Measuring increased engagement using Tablet PCs in a code review class

W. Fagen
Computer Science Dept.
University of Illinois
wfagen2@illinois.edu

S. Kamin
Computer Science Dept.
University of Illinois
kamin@illinois.edu

ABSTRACT

The Programming Studio in the University of Illinois Computer Science department is a required course in which small groups of students participate in weekly code reviews of each other's programs. To attempt to increase student engagement in the discussions, Tablet PCs were introduced for several weeks in the middle of the semester. By recording the discussions before, during, and after the use of tablets, we measure the effectiveness of this intervention. In doing so, we develop a simple metric to measure the "active engagement" of the participants. We found each section was significantly more engaging when using Tablet PCs ($p<0.0001$) and the large majority of individual participants were more engaged. This paper contributes both an objective measurement of "active engagement" and a successful intervention in a programming studio-type course.

Categories and Subject Descriptors

K3.2 [**Computer and Information Science Education**]: Computer Science Education

General Terms

Education, software engineering

Keywords

Code reviews, Tablet PC

1. INTRODUCTION

Students in the Computer Science department at the University of Illinois are required to take CS 242, Programming Studio. Normally taken in the first semester of the junior year — after the core programming courses — the Studio is intended to build individual programming skills. The overall approach is drawn from studio courses in architecture and fine arts: students present their work "in public," and get feedback. Specifically, students meet every week in discussion sections of 3 to 6 students, plus a moderator, where each student has twenty minutes to present their program. With some exceptions, the discussions focus on the code itself, rather than the functionality of the program. Assignments are not as structured as in traditional classes, but the students generally have an idea of what the other students' programs do. (The main exception is in the last four weeks of the semester, where student-selected term projects are presented; during that time, program demos are mixed with code discussions.)

The Studio course was described in 2007 [WK07]. Although the Studio seemed to be useful to most students, there were some difficulties. We noted, "We need to continually work on maintaining the quality of the discussion in the discussion sections." Programs are complicated, and it is hard to have a stimulating discussion on something that only one of the interlocutors understands. The fear was that the discussions would be closer to monologues, the critiques offered would be superficial, and the discussion would be dull and unengaging. [WK07] describes various mechanisms to help stimulate discussion. Using Tablet PCs is another such a mechanism. This paper discusses how they were used in the programming studio, and how we attempted to verify that they were having the desired effect.

Tablet PCs are ordinary PC laptops with the added feature of a screen that can be written upon, using Microsoft's digital ink API. These machines have an "active digitizer" screen which can only be written upon with a special stylus. In this paper, we refer to these devices as "tablets," but we caution the reader that these are not iPads or Android tablets, which mostly lack the active digitizer and are not as easy to write on. Writing on them is very natural, and students adjust to them easily. In the Studio, each student has a tablet, and all comments written by any student are instantly shared with all other students. The system we used provides in effect a shared whiteboard, with a student's code forming the background image.

Traditionally, the Studio operates by having each student explain his or her program as it is displayed on a large screen. Adding the tablets provides two benefits: First, writing, which makes it easier to ask and answer questions about code. Second, private copies of the presenter's code, allowing each student to explore it independently, rather than following the presenter. In introducing the tablets, we hoped these benefits would help make the discussions more engaging and efficient.

The study reported here was peformed in the Spring, 2011, semester. Its goal was to test whether the tablets had a positive effect on the discussions. We set up a simple experiment: In six sections, we had students alternate between tablet use and the traditional Studio. We gathered data by using personal audio recorders to record the interactions during each meeting of each of the six sections. In this paper, we report on the results of the analysis of this data.

This paper makes three contributions to the study of Computer Science education:

- We describe a Tablet PC system supporting code reviews in the Programming Studio, and explain why it is helpful.
- We give a method for mechanically measuring "active engagement" of the Studio discussions, based on analyzing audio recordings of the participants.
- We describe an experiment to test the hypothesis that the introduction of Tablet PCs lead to livelier discussions. We present data from voice recording analyses that confirm the hypothesis.

The next two sections of the paper describe the Programming Studio and the use of Tablet PCs. Section 4 describes the experimental set-up, and then sections 5 gives the results obtained from the recordings. These are followed by a discussion of related work, and our conclusions.

