An Open Letter to Undergraduate Research Assistants

Professor McGuire
Williams College Computer Science Department

Revised May 5, 2008

This document explains the role of an undergraduate research assistant in the graphics lab and what you can expect from the experience. Please read it carefully and talk to me about anything that you don't understand or that differs from your initial expectations.

Every professor's methodology is slightly different, but about 90% of the way that I manage undergraduate research is probably common with most science professors at undergraduate institutions. It is also likely similar to the experience that you'll have as an undergraduate intern at a company.

This document is specifically intended for research assistants, and especially summer RAs. I have different expectations and ways of working with thesis students, who are more independent and have a different schedule.

Working as an RA is an incredibly rewarding experience. It is something that all of your science professors did early in their careers and something that I recommend to all CS majors. This document focuses on expectations and responsibilities. While reading the details, don't lose sight of the big picture. You have an opportunity to work on world-class scientific research and the satisfaction that you can expect from being the first to explore a problem while interacting closely with your peers and faculty.

Why RA?

The College's mission is divided between undergraduate education (which you're very familiar with by now!) and professional scholarship. In the sciences, scholarship primarily means research. The research process involves solving new problems in new ways and publishing the results as peer-reviewed scientific papers. So, faculty members teach courses and write papers. Just as the College hires undergraduate teaching assistants for courses, it also hires undergraduate research assistants to help with research. A TA is an important part of the team for a course, but isn't ultimately responsible for teaching or managing the class. Likewise, an RA is an important part of the research team but isn't responsible for managing the whole process or writing research papers. An RA learns the process of computer science research and makes a meaningful contribution towards a result. (In contrast, a thesis student *does* write a thesis and manage most of the process independently.)

Working as an RA is a good way to form a close relationship with a faculty member. The experience extends your education in ways that are not possible within regular coursework, and will also help you decide whether you are interested in pursuing an academic career. Regardless of your career choice, that faculty member's support will probably be important for your job or graduate school applications.

Along with increased responsibilities you'll enjoy increased access to department resources to help with your role, increased access to and familiarity with faculty, and be invited to some special activities like the summer canoe trip and science lunches.

Most research assistantships are paid positions, but if you're interested in serving as an RA primarily for the financial incentive, it is probably a bad fit for you and I recommend another job. A CS student can make a lot more money as a programming

intern at a company or as an OIT employee than as an RA, and if you aren't interested in research you won't enjoy the position or appreciate your tasks very much.

Student Qualifications

To work on the kinds of problems that I'm interested in, you'll need a good working knowledge of how computers work at the level of CS 237, and the ability to program in an object oriented language like Java or C++. Most of my code is in C++, GLSL, Python, and Matlab so you should expect to learn those languages quickly if you don't know them already. Computer graphics relies heavily on linear algebra (as taught in Math 211), statistics (Stat 201), and calculus, so experience with those is a great advantage when applying for a research position. We use video, modeling, and image editing software like FinalCut, 3DS Max, and Photoshop heavily in the lab. I do not expect students to know those packages when they are first hired. If you are already experienced with them, that is a bonus and something that you should tell me about.

Most research assistantships in the CS department are paid positions for upper-class students. Other faculty members will occasionally take on students who have only completed CS 136. I will not hire students who have not completed CS 237, however I will occasionally bring students into the lab as volunteers. The volunteer position is a lower commitment for both of us. That increases your chance of getting a position in the lab and reduces the difficulty of your leaving if you change you mind.

What is Research?

Scientific research seeks to understand the world around us. It starts with a broad, open-ended question, like "can computers think?" or "how can we simulate light?" We then choose a sub-problem that will give us insight towards the larger issue but is more manageable, like "what error function produces the best object identification when using gradient descent optimization?" and work on that. Throughout, we follow the scientific method:

1. Choose Problem
2. Form Hypothesis ("guess the answer")
3. Experiment to Objectively test the Hypothesis
4. Submit Results for Peer Review

Steps 2 and 3 repeat a lot, since rarely is our initial hypothesis correct or the first experiment definitive.

Research is very different from course work. The biggest difference is that nobody knows the answer to the problems that we're looking at! In fact, we don't even know if we're asking the right questions, or whether it is possible to answer them. It is more like detective work than a homework assignment. You'll be a lot more independent than in a course, with less structure. This means that you have to be proactive about seeking documentation, devising smaller experiments, and structuring your workday. Often you'll be facing a problem and unsure how to even begin, so you'll just have to dive in and try something without any idea of whether it will work. As discussed below, the integrity standard in research is much higher than in a course.

