic, but not covering every textbook chapter.

By expanding the extended model to encompass the four combinations of depth and breadth, we can see in Figure 6b the relative advantage of each approach in units of high school years of study. We use as a baseline the combination of depth absent and breadth absent. Students who were in high school classrooms in which depth was present (but some topics were left uncovered) later performed at a level equivalent to having had one third more of a year of the science subject. Students in classrooms where all topics were covered and none received extended time later earned grades in college equivalent to students with one third less of a year of high school science.

Alternative Definitions of Variables

One concern regarding the findings on coverage is that they could be an artifact of our definition of depth and breadth. Because the quantitative analysis of breadth and depth conducted here is relatively novel, we could not rely on any standard definitions. At issue, therefore, is the appropriateness of the definitions we chose. Hence, we considered it useful to include the following special section to explore different operationalizations of our independent variables. By varying how these variables are defined, we can gauge the robustness of our conclusions.

– *Alternative cutoff for depth variable: From "a month" to "a semester."* This decreases the percentage of students for whom the depth variable is set at 1 from 67% to 58% in the sample. When this alternative operationalization is used, the effects of breadth and depth on performance in the models remain similar. There is no significant interaction term between depth and breadth in any of the three science disciplines. Thus, our findings were robust under two different variable definitions for depth.

– *Depth as duration of instruction versus "a recurring topic."* Another potential issue with the operationalization of the depth variable might lie with the impact of the choice, "a recurring topic," being counted as a marker for depth. Two new conditions, "depth:duration" and "depth:recurring" were formed by separating out the "recurring" response from the original depth variable. A combination "depth:both" was also created for subjects with both conditions. Statistical *contrast analyses* within the three-level hierarchical models (including the usual controls) were employed to examine empirically whether our originally chosen depth variable definition made substantive sense. The contrast analyses allowed us to determine if specific levels of the new categorical depth variables (no depth, depth:duration, depth:recurring, depth:both) differed from each other in predicting the outcome. We find that in none of the three disciplines were the outcomes different between students with any of the three new depth conditions. On the other hand, in all three disciplines, there were significant differences between students with no depth and those who reported depth in any form. Hence, one can conclude that the real "action" is between depth and no depth, whereas internal differentiations of depth—splitting the "recurring" level from time amount levels—are less important. This outcome supports the use of a single, combined depth dummy variable in the regression models.

– *Number of topics treated in depth.* A five-level ordinal depth variable was created, distinguishing students with 0, 1, 2, 3, or $\leq$4 topics receiving additional time. Assuming that some amount of depth in their high school science courses is beneficial for the students' performance in college-level introductory science courses, it appears reasonable to hypothesize that the graph of introductory college course grade versus a variable measuring how many topics were treated in depth would take the shape of an inverted U curve. Some amount of depth should be more advantageous than no depth at all, but at some point the beneficial effect of depth in additional fields should attenuate because students would have already reaped the benefits stemming from in-depth study, and the potential disadvantage of the students' narrowness of knowledge might become more salient. In graphs of all three disciplines, this type of shape was indeed found in a series of three-level hierarchical models (including the control variables mentioned above), although differences were so small that no firm conclusions can be drawn about the precise shape of the pattern. On the other hand, the relative similarity of the coefficients for various numbers of depth fields validates our approach of treating depth as a dichotomous variable that distinguishes depth in no field and depth in one or more fields.

- *Number of missing topics.* An analogous analysis was conducted for the breadth issue. A four-level ordinal breadth variable was created, distinguishing students with 0, 1, 2, or ≤ 3 missing topics. It is plausible to assume that the graph of introductory course grade versus the breadth variable would take the shape of an inverted U curve, as discussed above. In biology and chemistry, this general shape was found. Students with one missing topic had an advantage over those with complete coverage. Students with two missing topics reaped the largest advantage, and those with three or more missing topics had slightly lower grades than those with two missing fields. In physics, by contrast, students with three or more missing topics did the best. (An additional analysis showed that, in physics, diminishing benefits began only at four or more missing fields.) However, none of the physics coefficients reached significance. The differences found between differing numbers of missing topics were again small and validate our approach of conceptualizing breadth as a dichotomous variable that distinguishes students with complete coverage of all topics from those with one or more "holes."

