Chapter 1

What Is Undergraduate Research, and Why Does It Matter?

CONDUCTING RESEARCH IS an important culminating experience in the education of many undergraduate science students in the United States. This book describes the outcomes of undergraduate research (UR) experiences, the processes by which these outcomes are achieved, and the meaning of these outcomes for both students and the mentors who work with them on scientific research projects, based on our findings from a multi-year study of undergraduate research and its role in science education. An overarching theme in these findings is the notion of "real science," which recurs throughout the comments of undergraduate research students and their advisors. Their work together on scientific research projects provides the experiences and observations that form the backbone of this book. The importance of "real science" for students' educational and professional growth is evident in their own words:

> *It's kind of scary, especially at the beginning. I was like, "How can someone like me be doing this?" [But now] I'm coming up with valuable information and it's great. I mean, actually producing data and actually doing it, I felt like a scientist. But you really feel more like a scientist when you have something good! (female UR student, biology)*

• • •

> *Once your superiors—whom you admire and look up to as scientists—start asking your opinion on a scientific matter. . . . Personally, it made me feel like I was actually a real physicist. (male UR alumnus, physics)*

• • •

> *Presenting at a conference made me feel like I was a part of the scientific community. . . . I have been able to talk about my work and feel like an equal [with my advisor], and do it with other people [at my school]—but being able to do that with a total stranger was a really, really neat experience. It gave me a lot of confidence and made me feel like I was a real chemist! (female UR alumnus, chemistry)*

• • •

> *A lot of things you do in school, like you do homework or whatever, and you never feel like you're really doing something real. And this was one of the first things that I did that, like, really encompassed everything and really brought things together. It was one of the first times I really felt like I was really doing something. (male UR student, engineering)*

Clearly, being "real" is important to students. So what makes a research project "real"? As we will show, real research is an investigation whose questions, methods, and everyday ways of working are authentic to the field. The research questions are well defined so that they can be systematically investigated, but, importantly—and unlike most questions in a classroom—their answers are unknown. Research results may not be quickly forthcoming, but they constitute a genuine contribution to the field if and when they do emerge. The research methods are ones used in the discipline and seen as valid by disciplinary experts. As in any other research project, the choice of methods may be constrained by intellectual, technical, or financial resources. For an undergraduate research project, such constraints may arise from the involvement of novices and the educational mission of their institution—but the term *undergraduate research* does not inherently rule out particular approaches to the research question. Perhaps most important, as we shall see throughout this book, students and faculty work together in ways that are typical of their field and authentic to the profession. Thus, students learn the intellectual and social practices of science by doing it. By engaging deeply themselves in a particular question, they begin to understand more generally how scientists engage questions and construct knowledge, and that this is a human activity in which they too could participate.

As Merkel (2001) points out, the use of the term *undergraduate research* has not always been clear—indeed, the term *research* itself has different meanings in different disciplines and settings. The Council on Undergraduate Research (CUR, n.d) offers a broad-based definition: "An inquiry or investigation conducted by an undergraduate student that makes an original intellectual or creative contribution to the discipline" (see also

Wenzel, 2003). This language is inclusive of CUR's multidisciplinary audience, but in its lack of mention of faculty guidance or mentoring, it does not fully describe UR as typically practiced in the sciences. As we shall describe, the research advisor's role is critical in guiding students' work and inducting them into the intellectual and social ways of the profession. The way that UR advisors work with students parallels the master-apprentice relationship that is traditional in many professions, including graduate education in science.

A note is in order to clarify our choice of language. Throughout this book, we commonly use the terms *science*, *scientist*, and *scientific* with the intent to include psychology, mathematics, and engineering, at least with respect to UR in these fields. The acronym STEM, standing for science, technology, engineering, and mathematics, is also used, but this acronym is sometimes inelegant and comes with neither a corresponding adjective nor a term for the individuals who practice it. The studies that we discuss in the bulk of the book involve mainly students and faculty in the natural sciences, but they also include mathematicians, engineers, computer scientists, and psychologists. Our intent is to be fully inclusive while avoiding unwieldy language.

In this book, we restrict our discussion to intensive, multiweek research experiences in the sciences, mathematics, and engineering that involve student collaboration with faculty or other experienced scientists, and we refer to this as the *apprenticeship model*. Moreover, we argue that the goals and practices of apprentice-model UR are shaped and sustained by its value as both an educational activity for students and a scholarly activity for their research advisors. Because course-based inquiry is generally driven by educational concerns only, we intentionally exclude it from our definition of undergraduate research. Although course-based inquiry is important and still too uncommon in undergraduate STEM education, it should not be conflated with the apprenticeship model of undergraduate research, for reasons that we hope become apparent in this book.

Undergraduate research is widely conducted in the sciences, led by faculty at primarily undergraduate institutions (PUIs) across the United States. At research universities too, faculty whose laboratories include graduate students, postdoctoral researchers, and technicians often also host undergraduate researchers. We use the term *faculty-led UR* to refer to all such research experiences that are largely initiated and directed by faculty themselves and hosted by individual research groups, with modest or no coordination at the departmental or institutional level.

