Experimenting with In-person Grading

J. Philip East
Computer Science Department
University of Northern Iowa
east@cs.uni.edu

Abstract

In this paper I describe an ongoing investigation of in-person grading (IPG). Having determined that IPG is reasonable from theoretical and instructor points of view [2], the study reported on here was conducted to determine if IPG was reasonable from the student perspective. The assessment of student reactions is described in some detail. The initial use of IPG and the work described here serve to pilot-test the concept and set the stage for a more formal future experiment. The work here also illustrates a multi-step approach to pedagogical study considered useful for both informing practice and leading up to more formal experimentation.

Introduction

During the past three years, I have been experimenting with the use of in-person grading (IPG) as an alternative to traditional grading of student programs. I encountered this concept via the SIGCSE discussion listserv and Doug Cooper's statement, "If you've never tried it [face-to-face grading], let me encourage you to take the plunge—it is probably the single most important improvement that can be made in course management." [1, p.130].

With this encouragement, I embarked on a four-step examination of IPG. The first two steps (an assessment of face validity and a pragmatic examination from my point of view) are discussed below. The third step (a pragmatic examination of IPG from the student point of view) is a principal subject of this paper. The fourth step (a controlled experiment of IPG) is expected to be the subject of a future paper.

This paper has two purposes. One is to report on my experience with IPG. The other is to illustrate a process of reflective practice [6] and experimentation in computer science education.

In the next section, I briefly describe IPG as I have used it and the first two stages of my model for reflective practice and experimentation. The third step in the process—an examination of student perceptiions— is presented and I discuss the method and results of the recently completed study. That discussion is followed by my interpretation of those results. In closing, I summarize the paper and review my suggested process of reflective practice and experimentation.

Instructor Perceptions Of In-Person Grading

The Practice of In-person grading

The process of IPG is relatively simple. A student works on a programming assignment and signs up for a meeting with the instructor (or TA). The instructor examines the program and its results in preparation for the meeting. At that meeting, the student's submission is discussed. The student receives feedback in person and can probe for additional information. The instructor is able to ask questions to ascertain and clarify student understanding.

Specific logistics of IPG can (and probably should) vary from instructor to instructor and course to course. I have used the technique in Algorithms and COBOL courses. In an effort to develop student judgment about program quality, I have typically asked students to submit self-evaluation forms with their programs. The forms examine correctness, specification fulfillment, design, layout, and documentation. I mark my assessments on the student-submitted form and it serves to focus part of the IPG meeting. We also discuss student questions and I try to determine the student's understanding of language and programming concepts addressed by the assignment.

The meetings take from 15 to 30 minutes. With 20 to 30 students in the class, a week in which I conduct IPG has little or no unscheduled time. I spend evenings preparing for the next day's meetings. I cannot conduct IPG for every student for every assignment for every class and perform any scholarship or service activity (or have any mental down-time during the day). As a result, I limit IPG to no more than half the weeks in a semester. Thus, not every assignment in a particular course or even every course is subject to IPG.

My preparation for each meeting includes a relatively quick perusal of a student's program and results of its execution. While doing that, I note questions to ask and comments to make regarding the program. I have found that consistency in grading seems less critical when I can discuss the

program with the student. This means that while grading actually takes more time when using IPG, the time is less intense and chunks of time need not be blocked out to do the entire set of assignments in order to provide consistency. At the end of an IPG meeting, I provide the student with a general assessment of assignment quality (poor, okay, good, excellent) but assign numerical grades after the meeting. I review the scores after all meetings have been conducted in an effort to ensure consistency and to avoid pleas for better grades.

Initial Assessment of Face Validity

The initial test that the concept of IPG had to pass was that of face validity, which is similar (if not identical) to what Kerlinger [5] calls construct validity with respect to empirical research. When performing empirical behavioral research, construct validity is that which addresses whether the data being collected actually measure the construct or property posited. When discussing teaching practice, we are not necessarily measuring anything. However, it is still critical that we have a theory-based rationale for what we do. Face validity refers to the reasonableness of a concept, construct, or practice on theoretical grounds. One argues the existence or quality of face validity using beliefs about learning and teaching. Before deciding to experiment with a new pedagogical practice, a reflective practitioner [6] will examine the practice for its fit with theory. To me, this means seriously examining whether there are reasons to believe the practice will help and what any such reasons might be. Without face validity, the foundation for pedagogical practice is sand rather than bedrock.

I examined the practice of in-person grading to see if my view of theory (published theory as well as personal beliefs about learning and teaching) supported the practice. Space limits prohibit a full discussion of theory and IPG's face validity, but major aspects are included below.