Special Populations

One important issue that arises with significant findings across the three science disciplines is that the results may be dependent upon some other factor. In particular, depth and breadth may be related to some particular subgroup within our sample. We explore the relationship between depth and breadth and students who attended AP courses as well as highly proficient students. We also consider teacher characteristics.

- *AP students.* Because of the set nature of the AP science curricula and the fact that for most, it is their second high school course in the subject, we considered whether AP science students should be excluded from the analyses. In conducting a series of parallel analyses without the AP students, we find that the results remained mostly unchanged, when using both the original and the alternative cutoff for the depth variable. The general pattern of the depth and breadth effects and of their significance was the same. Hence, it appeared appropriate to include science AP students in our analyses, while using AP as a control variable. Interactions of AP with depth and with breadth were also estimated in the extended model, but neither was significant.
- *Students with strong academic backgrounds.* Do the more proficient students benefit the most from a high school science course that goes into depth, compared with

academically weaker students? We systematically investigated the interactions between performance variables and the depth and breadth variables and found only two interactions significant below the .05 level. Contrary to the main effects analysis results, the significance of the interactions was not consistent and did not provide a clear picture of the connections between the outcome and predictors. The results suggest that the associations for high- and low-achieving students are similar with respect to depth and breadth of content coverage.[4]

– *Students of highly rated teachers.* Our survey asked students to rate several teacher characteristics. If depth and breadth were to a considerable extent correlated with those teacher characteristics, this would make the results harder to interpret because of the conundrum of colinearity: that particular teacher characteristics (rather than depth or breadth of coverage) might be the actual determinants of student outcomes, and that depth or breadth themselves might merely reflect those teacher characteristics by association. Teacher characteristics, however, were found to be only minimally correlated with breadth and depth. Among the teacher characteristics, only very weak correlations are found between breadth and the teacher's rated ability to explain problems in several different ways ($r = .04$). Depth correlated weakly with the ratings of the teacher's subject knowledge ($r = .05$). It appears that highly rated teachers are no more likely than others to teach with depth or breadth.

DISCUSSION

The baseline model (Table 2) shows the association between breadth and depth of high school science curricula and first-course introductory-level college science performance. The most striking analytical characteristic is the consistency of the results across the three baseline models, revealing the same trends across the disciplines of biology, chemistry, and physics. In addition, the contrast of positive associations between depth and performance and negative associations between breadth and performance is striking, though not significant in all cases. This replication of results across differing disciplines suggests a consistency to the findings. However, we caution that the baseline model offers only outcomes apart from the consideration of additional factors that may have some influence on these associations. The robustness of these associations was put to the test, as additional factors were included in an extended model with additional analyses.

The extended model, shown in Table 3, included variables that control for variations in student achievement as measured by tests scores, high school course level (regular, honors, AP), and high school grades as well as background variables such as parent's educational level, community socioeconomic level, public high school attendance, and year in college. With the addition of these variables in the extended model, we find that the magnitude of the coefficients decreased in the majority of instances, but the trends and their significance remained. Keeping in mind the generalized nature of the analyses we have undertaken, the most notable feature of these results stems not simply from the magnitude of the coefficients, but rather from the trends and their significance. Here, the positive associations between depth of study and performance remain significant whereas the negative association between breadth of study and performance retains its significance in the biology analysis. Although this association is not significant in the chemistry and physics analyses, the negative trend is consistent with our result for biology.

---

[4] A significant interaction ($p = .03$) was found between the math SAT score and depth. In addition, a second significant interaction ($p = .02$) was found between breadth and most advanced mathematics grade. Apart from these two significant interactions, all others were nonsignificant.