More recently, universities and government laboratories have sponsored structured research programs, sometimes with the goal of recruiting students from groups that are nationally underrepresented in fields. We call these *structured UR programs* because they often include UR along with organized training, presentation and professional development activities, and other kinds of academic and financial support. Many involve a particular cohort of students who enter the program together and participate for longer than just one summer. While some practices differ in these varied contexts and, to a lesser extent, by discipline, UR experiences in the United States appear to have in common several features:

- A well-defined research project designated to the student or a student team, connected in some way to an ongoing effort in the research group or to an area of scholarly interest of the supervising researcher
- Multiweek immersion—often full time for ten weeks during the summer, though UR may also be carried out through the academic year
- Individualized guidance from an experienced scientist

There is growing interest in earlier entry to UR, but at this time, most students participate in UR as college juniors, seniors, or rising seniors in the summer between the junior and senior years (American Society for Biochemistry and Molecular Biology, 2008; Russell, 2006).

History of Undergraduate Research

The idea that undergraduates should conduct real investigations is not new. The California Institute of Technology traces the origins of its undergraduate research program to Arthur Noyes's tenure as chemistry department chair beginning in 1920, touting an early publication by two students who later became Nobel laureates (McMillan & Pauling, 1927; Merkel, 2001). A century ago, Drinker (1912) surveyed the practice of UR at undergraduate medical colleges, one of which dated its own UR efforts to 1895. A proponent of UR, Drinker argued that medical students "have a right to gain some notion of what investigation entails," but "the doing of fixed experiments in fixed hours does not entail the exercise of investigative faculties other than those of the most mechanical nature" (p. 730).

Drinker and his survey respondents postulated outcomes of UR little different from those claimed by practitioners today: in doing research, students must bring to bear both "imagination" and "high scientific accuracy" (p. 730). Students learn "the difficulty of putting a problem on a working basis" (p. 730) and experience "an intellectual awakening" (p. 736) that is as valuable to the "practical man" as to the "laboratory man" (p. 732). Respondents presumed that doing research helped to recruit students into the profession of research, but also argued that research-derived critical thinking skills transferred to other fields. "All of us believe in its value," wrote one dean, "otherwise we would discourage it—not, I fancy, for the value of the scientific results obtained, but for its educational value on the picked men and the belief that the group of the serious workers in medical science will be recruited from this body of students" (p. 736). A follow-up report (Starr, Stokes, & West, 1919) indicated that opportunities for undergraduate research had "increased greatly since 1912" (p. 311).

In her review of the history of UR, Merkel (2001, 2003) traces the beginnings of organized UR activities at research universities to MIT's program, started in 1969 (Massachusetts Institute of Technology, 2000, n.d.). At liberal arts colleges, undergraduate research was under way, at least in chemistry departments, by the postwar science boom of the 1940s and 1950s, further spurred in the 1960s by post-*Sputnik* concerns about American competitiveness in science and technology (Bunnett, 1984; Craig, 1999; Neckers, 2000; Trzupek & Knight, 2000; see also Crampton, 2001; Hansch & Smith, 1984; Pladziewicz, 1984). Participants in a 1959 conference on teaching and research debated whether scientific research was an appropriate activity for undergraduate colleges, or instead a cost- and time-intensive distraction from faculty's main work of teaching (Spencer & Yoder, 1981). In the mid-1980s, college presidents met at Oberlin College to draw attention to the success of liberal arts colleges in producing large numbers of science majors who went on to science careers and science Ph.D.s. Prompted by findings such as Spencer and Yoder's (1981) analysis of research activity in chemistry departments at liberal arts colleges and the number of their graduates who earned Ph.D.s in chemistry, the Oberlin report lauded student-faculty collaborative research as a major contributor to strong science education at these schools (Crampton, 2001; Gavin, 2000;). Accounts of UR in this era are consistent in portraying UR as a form of faculty scholarship particular to PUIs, initiated and sustained by individual determination, scrappy grantsmanship, and grassroots networks (in addition to sources cited above, see Doyle, 2000; Mohrig & Wubbels, 1984; Pladziewicz, 1984). Faculty valued research as a means to stay scientifically up to date and connected to their discipline, and thus fresh in the classroom; obtain equipment useful in laboratory courses; and build a positive reputation for their department. They recognized UR's positive side effects for students, but had not claimed them in public until withdrawal of National Science Foundation (NSF) funding for undergraduate science education in 1981 forced them to reconsider how they might finance faculty development, course improvement, and student research activity (Mohrig & Wubbels, 1984; National Science Foundation, n.d.).