2. PROGRAMMING STUDIO

CS 242, Programming Studio [WK07], is a required class for CS majors at the University of Illinois, providing individualized instruction on programming by having students critique each other's programs in a code review setting. The class grew out of a concern that, despite many hours spent programming in CS courses, not all of our students were becoming proficient at it. A student could get through our program — we are undoubtedly not unique in this regard — without ever having a detailed, one-on-one, discussion about a program he or she had written.

The Studio is taken in the junior year, after all the core courses have been completed. Students get fairly open-ended assignments each week, and are expected to show up at the weekly discussion with a working program. In a group of three to six students, plus a moderator — a graduate student or upper-level undergraduate — each student presents his or her program for 20–25 minutes. The presentation may include a demo, but usually it is simply a discussion of the code. (This changes in the last few weeks of the semester, when students do self-chosen projects; then, more of the discussion is devoted to demos.) In the presentation, the presenter stands in front of the group and explains his or her program, while showing it on a large display. The arrangement is shown in Figure 1.

Figure 1: Non-Tablet PC Studio structure: students follow presenter

Most students find the studio beneficial, and many report that it played a key role in their development as a Computer Scientist. However, others complain that the discussions tend to focus on superficial aspects of their programs. This points to a fundamental issue that we identified in [WK07]: how can a discussion of programs be kept lively and engaging? Some mechanisms for addressing this issue are described in [WK07]. The tablets are our latest effort in this direction.

Terminology: We refer to the presenting student as the *presenter*, non-presenting students as *reviewers*, everyone in the studio, including the moderator, as *participants*. We also will refer to the group of participants, excluding the moderator, as *students*. A *section* is a regular meeting of a particular group of students; the experiment involved six sections (Thursdays at 5PM and 7PM, Fridays at 9AM, 11AM, 1PM, and 3PM). A *meeting* is a single meeting of a section, lasting from one to two hours depending upon the number of participants.

3. TABLET PCS IN THE STUDIO

The system used in the Studio is a straightforward "shared whiteboard" application. It was developed using the SLICE framework [Kam08,Slice], and is customized for our course. As each student takes a turn as presenter, he or she loads the program files into the system, where each is represented by a tab. Within each tab, the text of the file forms the background of the whiteboard; the text cannot be directly edited, but simply annotated upon. Figure 2 shows the SLICE screen during a studio discussion.

Each student in the discussion, as well as the moderator, has a tablet, and all share the same data. That is, any program loaded into the presenter's tablet, and any stroke made by anyone, is immediately sent to all others. The arrangement is otherwise identcal to that shown in Figure 1.

The tablets provide two capabilities: (1) The abil-

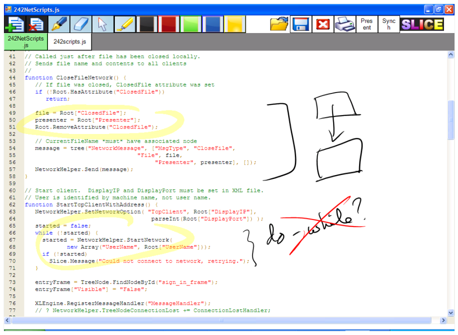

Figure 2: SLICE screenshot

ity to write, draw, and point. Drawing is useful to explain the deep structure of a program, data structure, or algorithm; writing makes it easier for reviewers to ask "what if you did it this way?" questions; pointing makes it easier to draw attention to a particular spot in the program. In general, writing helps make the discussion more efficient. (2) The ability to explore the presented program independent of the presenter. Each student can switch tabs and scroll through the files in the presented programs. This allows a reviewer to spend time understanding a section of code after the presenter has moved on, or simply to explore it in a different order than the presenter uses. (A button on the interface allows a student to synch up with the presenter.) This can allow a reviewer to gain a better understanding of the presented program — or at least of some part of it - and thus make more substantive comments about it.[1]

4. EXPERIMENT

The Studio discussion can switch easily from the traditional structure to the tablet structure. The design of our experiment was to have the discussions operate first in the traditional setting, introduce tablets for several weeks in the middle, and then move back the traditional setting towards the end of the semester. During this entire time, we took audio recordings of each meeting.

We ran the experiment in the Spring, 2011 semester for ten weeks, with six sections participating, with a total of 17 students and 4 moderators. We began recording students and moderators, using small personal recorders with lapel microphones on each participant, in week 5 and continued through the end of the semester (week 14). In weeks 5–6 and 11–14, the Studio ran in its traditional structure; in weeks 7–10, the tablets were used.