What these outcomes reveal is that although the choice to pursue depth of study has a significant and positive association with performance, the choice to pursue breadth of study appears to have implications as well. These appear to be that students whose teachers choose broad coverage of content, on the average, experience no benefit. In the extended model, we arrive at these results while accounting for important differences in students' backgrounds and academic performance, which attests to the robustness of these findings. The findings run counter to philosophical positions that favor breadth or those that advocate a balance between depth and breadth (Murtagh, 2001; Wright, 2000).

Finally, a particularly surprising and intriguing finding from our analysis indicated that the depth and breadth variables as defined in this analysis appear to be uncorrelated; additionally, the interactions between both variables are not significant in either model. Figure 6 offers an analysis of the outcomes treating breadth and depth as independent characteristics of high school science curriculum. The first panel reveals that over 40% of the students in our survey fall within the "depth present–breadth absent" grouping, whereas the other three varied between 14% and 23%. The distribution across these four groupings shows that the teachers of students in our survey are making choices about which topics to leave out and which to emphasize. These choices have consequences. The second panel indicates that those students reporting high school science experiences associated with the group "depth present–breadth absent" have an advantage equal to two thirds of a year of instruction over their peers who had the opposite high school experience ("Depth Absent–Breadth Present"). This outcome is particularly important given that high school science teachers are often faced with choices that require balancing available class time and course content. It appears that even teachers who attempt to follow the "best of both worlds" approach by mixing depth and breadth of content coverage do not, on average, provide an advantage to their students who continue to study science in college over their peers whose teachers focused on depth of content coverage.

Rather than relying solely on one definition of depth and breadth, we examined alternative operationalizations of the breadth and depth variables and replicated the analysis. The findings remained robust when these alternative operationalizations were applied.

However, a limitation in our definition of depth and breadth stems from the smallest unit of time we used in the study to define depth of study, which was on the order of *weeks*. Our analysis cannot discern a "grain size" smaller than weeks. Teachers may opt to study a single concept in a class period or many within that period. Likewise, teachers might decide that a single laboratory experiment should go through several iterations in as many days, while others will "cover" three different, unrelated laboratories in the same time span. These differences are beyond the capacity of our analysis to resolve. In the end, our work is but one approach to the analysis of the concepts of depth and breadth generated from students' self-reports of their high school science classroom experiences.

Findings in Light of Research From Cognitive Science

Most students hold conceptions that are at odds with those of their teachers or those of scientists. In general, students cling tenaciously to these ideas, even in the face of concerted efforts (Sadler, 1998). These "misconceptions" or "alternative conceptions" have been uncovered using several research techniques, but qualitative interviews, traceable to Piaget, have proven highly productive (Duckworth, 1987; Osborne & Gilbert, 1980). A vast number of papers have been published revealing student misconceptions in a variety of domains (Pfunt & Duit, 1994).

Students do not quickly or easily change their naïve scientific conceptions, even when confronted with physical phenomena. However, given enough time and the proper impetus, students can revisit and rethink their ideas. This change can often be painstakingly slow, taking much longer than many teachers allow in the rapid conveyance of content that is the hallmark of a curriculum that focuses on broad coverage. In the view of Eylon and Linn (1988), “in-depth coverage can elaborate incomplete ideas, provide enough cues to encourage selection of different views of a phenomenon, or establish a well-understood alternative” (p. 263). Focused class time is required for teachers to probe for the levels of understanding attained by students and for students to elucidate their preconceptions. It takes focused time for students to test their ideas and find them wanting, motivating them to reconstruct their knowledge. Eylon and Linn note, “Furthermore, students may hold onto these ideas because they are well established and reasonably effective, not because they are concrete. For example, naïve notions of mechanics are generally appropriate for dealing with a friction-filled world. In addition, students appear quite able to think abstractly as long as they have appropriate science topic knowledge, even if they are young, and even if their ideas are incomplete. Thus, deep coverage of a topic may elicit abstract reasoning” (p. 290).

