International Journal of Technology Enhanced Learning (2011), 3 (4), 336-376.
DOI: http://dx.doi.org/10.1504/IJTEL.2011.041280

Course Quality Improvement using Mid-semester Feedback

Abelardo Pardo, Iria Estévez-Ayres, Pablo Basanta-Val and Damaris Fuentes-Lorenzo

Department of Telematic Engineering,
Carlos III University of Madrid, Spain,
E-mails: {abel, ayres, pbasanta, dfuentes}@it.uc3m.es
www.it.uc3m.es

Abstract: Quality control mechanisms are becoming more important in higher educational institutions. Student evaluation of teaching is typically used to obtain feedback from students about a learning experience but its effect in the course may take too long. Fast feedback mechanisms, in exchange, look at obtaining feedback in a way that corrective measures can be applied quickly. In this paper a process is described to obtain feedback from the students about a course, analyze the received results, and identify the most significant aspects. The process has been applied to a course and led to some adjustments that had immediate impact on the course.

Keywords: student evaluation of teaching, quality improvement, course evaluation.

Reference to this paper should be made as follows: Pardo, A., Estévez-Ayres, I., Basanta-Val, P. and Fuentes-Lorenzo, D. (2010) 'Course Quality Improvement using Mid-semester Feedback', *Int. J. of Technology Enhanced Learning* , Vol. ?, Nos. ?/?, pp.???–???.

Biographical Notes: Abelardo Pardo is an Associate Professor of Telematics Engineering at the Carlos III University of Madrid. He received his Ph.D. in Computer Science from the University of Colorado at Boulder. His research interests are in the area of computer-supported learning, adaptive hypermedia and multimedia content creation. He is member of the .LRN community where he participated in the implementation of the Learning Design support within the platform. He has also participated in several international research projects and is the principal investigator of the *Flexo* project on adaptive elearning platforms.

Iria Estévez-Ayres was born in Vigo, Pontevedra (Spain) in 1978. She has got her Telecommunication Engineering degree at Vigo University in 2001 and her PhD at Universidad Carlos III de Madrid in 2007. Currently, she works at the Telematic Engineering at Carlos III of Madrid as a member of GAST (Grupo de Aplicaciones y Servicios Telemáticos), within the Real-Time Systems lab. Her research interests are in the dynamic and composition of distributed real-time applications, an area where she carried out her doctoral thesis.

Pablo Basanta-Val was born in O Valadouro, Lugo (Spain) in 1978. He has got his Telecommunication Engineering degree at Vigo University in 2001 and his PhD at Universidad Carlos III de Madrid in 2007. Currently, he works at the Telematic Engineering at Carlos III of Madrid as a member of GAST (Grupo de Aplicaciones y Servicios Telemáticos), in the Real-Time Systems Lab. His research interests are in Real-Time Java technologies and general-purpose communication middlewares, an area were he carried out his doctoral thesis.

Damaris Fuentes-Lorenzo holds an M.Sc in Computer Engineering, Development of the Enterprise Information System Specialization from the Universidad Carlos III, Madrid, Spain, and an M.Sc in Computer Science and Technology, specialized in Software Engineering. She has worked as research technician in both the Departments of Telematics and Computer Science, and she has worked as a researcher at IMDEA Networks. Currently, she is involved in several projects and works as teacher assistant while preparing her Ph.D in Telematics Engineering. Her research interests focus on web technologies, semantic web and web usability.

1 Introduction

Under pressure to increase the quality of their services, higher educational institutions are exploring the implementation of total quality management procedures as well as continuous quality improvements. Student Evaluation of Teaching (henceforth simply SET) is typically included among the procedures that are part of quality assurance programs.

However, when trying to precisely assess all the factors that are affecting the development of a course, institutional SET may have disadvantages. The questions are typically about the teacher behavior (e.g. “presents clearly in class”, “makes attempts to clarify the material”, “addresses the students correctly”, etc.). This is because results are usually taken into account for teacher promotion or salary increase processes. Also, the overall feedback process has a cycle larger than the course. Forms are typically filled by the students at the end of the semester, results are sent to the teaching staff, and any possible adjustment is considered for the next course edition.