As the arguments caught on that UR was not only important as scholarship for faculty at PUIs but also high-quality science education for students, the profile of UR rose among funding agencies and professional organizations. In the mid-1980s, the NSF initiated the Research at Undergraduate Institutions program to support UR through single-investigator grants from the research directorates (Council on Undergraduate Research, 2006). This was followed by the Research Experiences for Undergraduates (REU) program, now in its third decade, which supplies site grants to support undergraduates to work on research (National Science Foundation, n.d.). (Both programs were predated by NSF's Undergraduate Research Program, which made awards between 1971 and 1981.) The Howard Hughes Medical Institute began to award undergraduate science education grants that often supported UR programs, and the American Chemical Society's Petroleum Research Fund, the Camille and Henry Dreyfus Foundation, and Research Corporation all offered research grant programs with tracks targeted to faculty working primarily with undergraduates. CUR was founded by chemists in 1978 as an organization to promote and support student research in PUIs. The National Conference on Undergraduate Research began in 1987 to provide an opportunity for student researchers to present their work, and disciplinary professional societies began to include poster sessions for undergraduate research student presenters as part of their conference programs.

In the 1990s, national reports such as the Boyer Commission report (1998) cited UR as a practice that could contribute to improving undergraduate science education, move students from didactic to inquiry-based learning experiences, and reduce the dichotomy between teaching and research (see Katkin, 2003; Merkel, 2001, 2003). The 1990s also marked the accelerated development of programs to recruit and retain students from underrepresented groups, which often incorporated undergraduate research. If the early decades were the years for grassroots growth of UR, the 1980s the decade of its professionalization among faculty, and the 1990s the decade of recognition by policymakers of UR as an educational practice, then the 2000s appear to begin the era of evaluation and research. After "decades of blind faith" in the benefits of UR (Mervis, 2001a), researchers and evaluators have begun to identify its outcomes, assess their prevalence, and examine how they come about. We review these studies in detail in Chapter Two.

Current National Context for Undergraduate Research

In this book, we examine UR at the local level as an educational experience for students and as an educational and scholarly activity of faculty and departments. However, this local practice takes place in a national context of high interest in UR as an educational strategy, influenced by the traditional role of the research apprenticeship in scientists' education and by growing interest in students' development of thinking skills important for public science literacy.

Scientists, educators, and government and industry leaders have raised concerns over the supply and quality of STEM-trained workers needed to maintain American technological and economic leadership in a globally competitive economy (for a recent high-profile report, see National Research Council, 2007; for a summary of such reports, see Project Kaleidoscope, 2006). Since 1980, the number of nonacademic science and engineering jobs has grown at more than four times the rate of the U.S. labor force as a whole (National Science Board, 2008). Increasing the diversity of the science workforce is another "urgent need," given changing demographics, decreasing numbers of foreign citizens entering the U.S. STEM workforce, and growing international competition for scientific and engineering talent (Committee on Equal Opportunities, 2004). Equally important, concerns for equity and justice demand that all Americans have equal opportunities to enter the high-status, well-paid positions typically offered by science and engineering careers. Economic competitiveness too depends on a diverse workforce, because diversity fosters greater innovation and problem solving (Chubin & Malcom, 2008; Page, 2007). However, at higher levels of STEM education in many fields, the proportion of both women and people of color declines sharply—the so-called leaky pipeline—and progress in bringing their representation up to match the general population has been slow (National Science Foundation, 2007b). Thus, availability and access to high-quality STEM education remain critical for meeting U.S. workforce needs and providing equal opportunity for all citizens.

While multiple solutions to these pressing problems lie throughout the spectrum of K–12 and higher education, many calls for reform have focused on making undergraduate STEM education more practical, relevant, engaging, and grounded in research on how people learn (Bransford, Brown, & Cocking, 1999; Handelsman et al., 2004; Project Kaleidoscope, 2006; Seymour, 2002; Wieman, 2007). For example, the American Association of Colleges and Universities has called for higher education institutions to foster more "empowered, informed, and responsible learners" (Greater Expectations National Panel, 2002). The Boyer Commission (1998) urged that research-based learning become the standard in undergraduate education, particularly at research universities. National bodies have called for increased opportunities for student-centered, inquiry-based learning, including undergraduate research, in the STEM disciplines (Kuh, 2008; National Research Council, 1999; National Science Foundation, 1996). Many faculty and institutions are exploring the addition of "research-like" components to regular courses and labs (see DeHaan, 2005). Although different wording is often used, these efforts in undergraduate STEM education parallel efforts in K–12 education to incorporate scientific inquiry as both a strategy for teaching scientific concepts and an element of the curriculum. The aim is for students to develop not only conceptual understanding of the big ideas of science, but also the abilities to conduct an investigation and the understandings of science as a human process of constructing scientific knowledge (National Research Council, 1996; see also Laursen, 2006).