When we first introduced tablets into the course, in a pilot projectin Fall, 2010, we assumed the digital communication channel provided by the tablets would *displace* the traditional oral channel. However, we began to observe that the tablets might have the *opposite* effect, actually enhancing oral communication in the meetings.

[1]We made little attempt to distinguish the impact of these two features. It would be worthwhile to do so, in particular, because this second capability could be provided with ordinary computers, which may be more readily available than tablets.

Week	All	5–6	7–10	11–14
Tablet usage?		No	Yes	No
Recorded meetings	49	11	20	18
Total recordings	222	47	97	78
Hours	195.1	53.81	72.3	69.0
Size (GB)	10.46	2.89	3.88	3.70

Table 1: Basic information on audio recordings

That observation gave rise to the current study.

In particular, we postulated that the tablets would help make each student more individually engaged in the discussion. That is, each student would contribute to the conversation more often when the class used tablets than when it did not.

5. AUDIO RECORDINGS

To measure this form of engagement, which we call "active egnagement", we wanted to capture how long has passed since a given participant spoke. This metric would need to rate a conversation poorly if only a single speaker spoke for minutes on end without any participation by others and rate a conversation as productive if the entire class was engaged in the dialog.

In this section, we present our analysis. Following an explanation of the data collection process, we present a mathematical measure of "active engagement" of a meeting, which we call ϵ, and show strong, statistically-significant results showing a substantial increase in ϵ for tablet vs. non-tablet sections. We then define a related measure of "active engagement" of individuals, called γ, and give evidence that almost all participants showed increased γ in the tablet sections.

5.1 Data collection

To gather audio data, each participant was asked to wear a small, personal voice recorder with a lapel microphone. We used the VR5220 RCA Digital Voice Recorder, a consumer-level voice recorder that captures audio in a single-channel format at 8000 Hz.

To have minimal impact on the meetings, we attended only the first few minutes of the first meeting where recorders were used, to show the participants how to use them. We instructed the students not to worry, or stop the class, if the recorders malfunctioned. After each week, we collected the audio from the devices for storage. The stored recordings were tagged with the participant's name, and the time and date of the meeting.

The large majority of participants from each section had successful recordings. However, for both mechanical reasons and human error, we did not always obtain a complete recording for every participant in every section. Table 1 presents the basic statistics about the recordings.

5.2 Data processing

After the semester concluded, we had collected over nine consecutive days of audio recordings. Although each recording was made by a single student with an

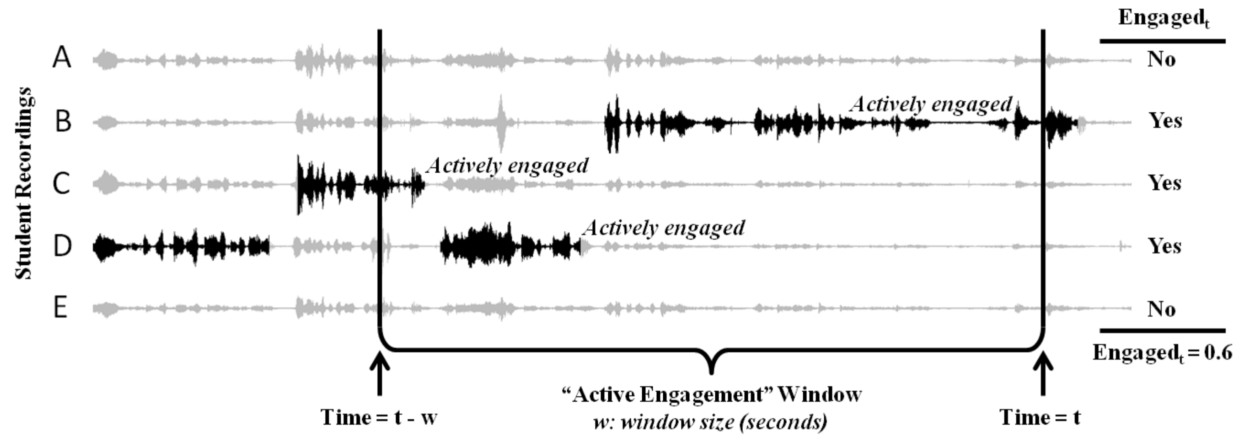

Figure 3: A visual overview of the "active engagement" metric, ϵ. Five participants in a studio meeting, labeled A–E, have their raw waveform displayed. The darkened areas of the waveforms were identified as turns in the discussion. Speakers B, C, and D had a speaking turn within the "active engagement window," resulting in a value of $\epsilon = 0.6$ at time t.

individual microphone, the recordings were corrupted by "cross talk" that sometimes made it difficult to isolate the actual speaker. We needed to determine who was actually speaking when.