A feedback mechanism with a shorter latency is what is known as “mid-semester feedback” or “fast feedback” (Angelo and Cross, 1993; Bateman and Roberts, 1993). As the name suggests, this technique consists on obtaining feedback from students at different times during the semester. Results supposedly reach the teaching staff

quickly enough to consider immediate adjustments. However, this type of SET is challenging to deploy because it needs an unobtrusive feedback gathering procedure, quick post-processing time and fast response by the teaching staff. Furthermore, no formal process to analyze the answers is used.

This paper describes a process to implement a quality assurance mechanism in a course. Feedback is obtained using multiple unobtrusive mid-semester student evaluations. Answers are then classified through a subjective categorization process. The most significant aspects of the course are then analyzed for potential corrective actions.

The process was applied to the first edition of a course that included significant changes: the teaching methodology shifted from conventional lecturing to active learning, some learning outcomes were included for the first time, grades were obtained with continuous evaluation paradigm, and a significant number of students were transferred from a similar degree. This scenario needed a procedure to effectively assess the course deployment and propose corrective measures if needed. The analysis of the obtained results have proved that this type of feedback is valuable also for future editions of the course.

Section 2 presents the most relevant work in the area of student evaluation of teaching. The details of the course where the process was used is described in Section 4. Section 3 describes the process. A detailed analysis of the obtained results is done in Section 5, and Section 6 presents some conclusions and future lines of research.

2 Related Research

Quality of services in higher educational institutions is becoming a very important issue. However, implementing measures to achieve this quality is a complex issue to the point that multiple organizations have joined efforts to describe and formalize quality management procedures (Dondi et al., 2007). Classroom assessment techniques (Angelo and Cross, 1993) are designed to gather information about the activities in the classroom so that instructors can know the most effective measures to improve the overall learning process. These techniques are themselves part of a wider area known as "Student Evaluation of Teaching" (SET).

Looking at the large body of literature about this area, the correlation between SET and the quality of a course is, to say the least, controversial (see Wilson, 1998, for an example). Numerous scientific studies praise the benefits of SET to improve a learning experience, and an equally important number of authors cast doubts about its true value. Two conclusions about SET appear consistently in the related literature.

The first one is that teaching effectiveness is a multidimensional construct, and as such, a reliable measure is elusive (Marsh, 1987). For example, Berk (Berk, 2005) presents a survey of 12 different strategies to measure teaching effectiveness, and concludes that student ratings are a necessary source of information specially for formative decisions. The second conclusion is that results obtained with SET highly correlate with other measures obtained from sources such as peers, alumni

or administrators (Felder, 1992). As a conclusion, it can be stated that SET is a reliable measure of teaching performance (Wachtel, 1998).

Student evaluations are typically used for two purposes (Loveland, 2007): for teacher promotion, and to obtain feedback for the teaching staff to improve future editions of the course. The second aspect is the one relevant for this paper. However, most of the SET deployed in educational institutions are usually forms containing a set of multiple choice questions that students answer, either on paper or through a web application, at the end of the semester, and are mainly about the teacher (Lang, 2007). This is because studies have analyzed all the different dimensions affecting a learning experience, and the teacher has been identified as the most important of them (Wright et al., 1997). But due to the numerous aspects influencing a course, student feedback could be more useful if asked in terms of the whole course and not only the teacher behavior. For example, Sadoski (Sadoski and Sanders, 2007) presents a study that identifies the two aspects that influence the most in the overall quality of a course: course organization and clearly communicating course goals and objectives. Although related to the teacher, these aspects point more precisely the areas in which an instructor needs to focus.

A special case of SET are the mid-semester student evaluation, informal early feedback, or classroom assessment (Angelo and Cross, 1993) techniques. These are all techniques related to the concept of fast-feedback in the area of management practices proposed by Bateman *et al.* (Bateman and Roberts, 1993). The idea is to deploy a mechanism by which feedback (sometimes informal) is collected from the students with a frequency that allows the instructors to reflect upon the answers and deploy, if needed, any corrective measures.