The difficulty of promoting conceptual growth is well documented in neo-Piagetian models (Case, 1998; Fischer & Bidell, 2006; Parziale & Fischer, 1998; Schwartz & Sadler, 2007; Siegler, 1998). These researchers argue that more complex representations, such as those prized in the sciences, require multiple opportunities for construction in addition to multiple experiences in which those ideas and observations may be coordinated into richer, more complex understandings of their world. This process is necessarily intensive and time consuming because the growth of understanding is a nonlinear process that is sensitive to changes in context, emotions, and degree of practice. The experience of developing richer understandings is similar to learning to juggle (Schwartz & Fischer, 2004). Learning to juggle three balls at home does not guarantee you can juggle the same three balls in front of an audience. Alternatively, learning to juggle three balls does not mean that you can juggle three raw eggs. Changes in context and variables lead to differences in the relationships that students have with ideas (both new and old) and their ability to maintain them over time (Fischer & Pipp, 1984; Fischer, Bullock, Rotenberg, & Raya, 1993). The process of integrating many variables that scientists consider in scientific models is a foreign and difficult assignment for anyone outside the immediate field of research.

As just noted, the realization in the cognitive sciences that the learning process is not linear is an important pedagogical insight. Because learning is a dynamic operation that depends on numerous variables, leaving a topic too soon deprives students and teachers the time to experience situations that allow students to confront personal understandings and connections and to evaluate the usefulness of scientific models within their personal models.

Implications in Light of High-Stakes Testing

We are currently in the “Age of Accountability” in U.S. education. The hallmark of this era is high-stakes standardized testing. The nature and consequences of these examinations influence the pedagogy that teachers use in the classroom. As Li et al. (2006) observed: “Academic and policy debates seem somewhat remote to practitioners on the frontline of science education. Science teachers, department heads, and instructional specialists need to survive and thrive in a teaching environment increasingly driven by standards and measured by accountability tests. They are the ones who must, here and now, find solutions to the pressing problems of standards-based reform” (p. 5).

Our concern stems from the connection between the extent that high-stakes state examinations focus on the “facts” of science, secondary school science teachers will focus their pedagogy on ensuring that students can recite these "facts." Clearly, high-stakes examinations that require recall of unrelated bits of scientific knowledge in the form of facts and isolated constructs will increase the likelihood that teachers will adjust their teaching methodologies to address these objectives. The more often discrete facts appear on high-stakes examinations, the more often we imagine that teachers will feel pressured to focus on breadth of knowledge. Conversely, we feel that the adoption of a different approach in high-stakes testing that is less focused on the recall of wide-ranging facts but is, instead, focused on a few widely accepted key conceptual understandings will result in a shift of pedagogy. In such situations, we envision teachers' instructional practice and curricula designed to offer students greater opportunity to explore the various contexts from which conceptual understandings emerge. We suspect that the additional focused time will allow students to recognize (and teachers to challenge) students' naïve conceptions of nature.

These concerns fall in line with Anderson's (2004, p. 1) three conclusions about standards-based reform in the United States:

- The reform agenda is more ambitious than our current resources and infrastructure will support.
- The standards advocate strategies that may not reduce achievement gaps among different groups of students.
- There are too many standards, more than students can learn with understanding in the time we have to teach science.

Anderson's (2004) final conclusion strongly resonates with that of the NRC (2007) and more specifically with findings, at the international level, by Schmidt et al. (1997, 2005). In a comparison of 46 countries, Schmidt et al. (2005) noted that in top-achieving countries, the science frameworks cover far fewer topics than in the United States, and that students from these countries perform significantly better than students in the United States. They conclude that U.S. standards are not likely to create a framework that develops and supports understanding or "coherence," a strategy that encourages the development of a deeper understanding of the structure of the discipline. By international standards, the U.S. science framework is "unfocused, repetitive, and undemanding" (p. 532).

From the perspective of our study, both Schmidt et al. (2005) and Anderson's (2004) conclusions are particularly relevant. Clearly, increasing the quantity of information required for examination preparation will lead to an increased focus on broader coverage of course content, an approach that our findings indicate is negatively (or certainly not positively) associated with performance in subsequent science courses. If teachers know that they and their students are accountable for more material, then the pressure to focus on breadth would seem like the natural pedagogical choice to make. Hence, teachers must decide whether they choose to maximize students' test scores or maximize preparation for success in future study. Although they have considerable feedback concerning the test scores of their students (from state tests, SATs, ACTs, and AP examinations), they have almost no knowledge of how the bulk of their students do in college science, save the few who keep in touch. Moreover, the rare college student who reports back to their high school science teacher is often the most successful and simply reinforces a teacher's confidence in his or her current methods.