Undergraduate research is relevant to these national concerns because it is commonly believed to be "invaluable" for "engaging, training and inspiring undergraduates (many from underrepresented groups) to pursue higher . . . degrees" (National Science Foundation, 2007a, p. 10) and to have "central importance" in "preparing scientists" (American Society for Biochemistry and Molecular Biology, 2008, p. 19). UR may be seen as one end of a spectrum of educational strategies that engage students, both a model for and a culmination of classroom-based inquiry (see, for example, Healey & Jenkins, 2009; Karukstis & Elgren, 2007). But there are substantial barriers to pedagogical change in undergraduate teaching, including the high autonomy of college instructors, their primary allegiance to their discipline, student and collegial resistance, and institutional barriers to research-based pedagogical reforms (Boyer Commission, 2002; DeHaan, 2005; Henderson, 2005; Henderson & Dancy, 2008; Kuh, 2008; Seymour, 2007; Walczyk, Ramsay, & Zha, 2007; Wieman, 2007). Thus, UR may be seen by funders, institutional leaders, and faculty developers as a path of lesser resistance to change in undergraduate STEM education than is classroom-focused reform. Indeed, a recent survey of members of a discipline-based scientific society, the American Society for Biochemistry and Molecular Biology (2008), highlights the seeming paradox that although faculty placed high value on "undergraduate research and integrative thinking" (p. 3), their classroom pedagogy was "not reflective of research on student learning" (p. 5)—fully 80 percent of their classes, at all levels, emphasized lecture. Thus, for all these reasons, undergraduate research is often viewed as a solution to national STEM education problems.

Scope of Undergraduate Research

If UR is in fact to aid in solving any of these problems, the numbers of students who participate will have to be substantial. However, that number is difficult to determine. In a survey by SRI International of thirty-four hundred students who received STEM bachelor's degrees between 1998 and 2003, just over half of respondents said they had participated in UR (Russell, 2005). The Boyer Commission (2002) offers the lower estimate that one-fifth of science and engineering students at research universities engage in UR. Results of the National Survey of Student Engagement indicate that 19 percent of all undergraduates participate in research with faculty (Kuh, 2008), including 39 percent of those with majors in the biological and physical sciences (American Council of Learned Societies, 2007). While Kuh's (2008) averages across broad institutional types and student characteristics vary surprisingly little, the participation rate is in fact quite variable from one school to another—higher at many smaller schools where faculty lead UR for their own students and lower where no on-campus opportunities are available. Wood (2003) cites 45 percent participation in UR for his biology department at the University of Colorado, while Merkel (2001) cites figures for student participation in UR of 80 percent at MIT, 60 percent at CalTech, and 22 percent for the University of Washington. Figures like these illustrate how departmental and institutional differences affect students' access to UR, even at schools that have established or are moving toward a "culture of undergraduate research," in Merkel's words. Most institutions do not systematically gather these data for themselves (Katkin, 2003). Participation also varies strongly by discipline; STEM graduates in the SRI survey reported participation rates near 30 percent for mathematics and computer science and up to over 70 percent for chemistry and environmental sciences (Russell, 2005).

These variable participation rates are one reason that it is difficult to tally the total numbers of UR participants. Russell (2006) has estimated that the NSF may support some fourteen thousand students per year, but Merkel (2001) reported thirty-two thousand students supported by NSF REU programs alone in fiscal year 2001. (We requested data from NSF on undergraduate research participation but were unable to obtain either agency-wide or individual division data from those contacted.) Whatever the numbers, it is likely that the number of UR opportunities is not enough to accommodate all students who seek the opportunity. A 2004 study reported that the NSF REU program in chemistry, which then supported about 650 students each year, could accommodate fewer than one in four students who apply (Henry, 2005).

Financial investment in UR by public and private foundations is substantial and supports students through both targeted UR programs and grants to individual investigators at PUIs. Again, numbers indicating the magnitude of this investment are difficult to come by. Academic Excellence, a study of undergraduate research at 136 PUIs, reported a ten-year total (1991–2000) of $682 million in funding for research and research instrumentation at these colleges, with 74 percent coming from federal and state government sources (Research Corporation, 2001). From the cost side, and taking the perspective that the faculty is an institution's primary investment, Gentile (2001) has estimated the projected investment in a faculty member over a thirty-year academic lifetime to be $4 million, including both research- and teaching-related costs. His worksheet enables this figure to be computed for a particular local setting. From a student perspective, funding for NSF REU awards in chemistry for 2009 averaged $10,000 per summer UR student, covering both direct student support and associated program costs (Colon, 2009).

Without good data about the participation level of students and faculty, the resources committed, or their cumulative impact, it is difficult to state whether the prevalence of UR is growing, shrinking, or staying the same. However, most sources agree that UR is on a rising trajectory. The SRI study (Russell, 2005) noted that participation rates in UR had increased from 48 percent among 1988–1992 STEM graduates to 56 percent for 1998–2003 graduates; concurrently, the proportion of respondents who said it had not occurred to them to participate in research declined from 24 percent to 15 percent. The Academic Excellence study found that the number of students engaged in summer research at the 136 PUIs in this study increased by 65 percent in the decade 1991 to 2000 (Research Corporation, 2001). In a follow-up study to the Boyer Report, Katkin (2003) reported that research universities had taken many steps to expand UR opportunities and raise the visibility of UR, often establishing centralized offices to support UR and advertise it to students, promote it in departments, and raise funds. Katkin's data also showed increases in the number and percentage of participating students and the number of faculty UR supervisors. However, the lack of systematic data collection by institutions is a problem: as Kenny (2003) points out, "A lot may be happening, but no one is charged with keeping score" (p. 105).