Due to the possibility of a quick response, this type of evaluation is specially recommended when teaching a course for the first time, or when there is a significant change in methodology, to detect strength and weaknesses during the semester (Felder, 1995). Additional advantages are the possibility to ask students to provide different type of comments than the ones derived from the end-of-semester multiple choice questions. To increase participation, the process must be kept as unobtrusive as possible (Davis, 1993).

Processing fast-feedback answers does not receive the deserved attention in the literature. Due to the type of the questions used, the suggestions are simply to browse through the answers, take student comments seriously and let students know what, if anything, will be changed. In techniques such as the one-minute paper (Angelo and Cross, 1993), students are asked about the aspects that were more unclear during a lecture, but once the answers are gathered by the instructor, not formal procedure is given to process these answers.

Instructors rarely have full awareness of what is truly happening in a course and are inclined to misread how are students perceiving the course as a whole, thus drifting away from a critically reflective mind (Brookfield, 1995). Fast-feedback answers are a tool to obtain an accurate and reliable measure of the course perception. But when open questions are used for this type of evaluation, and the number of answers is large (more than 20), informal analysis techniques might not reach information that is easily obtained with simple formalized procedures.

This paper describes a process in this context. A mid-semester student evaluation starts the process. The answers are then categorized and results are sorted to identify

those aspects that are having a significant impact in the course. The method to obtain the student feedback was inspired by the classroom assessment techniques described by Angelo *et al.* (Angelo and Cross, 1993), but instead of requesting feedback at the end of each class, the forms were requested three times during the semester in an approach similar to the one described by Felder (Felder, 1995).

A categorization step is then applied to classify the answers and identify those aspects of the course with a significant impact for the students. With this information, the instructors may consider adjustments to be applied to the course.

3 The Process

The proposed process consists of four steps as depicted in Figure 1.

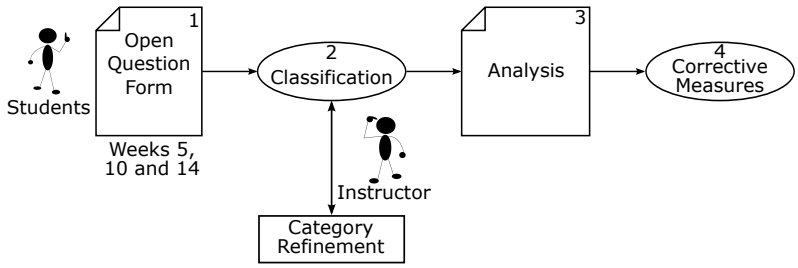

Figure 1 Structure of the process

Step 1: Obtaining Feedback from the Students

The form used in the first step to obtain the answers from the students was obtained as a compromise between the following aspects:

- The questionnaire should be brief and unobtrusive. Students were submitting exercises regularly during the course, so another submission request with too many questions would probably have a low chance of being answered (Davis, 1993).
- Given the course structure and context, there were numerous aspects to be monitored. Any attempt to enumerate them turned into a long list.
- Students should understand the need and purpose of this questionnaire.
- The responses should be simple as to speed up the post-processing.

The resulting form started with a succinct paragraph requesting the feedback and stating that the results would be kept anonymous, followed by two separated text boxes with the question "Describe briefly the most critically positive (respectively negative) aspect that you encountered in the course so far". The form was published as a HTML page in a survey platform with a mechanism to limit each answer to only 300 characters as a compromise between open answers and a simple post-processing procedure. The combination of open questions, requesting only one aspect of the

course, and limiting the answer size force students to briefly analyze all the aspects of the course, choose the most critical one (positive and negative) and state it succinctly.

As expected when using open questions, answers varied in type and length. Some of them focused only in a single aspect of the course, others mentioned several of them in the same sentence, and other mixed positive and negative aspects in the same answer. A sample of an answer for the most critically positive aspect of the course is "the course material is fairly well prepared. Same with the classes. Besides, I like the combination of previous and in class activities". Analogously, a sample of the most critically negative aspect of the course is "The time needed to prepare the previous activities to each class is very high. It takes time away to study the rest of subjects. Also, because of the technical problems in the lab, we could not finish the first activities there and had to do it at home".