My own views of learning theory are somewhat eclectic rather than along the lines of any single theory I have examined [2]. I happen to believe that learning at the physical level is a behavioristic or associationistic process. However, the myriad influences of neural connection formation are probably beyond our understanding and certainly beyond our control. This kind of theory offers little insight for adjusting teaching behavior. However, the associationistic aspect of it is compatible with constructivist theory.

Constructivist learning theory offers a better opportunity to address teaching activity. The primary tenet of constructivism is that individual understanding arises from personal experience. We cannot know a student's experience by simply examining their programs, thus, if we are to teach better, we must attempt to "get inside the student's head" and work to enhance or rectify the student's understanding of what we wish to teach. Traditional grading or feedback mechanisms offer little opportunity to know and correct student misunderstanding. To me, constructivist teaching requires that I strive to gain insight into individual student thinking. My examination of IPG with respect to learning theory led me to believe it should be able to help me better see inside student heads.

I also think programming is a skill at which we may eventually develop significant expertise. One of my beliefs about skill development is that coaching can enhance it [3]. A coaching-like atmosphere is more nearly provided with IPG than with other feedback mechanisms I have used.

It also seemed to me that IPG would encourage students to seek help from me. A common complaint of students is that they spend inordinate amounts of time on programming assignments. Much of it seems to be spent mulling over a problem without making progress and eventually asking a question. Being in a closer relationship with the instructor (presumably provided by IPG) could lead to greater (and sooner) student questioning.

IPG passed my first test—face validity. Based on my beliefs/thinking/theory about teaching and learning, IPG was reasonable. Thus, it was okay to begin serious study of the technique.

Pragmatics from My Point of View

Having decided that IPG met theoretical muster, the next step was to see if it was practical in my own instructional context. Armed with an exhortation to use IPG and a positive determination of face validity, I tried it [2]. It seemed useful in that:

- Students seemed to pay more attention to my comments.
- Some students requested a chance to revise the program given their new understanding of how to go about it.
- I had a sense that IPG allowed us (both I and the students) to better focus on assessing program quality.
- For me, grading was completed sooner (though it required more time) than with my previous practice.
- Even though I spent more time grading than before, I did not begrudge the time because I felt that I was teaching when I was grading. It became an instructional act rather than one of assessment.

At this point, there seemed to be ample reason to further investigate the use of IPG for programming-oriented computer science courses.

Student Perceptions Of In-Person Grading

Method

In the spring of 2002, while using IPG for the fourth time, I asked students to provide me with some feedback about IPG. This particular class was a sophomore/junior level class in Algorithms. My goal was two-fold. I wished to find out if students thought IPG contributed positively to their learning. I also wished to get their suggestions as to how I might better implement IPG in the future, assuming students thought its continued use reasonable. With the assistance of a colleague in the college of education I devised a two-part questionnaire designed to provide insight into both these goals.

The first part was a closed-response section in which students were asked to consider a statement and indicate their agreement or disagreement at one of five points along a continuum:

```
Strongly Somewhat   I am    Somewhat Strongly
Agree     Agree    Neutral  Disagree Disagree
O----------O----------O----------O----------O
```

Students could respond that they had no opinion. The items used are shown in table 1.

Five of the closed-response items were designed to test slightly different aspects of the efficacy of IPG and eight items addressed possible improvements in its implementation (they are marked with *** in table 1). Providing several slightly different items about efficacy was expected to better ensure actual student perceptions would be assessed. Other questions sought student reactions to my concerns about IPG or ideas I had for the possible improvement of IPG. For example, I had reservations as to whether I would be comfortable meeting and talking with the students and assumed students might feel similarly.

The second part of the questionnaire consisted of seven open-ended questions designed to identify positive elements of IPG that should be retained and negative elements that should be eliminated. Open-ended questions would allow students to suggest actions we had not thought of. The questions are shown in table 2.

Table l. Closed Response Results

Statement	Strongly Agree				Strongly Disagree	N.O.	Avg.
	2	**1**	**0**	**-1**	**-2**		
In principle or theory, I think IPG should improve student learning. ***	9	13	4	0	0	1	1.2
As employed in this class, IPG was "better" than grading I typically receive. ***	3	14	6	1	2	1	0.6
At first I was nervous about meeting with the instructor for IPG.	4	11	3	4	5	0	0.2
I was nervous every time I met with the instructor for IPG.	0	7	4	9	7	0	-0.6
The instructor was reasonably prepared for the IPG sessions I attended.	10	13	1	2	0	1	1.2
I found the combination of instructor's written and oral comments more beneficial than the written comments I normally receive on programming projects. ***	8	10	5	2	1	1	0.8
The self-evaluation is an important component of IPG.	3	9	6	6	2	1	0.2
I believe the IPG improved my learning. ***	2	13	5	6	0	1	0.4
The student should be more "active" during IPG, e.g., taking notes, explaining thought processes, asking questions	3	12	7	1	2	2	0.5
The IPG could be improved if the instructor asked more questions.	0	7	11	4	2	3	0.0
IPG should always be conducted in the lab so the programs can be demonstrated (even though the meetings would be less private).	2	6	8	4	3	4	0.0
The meeting time for IPG should be longer (30 minutes allowed and used if needed).	0	4	7	9	5	2	-0.6
The instructor should continue using, and improving his use of, IPG. ***	7	9	6	3	1	1	0.7

The questionnaire was given to students on the last class day of the semester. No names were associated with any questionnaire. The components of the questionnaire (open- and closed-response) can, however, be linked to a single, anonymous individual.