You must act ethically and carefully in all of your work. Having a project fail is an acceptable outcome, but falsifying or misrepresenting results is beyond unacceptable. Doing so will terminate your academic career. The standard for scientific ethics is different than the practices of other fields and what you might expect from everyday life, so read the section below on integrit closely.

Research is very different from commercial software development. That's because solving the problem in research isn't as important as *how* you solve it. A scientist has to seek theoretically rigorous solutions that advance our understanding of the problem and the field. We also interact closely with prior art, giving credit to previous related work and often choosing to explore or not explore a direction based on what has been done before. The company that produces a new word processor has no obligation to talk about previous software. A scientist must discuss how his or her work differs from what has been previously done, and to do that, must also be aware of a huge body of work.

The reason that methodology, rigor, and prior art are so important in research is that solving specific problems isn't really our goal. Specific problems provide context and motivation, but science is really about advancing the state of the art and understanding how the universe works. This is why *how* a solution works is more important than how well it works or how practical it is to apply.

Integrity

The integrity standard in Williams courses (i.e., the honor code) is very high. The integrity standard in research is even higher. Research works because everyone in the field reviews each other's work and trusts each other's published results. You absolutely must report results accurately and honestly. You must show result images and data as produced by your program. It is completely unacceptable to edit result images or alter a data set to conceal artifacts.

We have to be our own harshest critics, verifying all results and testing alternative hypotheses. If there is a case where our algorithm could fail, explore that just as diligently as the cases where our algorithm succeeds.

Plagarism and copyright violations are unacceptable in any form—all relevant work must be cited and used with permission, whether it is an idea, an algorithm, data, or code. Specific examples include:

- Only use content for which you have written permission from the copyright holder, and then clearly cite the origin. This includes:
    - Movies
    - Still images
    - Music (e.g., for the soundtrack of an animation)
    - 3D models
    - Text (e.g., quotations)
    - Equations
- When gathering images and other data, ensure that you have written permission of the people involved. For example, release forms for actors (even extras) in movies you make and release forms from the curator when photographing in a museum.

Note that for copyright, it does not matter whether a 3rd party work is on the Internet, or whether our application is non-commercial, or whether you are a student. Permission is still needed in all of those cases.

Until research is published, it is essential that the work is kept confidential. Do not discuss your project or unpublished results with anyone outside the College. Do not describe your work in a job application or on your website without discussing it with me. Do not discuss problems that you're working on using mailing lists or internet forums, and do not ask for external help with research problems without asking me first.

Likewise, it is important that our data remain private until published. Some of our data is not even our own. We work with assets provided by companies and other schools and have limited permission for their use. Direct all external inquiries to me and always ensure that data is protected on CS department machines.

Tasks

As a computer science research assistant working on computer graphics and vision, you'll work on many different tasks. Part of the appeal of this field is that it combines so many different techniques and disciplines. The kinds of tasks that my RAs perform are listed below. Although I primarily distribute tasks based on your existing skills, this is also an opportunity to stretch and learn some new ones. Those might range from learning a new proof technique to using a soldering iron.

- Programming
  - New experiments
  - Support code
  - Testing, debugging & maintenance
  - Scripts for data processing
- Mathematics
  - Checking and completing derivations
  - Checking and completing proofs
- Reading and Writing
  - Progress reports
  - Posters
  - Proof-reading papers
  - Literature search
  - Researching hardware purchases
- Assisting with physical experiments
  - Measurements and data entry
  - Constructing and designing apparati
  - Recruiting subjects and actors
  - Transporting and configuring equipment
- Supporting the lab
  - Presenting for visitors
  - Cleaning and maintaining the lab
  - Helping, teaching, and managing other research students!

Recognition

Another benefit of working as an RA is that it can be an important step towards a career in computer science or a related field. If your summer experience is a productive one, I'll not only be writing letters of recommendation for you in the future but promoting you within industry and academia. There are many ties between industry and academia, and a call to contacts at companies like NVIDIA, AMD, Pixar, Electronic Arts, or Google should help you to get an interview there. Likewise, I actively recommend my best students to other professors at graduate schools who are looking for new PhD students.