Step 2: Categorization

A large number of these answers may mention an equally large number of course aspects, thus a more systematic approach is needed. In the second step, an initial set of categories is proposed corresponding with different "aspects" of the course such as "teaching methodology", "lab organization", "course organization", etc. All answers are reviewed by an instructor and marked as belonging to one or more of these categories. A special "miscellaneous" category is considered to include comments that do not refer to any aspect of the course or are difficult to classify (for example "I joined the course late, and the first day I had to submit an exercise"). Although this classification step undoubtedly introduces some bias, the obtained results highlighted certain aspects of the course so clearly that it was perceived as an improvement over simply browsing the answers.

Step 3: Analysis

At the end of step 2 each answer is labeled with the mentioned course aspects. Then, for each category three measures are derived. The number of positive and negative appearances, referred as the "positive" and "negative" end respectively, and the difference between these two measures. With these numbers the issues are differentiated as follows:

- Issues with both high values on the negative end and the difference need to be analyzed for corrective measures.
- Issues with large positive and negative ends, but with a small difference are labeled as "controversial". They are considered to deploy adaptation measures.
- Issues with large positive end and large difference are ruled as positive.
- Issues with low values in all three numbers are ruled not significant.

Step 4: Corrective Measures

With this informal categorization and analysis, the issues that are most significant for the students are identified. Instructors may now decide if any corrective measures are needed.

4 Empirical Scenario

The described process was deployed in the first edition of the semester course "Systems Architecture", which is part of the degree in Telecommunication Engineering [1] The total number of students that initially signed for the course was 204 and were divided into four groups. Of those, only 166 (81.37%) remained after the drop-out deadline. The course contained the following four learning outcomes:

1. Design and development of applications in the C Programming Language.
2. Use proficiently the tools for application development.
3. Apply team working techniques to develop an application for a mobile device.
4. Use of self-learning techniques.

Learning outcome 1 includes aspects such as being able to analyze a set of requirements, derive a design and create a program with the proper data structures and algorithms to fulfill the given requirements. Outcome 2 is obtained when students use regularly in their development work tools such as integrated development environments, debuggers, cross-compilers, emulators and version control systems. Other learning outcomes similar to these two are present in other related courses of the same degree.

Outcomes 3 and 4 refer to generic methodological aspects. Team work was used during the second half of the course (six weeks) in which groups of four students were created by the instructors to work in a project. Several documents about team dynamics were requested as readings and a class session was devoted to analyze teamwork, agree on a team contract and discuss the different type of conflicts that may arise.

The measures to achieve outcome 4 were applied all throughout the course. Each session had two sets of activities, previous and in-class. The set of previous activities was conceived such that students apply self-learning skills such as information gathering, reflection, self-monitoring, using a strategy, etc. Most of these activities required an objective that would be reviewed in the following class. Students could obtain any help for these activities by posting their questions in a forum hosted in a virtual community, visiting the teaching staff during their office hours, or setting specific appointments with the instructor.

The specific context in which this course took place is worth mentioning. Outcomes 3 and 4 were included for the first time explicitly as learning goals. As a consequence, students found a methodology significantly different to those used in other courses. Also, due to academic circumstances, an important percentage of

students took the course after transferring from a similar degree, and therefore the methodology was a significant change for the resulting cohort.

The course followed a continuous evaluation scheme. Five partial examinations spread along the semester were combined with small exercise submissions. The goal was to engage students to regularly work in the course. The final course grade was simply the sum of all these partial scores; no final exam was given. Most of the students were taking other courses in which previous activities were also requested, and therefore, there was the risk of having a higher than usual overall student workload.