Caveats

There are several concerns we wish to clearly elucidate. A potential criticism of this study stems from a central element of this analysis, the length of time spent on a topic and how this variable might be influenced by the ability level of students to grasp the material. One might imagine classrooms containing many struggling students actually spending more time on particular topics and classes with generally high achieving students requiring less time to grasp the concepts and quickly moving on to other topics. In this scenario, depth of content coverage would be an indicator for remedial assistance by the teacher at the classroom level. However, we assume that students in our sample—those who went on to take introductory science courses—are typically well above average in their science abilities. Thus we would expect our depth measure, if it really signified a remedial scenario, to have a negative impact, if any, on student performance in introductory college science courses. As it was, we observed the opposite (i.e., depth was associated with higher performance, even when controlled for student achievement in high school). This argument suggests that the remedial depth scenario is unlikely.

Another issue in this analysis is the degree to which we might expect other variables in the FICSS survey to correlate with depth and breadth as defined. We identified one such item in our survey: "How would you best describe learning the material required in your [high school science] course?" The respondents were provided with a 5-point rating scale ranging from "A lot of memorization of facts" to "A full understanding of topics." This question is constructed in a manner that clearly alludes (at one end of the scale) to a type of memorization that presumes a cursory or superficial level of understanding. This was certainly our intention when we wrote this question. However, Ramsden (2003) notes that the act of memorization may help learners deconstruct the structure and connections in the body of information they are attempting to commit to memory. Thus, is memorization a feature of both depth and breadth in learning? On the basis of this more carefully considered characterization of memorization and taking into account the tone and clear intention of this questionnaire item, we hypothesized that the responses to this item would have a weak positive correlation, if any, with depth, and a weak negative correlation, if any, with breadth. In the analysis, we found the correlation with depth was $r = .15$ (very weak) and the correlation with breadth was $r = -.09$, a result commonly considered "not correlated." The results in this case suggest that memorization (as a new variable) is orthogonal with our definition of breadth and depth and is not subsumed by depth or breadth. However, it remains to be seen how well our definitions hold up as new variables are identified.

Further Research: Looking at Individual Subject Areas

A key issue for further research in the analysis of the constructs of breadth versus depth is whether there are specific topics that high school teachers should focus on or whether high school teachers might choose to focus on any topic and still potentially reap an advantage in terms of students' enhanced future performance. In other words, do students perform better in introductory college science courses if they were exposed to at least one science topic in depth, regardless of what it was, or does it help them if they specifically studied topic "A" in depth, but not if they studied topic "B" in depth? An analogous argument can be made regarding breadth. It therefore appears useful to examine the potential effects of depth in individual subject areas, and of the omission of individual subject areas, in future research.

CONCLUSION

The baseline model reveals a direct and compelling outcome: teaching for depth is associated with improvements in later performance. Of course, there is much to consider in evaluating the implications of such an analysis. There are a number of questions about this simple conclusion that naturally emerge. For example, how much depth works best? What is the optimal manner to operationalize the impact of depth-based learning? Do specific contexts (such as type of student, teacher, or school) moderate the impact of depth? The answers to these questions certainly suggest that a more nuanced view should be sought. Nonetheless, this analysis appears to indicate that a robust positive association exists between high school science teaching that provides depth in at least one topic and better performances in introductory postsecondary science courses.

Our results also clearly suggest that breadth-based learning, as commonly applied in high school classrooms, does not appear to offer students any advantage when they enroll in introductory college science courses, although it may contribute to higher scores on standardized tests. However, the intuitive appeal of broadly surveying a discipline in an introductory high school course cannot be overlooked. There might be benefits to such a pedagogy that become apparent when using measures that we did not explore. The results regarding breadth were less compelling because in only one of the three disciplines were the results significant in our full model. On the other hand, we observed no positive effects at all. As it stands, our findings at least suggest that aiming for breadth in content coverage should be avoided, as we found no evidence to support such an approach.