Several indicators reflect growing interest in UR by funding agencies. For example, NSF's Division of Chemistry has experimented with undergraduate research centers to explore novel forms of UR that might engage students at an earlier stage or from previously untapped populations, including UR at two-year colleges and curricular forms of research activity (*Exploring the Concept*, 2003). The National Aeronautics and Space Administration and NSF have supported "extreme research" opportunities for students, such as the chance to conduct engineering experiments in the weightless environment of the "Vomit Comet" research aircraft, use international telescopes at distant observatories, or make geoscience field observations from oceangoing research vessels, Iceland, or the South Pole (Service, 2002). Several private foundations that support undergraduate research signaled their interest in UR by commissioning the *Academic Excellence* study to address their concerns about declining research proposal pressure from these PUIs (Lichter, 2000; Mervis, 2001b; Research Corporation, 2001). Despite the foundations' observations, the study found that overall, the sciences were healthy at these schools, which educate a disproportionate share of the nation's scientific workforce. Research-related grant dollars awarded to these schools had increased, as had colleges' investment in faculty start-up funds and capital facilities for science (Abraham, 2001).

Another indicator of growing interest in UR is a proliferation of how-to resources that seek to help those initiating UR at an ever-widening group of institutions. The *CUR Quarterly* and the *Journal of Chemical Education* offer long-running article series. Books by Merkel and Baker (2002) and by Handelsman and colleagues (2005) offer advice on mentoring UR students (see also Pfund, Pribbenow, Branchaw, Lauffer, & Handelsman, 2006), while Hakim (2000) discusses the institutional development and implementation of UR programs. CUR recently compiled a compendium of practices to develop and sustain a "research-supportive curriculum" (Karukstis & Elgren, 2007). Kinkead (2003) has reviewed resources on UR programs and inquiry-based teaching approaches that support them. Gaglione (2005) and Brown (2006) offer advice to two-year college faculty on starting a UR program, and Ball and coauthors (2004) do the same for those at comprehensive institutions (see also Husic, 2003). While interest is growing in UR and other forms of scholarly and creative activity in disciplines beyond STEM (Karukstis & Elgren, 2007; Katkin, 2003; Merkel, 2003), most non-STEM fields do not yet have well-established UR traditions. Similarly, international interest in UR is growing in countries that do not currently have a UR tradition.

The niche of how-to resources for students is also increasingly occupied. WebGURU is an online clearinghouse for students with practical information on how to seek an undergraduate research position and what to do once they get one. At its Web site, CUR maintains a list of the growing number of online undergraduate research journals, which provide opportunities for students to publish their work and learn the skills of professional writing and peer review (Netwatch, 1998).

Finally, there is grassroots evidence that UR is gaining popularity among students. Some campuses document rising participation in UR by their own students (see, for example, Bhushan, 2007; Biggs, 2006; Singngam, 2007). Katkin (2003) observes an increase in the number and visibility of centralized UR offices on campuses to serve growing student demand. These offices typically advertise research opportunities to enrolled students and facilitate students' matchup with advisors, projects, and funding. As part of its much-publicized annual college rankings used by prospective students and families planning for college, *U.S. News and World Report* spotlights schools with strong undergraduate research programs. The growing popularity of UR is even captured in pop culture. As of early 2010, over fifteen hundred YouTube videos bore the tag "undergraduate research." These online videos enable students to share their UR experiences and institutions to market UR to prospective students as a distinctive educational experience.

Together these indicators suggest that interest in UR is increasing among many stakeholders. This trend is positive insofar as it leads to opportunities to provide the educational and professional benefits of UR to larger numbers of more diverse students and encourages faculty to integrate their teaching and research work, as some have argued (see Prince, Felder, & Brent, 2007, for a thorough review of this literature). However, this trend can also have negative impacts, introducing new political and financial pressures for institutions that have not considered research part of their educational mission (Husic, 2003) and placing additional strain on individual faculty who may already be stretched to their limits with teaching, service, and other scholarly work (Tobochnik, 2001), as we discuss in Chapter Nine.

Studying UR: The Nature of the Evidence Presented

The high participation and investment in UR signify widespread belief in the value of UR for students' educational and career development. However, only recently have researchers and evaluators begun to establish an evidence-based understanding of the character and range of benefits to students, faculty, or institutions that are generated by different types of UR experiences. Traditional institutional outcome measures, such as the fraction of UR students who later pursue a Ph.D. in science, do not reflect the potential value of UR as an educational and personal growth experience for students. Although such data are widely cited, few data exist to justify any causal connection between UR and career outcomes. This lack of knowledge about the educational outcomes of UR, and their meaning for students, faculty, and institutions, was the impetus for our work.