Aside from methodological aspects, the course also had some additional complications derived from outcome 2. The use of development tools, combined with the blended learning approach (students had a significant workload outside of the classroom) resulted in the need of tools properly configured as a development environment. In order to simplify administration related tasks, a VirtualBox[2] virtual machine was created and offered for the students to download. The use of this machine in both personal computers and university labs posed an additional challenge as the obtained results showed.

5 Experimental Results

Figure 2 shows the course structure and the weeks in which the feedback was requested from the students. In all three occasions (weeks 5, 10 and 14 of the semester) the system accepted answers during five days. A message was posted in the course forum stating the importance of the feedback, that the results would be kept anonymous, and inviting students to submit their answers. Instructors also reminded students about the form in regular sessions.

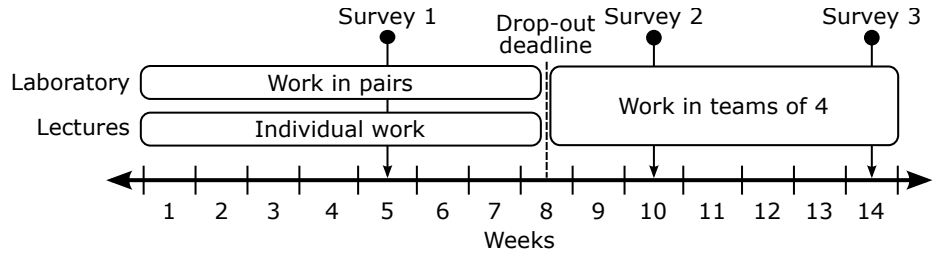

Figure 2 Course structure and survey deployment

Answering the feedback form was optional and the number of received answers in the three surveys were 67 (of 204 students, a 32.84%), 44 (of 166 students, a 26.50%) and 43 (of 166 students, a 25.90%) respectively. The higher sample for the first questionnaire, requested during week 5, is due to the deadline for course dropping (students sign out of the course with no effect in their academic record) which was on week 8 in the semester. The obtained results are shown in Table 1.

The second column is the description for each category. For each of the three surveys, two columns are shown: the number of appearances of the aspect mentioned as negative, and mentioned as positive. An empty cell means that the category

was not mentioned at all in the survey. As it can be seen, the number of aspects mentioned in each survey changes depending on the methodology being used at that moment. The three highlighted rows correspond to the three categories with the higher number of negative appearances.

In the first survey, the course workload was perceived as the most negative aspect by a significant number of students. At the same time, only one student considered the workload adequate. This difference among the positive and negative appearances points to an area that needs to be examined carefully. In the same first survey, the categories referring to the course evaluation policy and student participation in lectures had also a large number of appearance as negative, but in this case, there is also a significant number of students that considered these same aspects as the most positive ones. Another aspect with a significant number of negative answers is the use of the given development tools. Indeed, a set of configuration anomalies in the given tools prevented some students to execute some of the exercises requested in the labs.

The remaining categories in the graph had similar values for the positive and negative end and therefore were considered not significant.

In this second survey, the categories about course load, evaluation policy and student participation have a significant decline in both positive and negative counts, and new categories appear that mention the course project team work.

In the third and last survey only the course load and the evaluation policy maintain values similar to the previous ones, although both of them experience a decrease. On the positive column, two aspects have a significant increase: team work, and the presence of the course project.

5.1 Adjustments Derived from the Survey

After analyzing the results of the first survey, the following measures were decided by the teaching staff and applied immediately in the course:

1. An in-depth review of the workload estimations for the remaining activities.
2. Devote time in class to explain the type of working habits required by the course structure with special emphasis on the previous activities.
3. Reduce the frequency of evaluations in the schedule.

With respect to the technical problems with the tools in the development environment, the teaching staff decided that the measures already deployed would be enough to solve them, so no additional action was taken.

These measures were applied between weeks 5 and 10 of the semester. The results of the second survey show the effect of these adjustments. Although the same three categories collect the higher number of negative appearances, their magnitude have decreased more significantly than the number of received answers. Furthermore, the deployment of these adjustments was itself perceived as the most positive aspect of the course by six students. This result shows how the proposed process may also capture the effects of the measures derived from previous observations.