The "cross talk" turned out to be useful for this, as we could use it to precisely synchronize the audio recordings in each meeting. Effectively, this process gave us a multi-channel audio file where each channel was the microphone of one of the participants. The next phase of our analysis relied on the observation that the channel with the loudest volume (over a threshold of background static) was almost certainly the speaker at that time. To validate this assumption, we had three different humans tag a recording with the times that a participant spoke. We found that the small differences between the human interpretations of the audio was within the same margin as the differences between any human's interpretation and the algoirthm's result.

Finally, with individual "volume events" identified and attributed to a speaker, the final analysis was to combine adjoining "volume events" by the same person into a *turn*. This combines individual words, eg: "How did your algorithm sort this list?", into a single speaking *turn*.

As a whole, our treatment of the raw data transformed groups of 8000 Hz input waveforms into a series of turns of each speaker. For all the analysis from this point forward, we work only with the turns that we identified through this process.

5.3 Active engagement

Our hypothesis was that the tablets would have a positive effect on the oral channel. Specifically, we were interested in how often an individual in a meeting of the studio section would contribute to the conversation. To measure this, we developed an "active engagement" metric, which we call ϵ.

This metric asks a simple question: "At a given moment in time, how many of a meeting's participants had a speaking turn within the last *(*time window)?". ϵ is the ratio between the number of participants who had a speaking turn within the time window and the number who were present at that time. If our window was one minute, and all the participants spoke within the past minute, ϵ would be 1.0. If only two of the six participants spoke, ϵ would be 0.33. Figure 3 shows a visual overview of the ϵ metric.

We analyzed the average ϵ across the full meeting, for several window sizes. For each hundredth of a second (0.01s) in each meeting, ϵ is calculated. For an hour long section, this would result in 360,000 individual values. Effectively, we have performed a detailed approximation of an integration across the entire meeting.

Consider an average ϵ of 0.7 for a given meeting. This value would indicate that, across the entire meeting, an average of 70% of the participants contributed to the conversation within the last *(time window)* of time.

We have not yet said what a correct window size would be. Using a small window, on the order of only a few seconds, would be uninteresting as one wouldn't consider someone unengaged in a conversation if they went just ten seconds without speaking. On the other hand, a large window would be equally uninteresting as *everyone* is likely to get a turn in each window. We present data for windows ranging from 30 seconds to 300 seconds, but often focus on a window of one minute.

Figure 4 shows the average ϵ of all 49 meetings that were recorded, using a one-minute window. On average, the introduction of tablets significantly boosted the average ϵ from less than 70% to nearly 85%. Table 2 shows the average ϵ for various windows, showing the result remains statistically significant with windows from 30 seconds up to five minutes.[2]

[2]The numbers presented are from the analysis of all participants, including both students and moderators. We also performed analysis using only the students.