The authors thank the people who made this large research project possible: Janice M. Earle, Finbarr C. Sloane, and Larry E. Suter of the National Science Foundation for their insight and support; James H. Wandersee, Joel J. Mintzes, Lillian C. McDermott, Eric Mazur, Dudley R. Herschbach, Brian Alters, and Jason Wiles of the FICCS Advisory Board for their guidance; and Nancy Cianchetta, Susan Matthews, Dan Record, and Tim Reed of our High School Advisory Board for their time and wisdom. This research has resulted from the tireless efforts of many on our research team: Michael Filisky, Hal Coyle, Cynthia Crockett, Bruce Ward, Judith Peritz, Annette Trenga, Freeman Deutsch, Nancy Cook, Zahra Hazari, and Jamie Miller. Matthew H. Schneps, Nancy Finkelstein, Alex Griswold, Tobias McElheny, Yael Bowman, and Alexia Prichard of our Science Media Group constructed of our dissemination website (www.ficss.org). We also appreciate advice and interest from several colleagues in the field: Michael Neuschatz of the American Institute of Physics, William Lichten of Yale University, Trevor Packer of the College Board, Saul Geiser of the University of California, Paul Hickman of Northeastern University, William Fitzsimmons, Marlyn McGrath Lewis, Georgene Herschbach, and Rory Browne of Harvard University, and Kristen Klopfenstein of Texas Christian University. We are indebted to the professors at universities and colleges nationwide who felt that this project was worth contributing a piece of their valuable class to administer our surveys and their students' willingness to answer our questions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the U.S. Department of Education, or the National Institutes of Health.

REFERENCES

American Association for the Advancement of Science. (1989). Project 2061: Science for all Americans. Washington, DC: Author.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

Anaya, G. (1999). Accuracy of self-reported test scores. College and University, 75(2), 13–19.

Anderson, C. W. (2004). Science education research, environmental literacy, and our collective future. National Association for Research in Science Teaching. NARST News, 47(2), 1–5.

Anderson, R. D. (1995). Curriculum reform. Phi Delta Kappan, 77(1), 33–36.

Baird, L. (1976). Using self-reports to predict student performance. Research Monograph No. 7. New York: College Entrance Examination Board.

Beittel, K. R., Mattil, E. L., Burgart, H. J., Burkhart, R. C., Kincaid, C., & Stewart, R. (1961). The effect of a "depth" versus a "breadth" method of art instruction at the ninth grade level. Studies in Art Education, 3(1), 75–87.

Bloom, B. (Ed.). (1961). Taxonomy of educational objectives: The classification of educational objectives. Vol. Handbook 1: Cognitive Domain. New York: David McKay Company.

Bradburn, N. M. (2000). Temporal representation and event dating. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), The science of self-report: Implications for research and practice (pp. 49–61). Mahwah, NJ: Lawrence Erlbaum Associates.

Butts, M., & Precott, S. (Eds.). (1990). Science framework for California public schools. Sacramento: California Department of Education.

Case, R. (1998). The development of conceptual structures. In D. Kuhn & R. S. Siegler (Eds.), Handbook of child psychology (Vol. 2). New York: Wiley.

Clark, D., & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. Journal of the Learning Sciences, 12(4), 451–493.

Committee on Science, Engineering, and Public Policy. (2007). Rising above the gathering storm: Energizing and employing America for a brighter economic future. Washington, DC: The National Academies Press.

Dorans, N. J. (1999). Correspondence between Act and SAT I scores (College Board Research Report 99–1). New York: College Board.

Dorans, N. J., Lyu, C. F., Pommerich, M., & Houson, W. M. (1997). Concordance between ACT assessment and re-centered SAT I sum scores. College and University, 73, 24–35.