The findings presented in this book are based on two studies that together involved nearly six hundred interviews, conducted over five years, with students, their research advisors, other mentors, UR program directors, administrators, and staff. Most of the chapters are based on data from a large interview study, comprising over 360 interviews, of faculty-led summer UR as conducted at four colleges (thus, "the four-college study"). That study was developed to examine fundamental research questions about the nature of UR and its role in undergraduate science education:

- What are the benefits to students of conducting UR—both immediate and longer term, and as viewed by both students and their research advisors?
- What, if anything, is lost by students who do not participate in UR?
- What are the processes by which gains to students are generated?
- What are the benefits and costs to faculty from their own engagement as UR advisors?

In addition to the four-college study, we draw on findings from a program evaluation of a structured summer UR program that was conducted at a research laboratory for students from groups underrepresented in the sciences. Although that study evaluated a particular program to provide feedback to its developers, it serves here as a case exemplar of structured programs targeted to minority students. That study, based on over two hundred interviews, is the focus of Chapter Six. We also draw on findings from evaluation studies of structured UR programs at two research universities (Hunter, Thiry, & Crane, 2009; Thiry & Laursen, 2009).

In this section, we describe the design of the four-college study and the process by which data were gathered and analyzed, so that readers will understand the nature of the evidence presented. Further details of the study methods and interview samples are given in Appendixes A, B, and C. Our group's other studies that are mentioned were also interview studies that followed similar methods to those described.

Selecting a Model of UR: The Study Sites

We chose to study UR in a best-case scenario represented by the summer apprenticeship model. Our sites were four undergraduate liberal arts colleges: Grinnell College (Iowa), Harvey Mudd College (California), Hope College (Michigan), and Wellesley College (Massachusetts). Because these schools do not have graduate programs in the sciences, their faculty have uniformly committed to teaching and mentoring undergraduates. Many departments had a long tradition of summer UR and hired science faculty with the expectation that they would involve students in scholarly work. Full-time summer UR work was the model chosen for study because it is a common and widely practiced model, less variable in structure and implementation than academic-year UR activities, and because this intensive form was expected to provide the strongest and most distinctive benefits to students that might most easily be attributed to UR or other educational experiences. These choices do not mean that our findings apply only to these settings; on the contrary, we have evidence that well-implemented UR in other forms can achieve the same results. However, the relatively homogeneous model of summer UR at these colleges enabled us to define with clarity the phenomenon under study and to attribute student outcomes to that phenomenon. Thus, this study addresses the question of what is possible from well-designed, well-implemented, apprentice-model UR. Studies examining other research questions—including the outcomes of other UR models, comparison of outcomes across different UR models, and characterization of the wide range of activities that have been labeled "undergraduate research"—are still needed.

Our four-college interview samples included essentially all rising seniors who were participating in UR in summer 2000 and their faculty research advisors; a few were visiting students from other campuses. Comparison samples were developed of nonparticipating faculty and of nonparticipating student majors from the same class year as the UR students. As we discovered, colleges did not often track UR participation, and departments often did not know that students they had identified as not participating in UR had in fact pursued UR opportunities off campus.

Although the number and organization of departments varied across these campuses, all STEM departments that had summer UR students were included. The study thus spans the natural sciences and includes psychology, mathematics, computer science, and engineering in schools where these majors were offered and faculty in these disciplines also conducted UR. We make no claims about the applicability of specific findings beyond these fields, although we believe that our general emphasis on authenticity offers lessons relevant to other disciplines. The largest numbers of interviewees came from biology and chemistry, and smaller numbers from physics, mathematics, computer science, psychology, and engineering.

Gathering Multiple Perspectives on UR: The Study Samples

The four-college study examined UR from the distinctive perspectives of several groups: students who conducted research and students who did not, faculty advisors of UR and faculty who did not participate, and some administrators whose roles included institutional oversight of UR. The seventy-six UR students were science majors who as rising seniors—about to enter their senior year—participated in summer research in their discipline. Students from this group were interviewed three times: near the end or soon after the summer UR experience, at the end of their senior year, and about two years after graduation. These interim interviews with UR students did not produce notably different findings from the first interviews and are not discussed in any detail in this book.

A group of sixty-two comparison students was interviewed to investigate whether the gains from UR were unique or could be achieved through other educational experiences. This group came from the same departments and the same graduating class as the UR student sample, but included students who were not participating in UR in summer 2000 for a variety of reasons: some chose not to pursue UR, some applied but did not get a summer UR position on their campus, and some undertook other forms of UR as seniors during the regular academic year. These students were interviewed twice: at the end of their senior year, in order to allow for gains, if any, to emerge from their entire undergraduate experience; and again about two years later. These students pursued a variety of internships, work, senior theses, and off-campus UR, in addition to the classroom and campus experiences they shared with the UR students. Thus they cannot be viewed as an idealized control group—which is seldom available in any educational research. Rather, they serve as a comparison group whose range of experiences realistically reflects the rich array of undergraduate experiences that science students may undertake and to which UR may be compared. As we discuss in the chapters, we can detect but not control for incoming differences in these students' goals and interests, as well as in their outcomes. For both UR students and comparison students, the numbers of interviewees in the later rounds declined, as not all alumni could be reached or chose to participate.