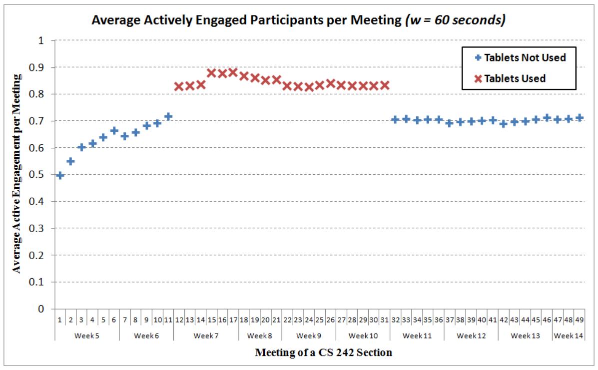

Figure 4: Average "active engagement" (ϵ) per meeting

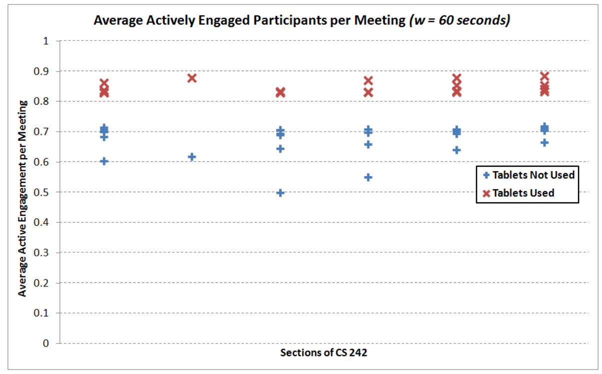

Figure 5: Average ϵ per meeting, grouped by the six sections that were recorded as part of this study

Window	No Tablets		Tablets		p
	Mean	StDev	Mean	StDev	
30s	0.5682	0.0508	0.7401	0.0230	<0.001
60s	0.6765	0.0519	0.8446	0.0189	<0.001
120s	0.7902	0.0438	0.9292	0.0135	<0.001
180s	0.8526	0.0343	0.9607	0.0095	<0.001
300s	0.9190	0.0209	0.9818	0.0056	<0.001

Table 2: Average ϵ across all meetings for various windows

Looking deeper into the data, we compared the meetings on a section by section basis. As the CS 242 studio course was a graded course and students signed up for a single, specific section, there was little variance week-to-week of the students who attended each section.[3]

Figure 5 shows the same ϵ values grouped by section. We find that the sections are very similar, and all showed improvement when using tablets.

We next asked if there were significantly more *turns* in the conversation. An increase in the number of turns might signify that a different type of conversation occurred, where each person spoke for shorter segments of speech. However, an analysis on the average number of turns per minute, displayed in the table below, shows no meaningful difference:

	No Tablets	Tablets	p
All	10.5934	10.4143	0.9278

Further analysis of the data allows us to look at individual participants. For individuals, we use a metric related to ϵ, which we call γ: the percentage of time in the meeting that the individual is "engaged" in the sense of having had a turn within the the previous time window. For example, a γ of 0.8 indicates that, at any given time in that meeting, there was an 80% chance that this participant had spoken within the previous window. In the following, we again use a 60–second window.

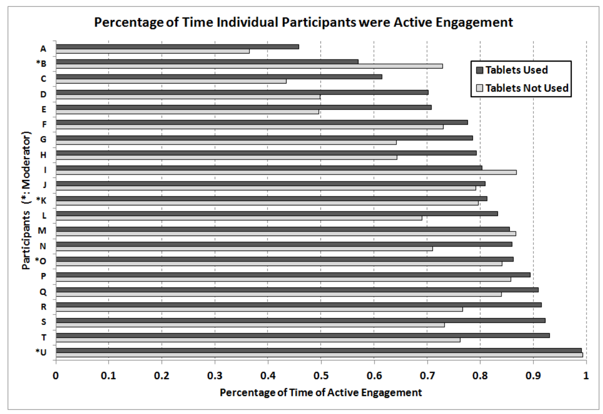

Figure 6: Average "individual active engagement" (γ) per participant (w = 60 seconds)

Figure 6 graphs each participant's average γ over all meetings they attended, grouped into tablet and non-tablet meetings. We make several observations on this figure:

1. The vast majority of students (15 of 17) were more "actively engaged" using tablets. The average student saw a boost to their γ of 8.34% with tablets.
2. Of the students who were least engaged without the use of tablets (participants A, C, D, and E all had $\gamma < 50\%$), all saw a significant boost from the tablets (an average of +17.26%).
3. The moderators were largely indistinguishable from the students. One exception is participant B, a moderator, whose γ dropped from 73% to 57% with the introduction of tablets. For moderators, this drop may be a positive thing, as one may prefer to see more discussion among *students*.

Both forms of analysis resulted in the same trends when graphed, nearly identical raw values, and the same level of statistical significance. Later in this paper, we examine a per-person "active engagement" and find that the moderators were largely indistinguishable from the students.

[3]When students were unable to attend their assigned section due to a conflict, they were allowed to attend other sections. Our analysis used all students present at the section, even when a student only appeared once.

4. Performing a paired, two-tailed t-test on the participants' data graph in Figure 6, we find these results are statistically significant (p=0.000914). If we choose to exclude the moderators, the p-value decreases to 0.0000416.