Results of the questionnaire were tallied and recorded. Responses to closed-response questions were given numerical values and summed and averaged. Responses to open-response questions were recorded and examined.

The purpose of this study was to see if my students were amenable to the use of in-person grading. If students voiced significant dissatisfaction with the technique it would make little sense to continue. Active dislike on the part of students would likely mean that any benefits, either real or perceived, would likely not materialize in practice over time. I decided that the student response needed to be better than neutral. Only a relative clear, though perhaps weakly positive, result would justify my continuation of IPG.

Results

Table 1 shows how students responded to the closed-response questions. The column headed by N.O. indicates students with "no opinion". The Avg. column indicates the average response to each question. Calculation of the average assumes strong agreement is worth two points (strong disagreement is minus two points) while mere agreement and disagreement count one and minus one point respectively. A response of "I am neutral" was given zero points. No-opinion responses are excluded from average calculations.

The average item score on the five efficacy items (those with asterisks) was 0.75 (n=5). The average response score on these items was 0.73 (n=5). Both these values are strongly positive on the -2 to +2 range of possible scores.

Viewing the results of each student provides an interesting alternative perspective. By examining student averages rather than question or response averages we can infer which students were satisfied and which were dissatisfied with IPG. Seven students had negative results

```
-0.77 -0.29 -0.45 -0.15 -0.08 -1.00 -0.62
```

for an average negative score of -0.48. Nineteen students had positive results

```
0.25 0.54 0.25 0.92 0.31 0.38 0.73 0.55 0.69 1.08 0.31 1.25 0.54 0.92 0.77
1.23 0.38 0.18 1.23
```

yielding an average positive score of +0.63. One student had a zero result. The overall average score was +0.34. Students supportive of IPG expressed stronger feelings that students who did not support IPG. Additionally, positive students outnumbered negative ones by more than two to one.

Table 2 shows the number of responses to various items in the open response portion of the questionnaire. Space limitations preclude providing the responses. They were examined to see how many were relevant to IPG (all were relevant to some aspect of the class but some did not pertain to IPG). The count of "relevant" responses is included in table 2. One example of item irrelevancy relates to item 4, negative aspects of instructor's actions or attitude. Three students disliked the grading scheme used by the instructor. It was identical to the grading process for non-IPG grading and therefore irrelevant to IPG. Several irrelevant responses alluded to the programming languages allowed on the assignments. One student response to identifying positive and negative aspects indicated everything was positive and nothing was negative. While nice to see, these responses were also irrelevant to the questions asked.

The first four items allowed students to indicate whether they felt positively or negatively about IPG (first and third, positive; second and fourth, negative). Clearly, there were more positive, relevant comments than negative relevant comments.

Table 2. Open response questions, response counts, and responses deemed relevant to IPG

Question	Responses	Relevant Responses
What aspects of IPG do you feel are noteworthy?	26	26
What aspects of IPG do you feel are negative?	18	12
What aspects of the instructor's actions or attitude do you feel were positive or noteworthy?	20	19
What aspects of the instructor's actions or attitude do you feel were negative?	6	3
What specific suggestions do you have for improving IPG as performed by this instructor?	14	11
What student activity would make IPG more effective or beneficial to learning?	12	11
Do you have any other comments or suggestions?	11	11

Conclusions

I interpret these results as indicating a quite positive student perspective toward the use of IPG.

All five questions relating to the efficacy of IPG had positive scores. This suggests that students did indeed consider IPG useful to learning--in general and for them in particular. While most any positive result would have been interpreted as justifying further efforts, these scores were deemed to show strong student support for IPG.

It is also informative to examine the numbers of students disagreeing with the IPG's efficacy. Of particular note is the fact that every student thought IPG was "in theory" a good idea. Only 3 of the 27 students had any disagreement at all with the notion that IPG was better than other grading and feedback procedures they had experienced.

Similarly, when all the closed-response items (rather than just those concerning IPG efficacy in general) were examined, students were relatively strongly in favor of IPG. Only 7 of the 27 students had a "negative" average score. Students were 3 to 1 in favor of IPG and their scores indicate a greater approval than the nay-sayer's disapproval.