To hear the faculty perspective, we interviewed a group of fifty-five faculty who were the research advisors of the UR students we interviewed. Their comments collectively reflect their many years of research experience with students. In addition, we interviewed thirteen faculty who had previously led UR but had temporarily or permanently discontinued their work with student researchers. Like the UR-active faculty, they provided an experienced perspective on UR and offered additional commentary on their reasons for discontinuing UR work. Interviews with twelve administrators—department chairs, deans, provosts, and UR program directors, all of whom also were or had been active UR advisors—provided information on UR in a broader institutional context.

In each chapter, we note the interview groups that offered the evidence we present. We consistently use the following terminology: *UR student* refers to a participant in the first interviews with student researchers and *comparison student* to a participant in the first interviews with the comparison group. Where relevant, observations made by comparison students are also identified according to their other educational experiences, such as internships or courses. *UR alumni* and *comparison alumni* refer to individuals from the two student samples interviewed again two to three years after their college graduation. *Research advisors* are all faculty and administrators, active or inactive, who had supervised UR students. We use this term when referring to their UR role, and the term *faculty* when discussing their other functions within their colleges, such as teaching and collegial interactions. Our choice of the term *advisor* is deliberately inclusive: at these four colleges, research advisors were faculty, but UR conducted at universities may involve advisors who are graduate students and postdoctoral researchers; at nonacademic laboratories, advisors may be working research scientists. The term thus emphasizes the functions of UR advisors, not their job title. Full details of the four-college interview samples and their makeup by gender, discipline, and other variables are provided in Appendix A.

Working with Evidence from Interviews: The Study Methods

To address our major research question about the benefits of UR to students, we began by gathering published studies and descriptive accounts of UR and built from these a checklist of possible benefits to students. These covered a range of possible changes in knowledge, skills, behaviors, and attitudes: scientific knowledge in and beyond their discipline; understanding of the process of science; growth in practical, intellectual, and teamwork skills; changes in confidence and attitude; career preparation; career choice; and others. An earlier report (Seymour, Hunter, Laursen, & DeAntoni, 2004) includes our analysis of this literature.

We then used this checklist (Appendix C) in our interviews, querying students about these possible gains as areas where "faculty think students may gain from doing undergraduate research." Interviews with UR students tended to focus on the UR experience itself, but we checked with students about whether the gains they described came from UR or other sources. Similar language, without reference to research, was used to probe the same gains among comparison students, whose accounts of the sources of particular gains are discerning and informative (Chapter Four).

In responding to the gains checklist—or in spontaneous narratives elsewhere in the interview—students described benefits they made, did not make, or made to some extent. Beginning with the UR student interviews, we coded the interview transcripts for these gains (and for other concepts) in detail, including gains that matched the checklist as well as additional gains that students raised. Coding interview data is a painstaking process: the coder reads the written transcript, tagging each separate idea raised with a code. When the same idea is raised again in a later comment or by another person, the same tag is reused; as new ideas are raised, each is given a distinct code. Each code may tag one or several distinct observations. Codes also record whether the gain was positive, negative, or mixed: gained, not gained, or partially gained. When the coding is complete, the set of codes or tags—known as the codebook—reflects the overall content of the data set. The codes are then sorted and categorized under broader labels, called *parent codes*, and the analyst searches for patterns in these parent codes, the frequency with which they are used, and linkages between codes and speakers, such as by gender or role. Some patterns may be noticed and explained by interviewees themselves, while others emerge from the data and are discerned by the analyst without interviewees' explicit awareness.

We sometimes liken the analysis process to disassembling necklaces and organizing the loose beads. In this analogy, each interview is a necklace, a string of many individual observations by a speaker. Coding labels each bead in detail: a round red plastic bead, a shiny yellow oblong bead. Analysis thus resembles sorting beads into jars: red or yellow, shiny or not; beads from short necklaces or long ones. Using powerful text analysis software packages to stand in for jars, codes can be sorted simultaneously across multiple dimensions.

We can then describe the jars—the codes, parents, and broader domains that together constitute the qualitative content of the interviews. We can also count the beads in each jar or set of jars. Counting observations is one way to estimate the relative weight of opinion about a set of topics or identify differences in the weight of opinion among different interview groups. We count conservatively to avoid overrepresenting any views, such as when a single speaker makes the same point several times. Most often, we report the number of observations, which far exceeds the number of people, as each person may make multiple comments on any given topic. Sometimes we report the number of speakers to indicate the occurrence of a particular phenomenon or view. These frequency counts are often informative, but interview data cannot be treated by statistical techniques as can responses to standardized survey items. Throughout this book, we refer to frequency counts in discussing the relative importance of ideas, but we remain mindful of the advice apocryphally posted on Einstein's office wall: "Not everything that can be counted counts, and not everything that counts can be counted." (We thank Richard Donohue for pointing out the appropriateness of this quotation to our work.) Appendix B includes a detailed discussion of our coding and analysis methods, including treatment of the frequency counts.