When I publish a paper that benefited from your work, I will acknowledge your contribution both in the paper and in talks I give about it in the future. On the graphics group web page I also list the current and previous research students. It is likely that some of the code you have worked on will be released when the work is published; in that case you will be credited in comments on the code. If you work on the G3D library you'll be listed in the documentation there as a contributor as well.

I often coauthor posters with students. Summer research students will always have the opportunity to present their work on campus, and most RAs will have a chance to present their work to visitors. When appropriate, I will take posters to a scientific conference as well. In that case I'll be presenting our work but your name will still be right under the title with mine.

I will occasionally invite a student to coauthor a paper when their independent contributions to a project are a substantive portion of the final algorithm design and that student has sufficient experience and time to work on the paper. Although this has happened several times, it is not the typical RA experience. Specifically, you should not expect to be listed as a coauthor on papers reporting the experiments that you worked on, and you should not feel bad if we do not coauthor a paper together. Writing a scientific paper is a very different skill than programming, reviewing literature, or performing experiments. It is not a skill that you are expected to have at this point in your career. I do hope that by reading other papers and writing reports will begin to develop that skill, which you can then exercise as a thesis student and in graduate school.

Intellectual Property

As a rule of thumb, assume that while you are employed by the College all of your work products related to your project and computer science are legally owned by either the College or me, even if you work on them outside of your regular hours. This is standard Massachusetts law and is what you'd encounter in any workplace. This is rarely an issue in academic research because our goal is usually to publish our work and not make a commercial product, however, very occasionally a student will work on a patentable technology and it is important to understand that you do not have ownership of that.

Many CS students also work on personal projects like video games or web applications. The department encourages that, and nobody wants to take that away from you or discourage you from using appropriate College resources to advance your personal projects. If you have concerns or are working on a third-party commercial product while an employee of the College, you should discuss the intellectual property policy with Keith Finan in the Provost's Office.

As part of my research I work with confidential data from other research and product groups. Do not use such data unless I explicitly direct you to do so and explain

the limitations on its use. All confidential data is contained in clearly marked directories to avoid confusion. You should take every effort to prevent accidental dissemination of this data.

Summer Hours and Interaction

I expect summer RAs to work a 40-hour week of five 8-hour days, from 9am-6pm. It is common to get excited about a project and work more than that, although you are not required or paid for doing so. We break for lunch from 12 pm-1 pm. On Tuesdays lunch is provided by Div III—you should attend the lunch *and* the following talk. Although the CS department is a very flexible and casual place to work, don't forget that this is a job and I'm your boss. During work hours you generally should not be playing video games, browsing the web for non-work topics, or having long phone conversations. Occasionally running errands and checking personal e-mail is perfectly normal, but don't let it get in the way of your productivity. We both want your recommendations to say that you're a responsive and hard worker, not that I could never find you and that you surfed the web all day.

You'll sit in the Unix lab, or the Graphics lab if your project needs special equipment. You will do additional research in the science library and some experiments around and off campus. I recommend getting out of the lab when reading papers or thinking over ideas—a computer room is good for working on a computer, and that's about it. The coffee house, library, and our lovely campus are much more stimulating for thought. Just send e-mail to gfx@cs.williams.edu if you're ducking out so that others will know where to find you.

We'll agree to milestones each week. If you're not going to make a milestone, let me know early so we can change your process to stay on target. Don't spend 40 hours being frustrated and making no progress!

You'll attend short daily status meetings with the other RAs and meet with me once a week. These meetings will be focused on milestones. You'll also maintain a log of your progress and stay in regular e-mail contact with me so that I can keep an eye on your work even when we aren't face to face.

Check your Williams e-mail for messages from me at least twice a day; I expect prompt responses, even if you just acknowledge the e-mail but don't have an answer to my question yet. (Lifelong productivity tip: Personally, I find that I'm more productive if I *don't* read e-mail or open a web browser first thing in the morning or after lunch—I like to dig in to my work, and then check e-mail at 10 am or 3 pm when I'm starting to slow down from my initial burst.)

Depending on the nature of your project and how it is progressing, we might also work side by side a lot more than just status meetings. You'll rely primarily on other RAs for moment-by-moment help with programming and using lab resources. The amount of time I spend with you and tasks that I ask you to take on are based on your skill set, the state of your project, and how your project aligns with the others that I'm working on. It isn't a reflection of how I feel about you or whether you are doing good work.

The head RA will be in charge of the lab. That RA will hold daily status meetings with other RAs, give me daily status reports, help and train more junior students, and be responsible for the progress of other RAs towards their milestones.