Duckworth, E. (1987). Having of wonderful ideas and other essays on teaching and learning. New York: Teachers College Press.

Duit, R. (2006). STSCE bibliography. Kiel, Germany: Leibniz Institute for Science Education, University of Kiel. Retrieved April 7, 2006, from http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html.

Eylon, B.-S., & Linn, M. C. (1988). Learning and instruction: An examination of four research perspectives in science education. Review of Educational Research, 58(3), 251–301.

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon & R. M. Learner (Eds.), Handbook of child psychology: Theoretical models of human development (6th ed., Vol. 1, pp. 313–399). New York: Wiley.

Fischer, K. W., Bullock, D., Rotenberg, E. J., & Raya, P. (1993). The dynamics of competence: How context contributes directly to skill. In R. Wozniak & K. W. Fischer (Eds.), Development in context: Acting and thinking in specific environments (pp. 93–117). Hillsdale, NJ: Erlbaum.

Fischer, K. W., & Pipp, S. L. (1984). Process of cognitive development: Optimal level and skill acquisition. In R. J. Sternberg (Ed.), Mechanisms of cognitive development (pp. 45–80). New York: Freeman.

Goforth, T. T., & Dunbar, J., A. (2000). Student response to quantitative aspects of instruction in an introductory geology course. Mathematical Geology, 32(2), 187–202.

Grannot, N., & Parziale, J. (Eds.). (2002). Microdevelopment: Transition processes in development and learning. Cambridge, UK: Cambridge University Press.

Hirsch, E. D. J. (2001). Seeking breadth and depth in the curriculum. Educational Leadership, 59(2), 22–25.

Jackman, W. (1894). Nature study for the common schools. New York: Henry Holt and Company.

James, W. (1890). Principles of psychology. New York: Holt.

Katz, L. G., & Rath, J. (1992). Six dilemmas in teacher education. Journal of Teacher Education, 43(5), 376–385.

Kennedy, K. M., Rodrigue, K. M., & Davis, S. F. (2000). So you want to teach less in hopes of teaching more. College Student Journal, 34(4), 626–634.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. Journal of Research in Science Teaching, 39(6), 522–549.

Kirst, M. W., Anhalt, B., & Marine, R. (1977). Politics of science education standards. The Elementary School Journal, 97(4), 315–328.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. Review of Educational Research, 75(1), 63–82.

Li, J., Klahr, D., & Siler, S. (2006). What lies beneath the science achievement gap: The challenges of aligning science instruction with standards and tests. Science Educator, 15(1), 12.

Linn, M. C., & Hsi, S. (2000). Computers, teachers, peers: Science learning partners. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Murtagh, T. P. (2001). Teaching breadth-first depth-first. Paper presented at the 6th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education, Canterbury, England.

National Center for Educational Statistics. (2006). Highlights from the TIMSS 1999 video study of eighth-grade science teaching (No. NCES 2006-017). Washington, DC: National Center for Educational Statistics, U.S. Department of Education.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, DC: U.S. Government Printing Office.

National Research Council. (1997). Introducing the National Science Education Standards. Washington, DC: The National Academies Press.

National Research Council. (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academy Press.

Newmann, F. M. (1988). Can depth replace coverage in the high school curriculum? Phi Delta Kappan, 69(5), 345–348.

Niemi, R. G., & Smith, J. (2003). The accuracy of students' reports of course taking in the 1994 National Assessment of Educational Progress. Educational Measurement: Issues and Practice, 22(1), 15–21.

Novak, J. D., & Gowin, D. B. (1984). Learning how to learn. New York: Cambridge University Press.

Onosko, J. (1989). Comparing teachers' thinking about promoting students' thinking. Theory and Research in Social Education, 17(3), 174–195.

Osborne, R., & Gilbert, J. (1980). A technique for exploring student's views of the world. Physics Education, 12(6), 376–379.

Pace, C., Barahona, D., & Kaplan, D. (1985). The credibility of student self-reports. Los Angeles, CA: UCLA Center for the Study of Evaluation.