6. RELATED WORK

The use of the studio concept in programming classes seems to be increasing [Na12]. There is a considerable literature on the use of Tablet PCs in classrooms, some with an emphasis on how they can increase student engagement [An05,DC08,HH08,L08]. We are not aware of any research on integrating Tablet PCs into a studio, nor of analyzing audio recordings to assess engagement. We drew the idea of using individual lapel microphones from "Conversation Clock" [BK07]; in that research, participants in a small meeting employed a system that visually displayed each individual's contribution to the conversation in real time; the researchers studied the subtle effects that this had on the conversation.

7. CONCLUSIONS AND FUTURE WORK

We have described how, in a course in which Computer Science students gather for mutual code reviews, Tablet PCs were introduced to help facilitate and enliven the discussions. Through the recording of ten weeks of weekly meetings across six different sections, we presented detailed analysis of the engagment of the various participants in each section.

To perform this analysis, we established an "active engagement" metric that evaluated how lively or engaging the discussions were in each meeting. This metric is defined simply as how many participants had a "turn" in the discussion within a window of time before the current time. Using this metric, we found the use of tablets, for every reasonable window of time (from 30 seconds to 5 minutes), significantly increased the "active engagement" in all of the discussion sections ($p < 0.001$).

Further, we examined the percentage of time each individual student was "actively engaged" in each of the meetings they attended. We found that, with the introduction of tablets, the large majority of students were more engaged (15 of 17), and, interestingly, the students who were least engaged gained the most in our metric (an average of +17.26%). We found strong significance in the increase in "active engagement" of the participants as a result of the use of tablets (p=0.000914).

In this paper, we did not perform any analysis that made use of the various roles of the participants, such as distinguishing which student was presenting at any time. It would be interesting future work to see how the interactions between the presenter, students, and moderator differ with and without the tablets.

Further, the analysis performed on this audio was done without reference to the topic of conversation. While the meetings are lead by a moderator (a member of the course staff) that kept the discussions largely on-topic, it would be of value to understand if the introduction of tablets aided to keep the discussion on-topic more often.

The quest for increased student engagement is a long one; a Google search for "strategies for increasing student engagement" gets over two million hits. In those circumstances where "active engagement" is a good proxy for increasing student engagement, we believe our work can contribute to studying engagement using more objective measurements.

Acknowledgements

We thank Chris Cortez for his help in running the experiments, and Studio instructors Mike Woodley and Charlie Meyer, and the CS 242 moderators, for their cooperation.

References

[An05] R. Anderson, R. Anderson, L. McDowell and B. Simon. Use of Classroom Presenter in Engineering Courses. Frontiers in Education, 2005.

[BK07] T. Bergstrom, K. Karahalios. Conversation Clock: Visualizing audio patterns in co-located groups. In Proc. 40th Annual Hawaii Intl. Conf. on System Sciences (HICSS '07), 2007.

[DC08] T. L. Derting, J. R. Cox, Using a Tablet PC To Enhance Student Engagement and Learning in an Introductory Organic Chemistry Course J. Chem. Educ., 2008, 85 (12), p 1638

[HH08] A. Hurford and E. Hamilton, Effects of Tablet Computers and Collaborative Classroom Software on Student Engagement and Learning, 38th ASEE/IEEE Frontiers in Education Conference, Saratoga Springs, NY, 2008.

[Kam08] S. N. Kamin, M. Hines, C. Peiper, B. Capitanu, A System for Developing Tablet PC Applications for Education. SIGCSE, Portland, OR, March, 2008.

[P08] P. Lafontant. Using Tablet PCs and DyKnow Vision to Enhance Student Engagement In A Human Anatomy Class. Workshop on the Impact of Pen-Based Technology on Education (WIPTE) West Lafayette, IN. 2008.

[Na12] N. H. Narayanan, C. Hundhausen, D. Hendrix, M. Crosby, Transforming the CS Classroom with Studio-Based Learning, SIGCSE 2012, Raleigh, North Carolina, USA, 165–166.

[Slice] SLICE website: slice.cs.illinois.edu

[Wi01] K. Wiegers, Peer Reviews in Software, Addison-Wesley Professional, 2001.

[WK07] M. Woodley, S.N. Kamin, Programming Studio: A Course for Improving Programming Skills in Undergraduates. SIGCSE, Covington, KY, March, 2007.