The open-response items were included to allow students to suggest improvements or problems I had not thought of. While student responses on these items do not in themselves indicate support or lack of support for IPG, I think the numbers of positive and negative responses can. The fact that 26 students identified something "noteworthy" about IPG and only 12 noted "negative" aspects is also seen to support IPG. I also interpret the fact that 11 students provided suggestions for improvements as positive support for the practice.

Finally, I consider it very significant that over half of the students agreed that IPG actually made a difference in their learning. I suspect that such a result would be difficult to achieve with most any other instructional intervention. Perhaps, Doug Cooper was right, "it is probably the single most important improvement that can be made in course management."

I am interpreting these results to mean that students will support IPG if we implement it reasonably. Thus, it is appropriate to continue the investigation of IPG as a possible improvement in the pedagogy of computer science.

Summary, Suggestions, & Plans

Two goals have been pursued in this paper. One relates to the particular pedagogical technique of in-person grading. The other relates to a suggested process for reflective practice and pedagogical experimentation.

I noted earlier that a college of education colleague assisted in this work. I sought her assistance, even though my training in computer science education had substantial educational research and theory components. As computer scientists, we often seek the eyes and opinions of knowledgeable colleagues to assist us in our program design, development, and testing. Similarly, we should seek the assistance of knowledgeable colleagues in our design and implementation of pedagogical studies.

Some aspects of this paper have focussed on the process of pedagogical study. I suggest that at least four steps or stages are necessary for quality research. Actually, the first three steps ought to be used in any serious examination of educational practice (even if no formal research study is intended). The steps involve determinations of:

- **Face validity for the practice.** Is the practice supported by the theory (both formal theory and personal beliefs) about learning and teaching? If so, the theoretical rationale for the practice should be communicated along with the practice. If not, the practice should probably be abandoned. Without an explicit discussion of face validity and the rationale for educational practice, all discussion of our practice will remain anecdotal. The importance of this step cannot be overstated if we are to advance our understanding of instructional practice.
- **Practitioner satisfaction.** This is determined by trying the practice in your own teaching. Just because a practice meets theoretical muster does not mean it can be implemented in a particular educational context. It should be reasonable to report conclusions to the computer science education community even if the practice did not work in an individual situation. It may be that others can offer a fix or that your findings will keep others from wasting their time. Note that often this step is the only one taken in the examination of pedagogical practice. Without reference to the theoretical basis for a pedagogical practice, we are forced to rely on personal opinion. Without the next two steps, we have little or no evidence beyond the enthusiasm of the practices' proponents.
- **Student satisfaction.** If students do not "buy in" to the practice it is unlikely to be successful no matter how good the theory or how strong the instructor enthusiasm. Again, it is reasonable to report both positive and negative results so that others can learn from an individual's experience.
- **"Experimental" validity.** If the practice of teaching computer science is to actually improve, we must eventually perform quality "research" to objectively examine our practice. It makes little sense to do research on a practice in which the first three steps do not have positive results. Without strong research support for our practice, however, we cannot advance the state of computer science education—it will remain art or craft guided by the seat of our pants. I enclose the word experimental in quotes because I think it important to note that pedagogical research should be as rigorous as any other research. However, I do not believe that statistically analyzed quantitative studies are the only means of rigorous research. Indeed, it is not uncommon that statistical analysis "proves" hypotheses totally inappropriately.

The other purpose of this paper was to report the results of a step 3 study that I conducted relative to in-person grading (IPG) of programs. The practice came highly recommended, made sense in light of my own beliefs about teaching and learning, and had met with satisfaction when I tried it in my own teaching. Student responses to a personally-developed questionnaire concerning student satisfaction and perception of utility have been analyzed as supporting the use and further study of the practice.

A colleague and I have begun to do so. We are presently conducting a study in which we will gather data in an attempt to determine if there are differences between three alternatives to providing student feedback.

References

[1] Cooper, D. (1995). *Teaching introductory programming (with Oh! Pascal!)*. W.W. Norton & Company, New York.

[2] East, J.P. (2001). Experience with in-person grading. Proceedings of the 34th Midwest Instruction and Computing Symposium (on CD-ROM). Cedar Falls, Iowa. [Available on-line via http://MICSymposium.org]

[3] Ericsson, K.A., and Charness (1994), N. Expert performance: Its structure and acquisition. *American Psychologist, 49*, 8, 725-747.

[4] Hilgard, E.R., and Bower, G.H. (1975). *Theories of learning*. Englewood Cliffs, New Jersey,: Prentice Hall.

[5] Kerlinger, F.N. (1973). *Foundations of behavioral research (2nd Ed.)*. Holt, Rinehart and Winston, Inc., New York,.

[6] Schon, D.A. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Acknowledgments

I wish to acknowledge the assistance of Katheryn Kinnard East in all aspects of this work and report. I might well not have done it without her encouragement and would have done it less well without her assistance.