Parziale, J., & Fischer, K. W. (1998). The practical use of skill theory in classrooms. In R. J. Sternberg & W. M. Williams (Eds.), Intelligence, instruction and assessment (pp. 96–110). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perkins, D. (1992). Smart schools: From training memories to educating minds. New York: The Free Press.

Pfunt, H., & Duit, R. (1994). Bibliography: Student's alternative frameworks and science education (4th ed.). Keil, Germany: Institute for Science Education at the University of Keil.

Piaget, J. (1983). Piaget's theory. In P. M. Mussen (Ed.), Handbook of child psychology (Vol. 1, pp. 103–128). New York: Wiley.

Raizen, S. A. (1997). Standards for science education. (Occasional Paper No. 1). Madison, WI: National Institute of Science Education.

Ramsden, P. (2003). Learning to teach in higher education (2nd ed.). London: Routledge Falmer.

Raudenbush, S., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. Journal of Research in Science Teaching, 35(3), 265–296.

Sadler, P. M., & Tai, R. H. (2001). Success in introductory college physics: The role of high school preparation. Science Education, 85, 111–136.

Sadler, P. M., & Tai, R. H. (2007). The two high-school pillars supporting college science. Science, 317(5837), 457–458.

Sawyer, R., Laing, J., & Houston, M. (1988). Accuracy of self-reported high school courses and grades of college-bound students. ACT research Report Series 88-1. Iowa City, IA: American College Testing Program.

Schmidt, W. H., Hsing, C. W., & McKnight, C. C. (2005). Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. Journal of Curriculum Studies, 37(5), 525–559.

Schmidt, W. H., McKnight, C., & Raizen, S. (Eds.). (1997). A splintered vision: An investigation of U.S. science and mathematics education (Vol. 1). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Schwartz, M. S., & Fischer, K. W. (2004). Building general knowledge and skill: Cognition and microdevelopment in science learning. In A. Demetriou & A. Raftopoulos (Eds.), Emergence and transformation in the mind: Modeling and measuring cognitive change (pp. 157–185). Cambridge, UK: Cambridge University Press.

Schwartz, M. S., Hazari, Z., & Sadler, P. M. (2008). Divergent views: Teacher and professor perceptions about pre-college factors that influence college science success. Science Educator, 17(1), 18–35.

Schwartz, M. S., & Sadler, P. M. (2007). Empowerment in science curriculum development: A microdevelopmental approach. International Journal of Science Education, 29(8), 987–1017.

Shubert, W. H. (1986). Curriculum: Perspective, paradigm, and possibility. New York: Macmillan Publishing Company.

Siegler, R. S. (1998). Children's thinking (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis. New York: Oxford University Press.

Sizer, T. R. (1984). Horace's compromise: The dilemma of the American high school. Boston: Houghton Mifflin.

Statler, T. S. (1997). Throw the book at 'em. Mercury, 26(5), 16–22.

Tai, R. H., Sadler, P. M., & Mintzes, J. J. (2006). Factors influencing college science success. Journal of College Science Teaching, 35(8), 56–60.

Van Driel, J., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. Journal of Research in Science Teaching, 38(2), 137–158.

VanSledright, B. A. (1997). Can more be less? The depth-breadth dilemma in teaching American history. Social Education, 61(1), 38–41.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Trans.). Cambridge, MA: Harvard University Press.

Westbury, I. (1973). Conventional classrooms. "open" classrooms, and the technology of teaching. Journal of Curriculum Studies, 5, 99–121.

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

Whitehead, A. N. (1929). The aims of education and other essays. New York: The Free Press.

Wineburg, S. (1997). Beyond breadth and depth: Subject matter knowledge and assessment. Theory Into Practice, 36(4), 255–261.

Wood, W. B. (2002). Advanced high school biology in an era of rapid change: A summary of the biology panel report from the NRC committee on programs for advanced study of mathematics and science in American schools. Cell Biology Education, 1, 123–127.

Wright, P. (2000). Balancing breadth versus depth—An impossible task. Studies in Science Education, 35, 187–188.