Trusting the Evidence: Validity and Reliability

We are often asked how we know that our data can be "trusted." In interview studies, as in all good scientific work, this depends on both gathering good evidence and interpreting it with an open and skeptical mind. First, our interviewees had little reason to dissemble to us as external researchers. Indeed, we were often struck by the candor and emotion with which they spoke, discussing failures or conflicts as well as successes and rewards. Speakers participated voluntarily and gave informed consent; their anonymity and confidentiality were protected by the ethical and professional standards of our field and formalized by human subjects research review. Thus, speakers could know that personal information would not be shared. We were careful to phrase questions neutrally so as not to lead respondents to an answer, for example, asking about "what gains you did or did not make." By conducting interviews with essentially all summer student researchers, their advisors, and nearly all nonparticipants we could identify in the same departments, we avoided biasing our samples.

We also built checks and balances into our data analysis. That the same ideas emerged over and over from so many separate interviews, where collusion was impossible, is one powerful indicator of their validity. Research team members continually discussed the work, reviewing the coding to define and refine categories and ensure that we agreed in how we coded similar information, and arguing about the meaning of our observations. Triangulation, or comparing different sources of information, such as from students and advisors, helped to refine the analytical themes and detect similarities and differences in perspective. When interviewing alumni, we conducted "member checks" to validate our findings by sharing them with alumni who could then agree, disagree, or offer commentary. Our multidisciplinary research team brought a variety of personal and professional perspectives to the project, including for one of us (Laursen) experiences as both a UR student and a faculty advisor of UR. While we distinguish evidence from interpretation as we do the work itself, for readability we have presented our findings in a form that does not sharply distinguish them.

Drawing Conclusions: Limitations and Strengths of the Four-College Study

The study examines UR as practiced at four highly ranked liberal arts colleges. These colleges attract a very capable student body that is largely middle class, white, and academically well prepared. The faculty are teaching oriented; they conduct research with students as a primary scholarly and educational activity, not as a satellite of graduate-level research. These choices allowed us to define a particular model of UR whose outcomes could be investigated and examine some local variations of that model. However, they also place some constraints on the extent to which the findings may be generalized to other forms of UR and other settings, and they leave some questions unanswered.

However, this choice does not mean that the findings are idiosyncratic. This study is very large for a qualitative study; the data were gathered from many research groups in several departments on four campuses. While some features of the organizational and cultural context of these sites are likely particular to liberal arts colleges, we shall argue that many of our findings are not particular to these settings. We offer evidence that both student outcomes and research advisors' strategies for working with UR students are not unique to these settings, and we have seen both in research universities and national labs. We also see little evidence of significant variations in student outcomes or advisor strategies by discipline. In several chapters, we discuss the extent to which the findings presented may be generalizable, and we return to this point in the concluding chapter.

Overview of This Book

This book is aimed at all those who are interested in UR in the sciences, including faculty who lead research groups with undergraduates, faculty and program developers interested in adapting the lessons learned from UR in science to other fields, academic administrators, policymakers, program officers, researchers, and evaluators. Research findings like these establish a knowledge base and begin to define effective practices from which variations and innovations can be created to achieve the same good outcomes in other ways.

Chapter Two provides a review of the literature on UR, comparing findings from published research and evaluation studies that provide evidence about the outcomes of both faculty-led research and structured UR programs. The body of the book is organized into two large sections: Chapters Three through Six focus on the student outcomes of UR, and Chapters Seven through Nine offer insight into the processes by which these outcomes are achieved and sustained, largely from a faculty perspective.

Chapter Three is the linchpin of the student section, as it defines the benefits to students of conducting research and provides a student perspective on how these gains arise. This discussion is extended in Chapter Four to consider how and whether these same benefits can also be gained from other college experiences, and in Chapter Five to elucidate the longer-lasting impacts of UR on students' postcollege work and educational paths. Chapter Six addresses the case of structured UR programs targeted to the recruitment and retention of students underrepresented in the sciences.

The perspective of research advisors is critical for understanding how student outcomes come about. Chapter Seven describes research advisors' everyday work with students as they make use of authentic problems to achieve the student outcomes that they desire. Chapter Eight takes a more global look at how advisors mentor students and assess their progress. And Chapter Nine examines research advisors' work within their institutional context, identifying the costs and benefits of their UR work and examining how advisors balance these in pursuing their own scholarly work while also attending to students' educational growth and professional development.

The concluding chapter revisits key findings and reflects on their implications for practitioners, leaders, and funders. The appendixes set out methodological details that provide transparency about the evidence we discuss without weighing down the narrative, and that may be useful to other researchers. Appendix A describes the interview samples in detail, Appendix B elaborates on the methods of the study, and Appendix C summarizes the interview protocols. The detailed table in Appendix D includes the frequency counts for each student benefits category, for all five main interview groups, to provide supporting detail for the quantitative evidence presented in each chapter.