N	Aspect	Week 5		Week 10		Week 14	
		−	+	−	+	−	+
1	Use of mobile devices in the labs	0	4			1	3
2	Student participation in lectures	24	14	5	5	3	1
3	Number of exercises solved in class	4	4	3	1	2	0
4	Course workload	29	1	9	7	7	3
5	Study new programming language	0	6	0	2	0	2
7	Use of course forum	0	3				
9	Course evaluation policy	10	13	8	3	5	0
10	Motivation received	0	3				
11	Laboratory infrastructure	4	1				
12	Teaching staff in the laboratories	5	4	1	0	3	2
13	Given development tools	8	1			1	0
14	Material written in English	1	0	2	1		
15	Organization of course material	1	7	0	1		
16	Availability of the instructors	0	4	0	2	0	1
17	Work in pairs in the lab exercises	1	0				
18	Work scheme in the lab sessions	0	5	1	2		
19	Exercise resolution in groups	0	1	1	2		
20	Documents on team work			1	1		
21	Team work as course outcome			1	4	3	11
22	The course project			1	5	0	14
23	Changes due to comments			0	6	0	1
24	Group creation criteria			1	3	2	0
25	Project Presentation session					1	4

Table 1 Results after the categorization step

In order to check the validity of the results obtained with this process, an interview with student representatives was held at the same time the first survey was being answered. The issues covered in that meeting were totally consistent with those shown by the survey: course workload was perceived as excessive, previous activities were not accepted, the evaluation scheme could be improved and the technical problems should be solved. This coincidence increased the perceived reliability of the process by the teaching staff despite the potential bias introduced with the classification step.

6 Conclusions and Future Work

In this paper a process to deploy a fast-feedback mechanism to improve course quality has been described. The obtained results showed valuable qualitative information about the course and prompted measures that were deployed immediately. The impact of these measures was also reflected in the results obtained by repeating this process during the semester.

By extending conventional techniques to obtain feedback before the semester end with a categorization step, instructors can easily identify those aspects in a course that are perceived by the students as most positive and negative. The classification procedure, although subjective, helps to handle a potentially large number of answers about numerous course aspects. The process has been used in a course, the structure of which included a significant amount of changes. Answers were obtained from the students using a simple, unobtrusive, open-ended based form. Answers were limited in size to encourage clarity and facilitate post-processing.

As future lines of work, we are currently exploring to increase the level of automation in the classification step by using techniques such as latent semantic analysis (Wild et al., 2005).

Acknowledgments.

The authors are truly indebted to Miguel Valero for his guidance during the design and deployment of this process. Work partially funded by the Learn3 project, "Plan Nacional de I+D+I TIN2008-05163/TSI", the Best Practice Network ICOPER (Grant No. ECP-2007-EDU-417007), the Flexo Project "Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica" (Ref. TSI-020301-2008-19), and the "Emadrid: Investigación y desarrollo de tecnologías para el e-learning en la Comunidad de Madrid" project (S2009/TIC-1650).

References

Angelo, T. A. and Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers (Jossey Bass Higher and Adult Education Series)*. Jossey-Bass.

Bateman, G. R. and Roberts, H. V. (1993). TQM for professors and students.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1):48–62.

Brookfield, S. (1995). The getting of wisdom: What critically reflective teaching is and why it's important. In *Becoming a Critically Reflective Teacher*. Jossey-Bass.

Davis, B. (1993). *Tools for Teaching*. Jossey-Bass, San Francisco.

Dondi, C., Hildebrandt, B., Husson, A.-M., Moretti, M., and Pawlowski, J. (2007). Workshop on learning technologies project team e-learning quality. Technical Report ICS 35.240.99, CEN.

Felder, R. (1992). What do they know anyway? *Chemical Engineering Education*, 26(3):134–135.

Felder, R. (1995). We never said it would be easy. *Chemical Engineering Education*, 29(1):32–33.

Lang, J. (2007). Did you learn anything? *The Chronicle of Higher Education*, 53(27):C1.

Loveland, K. A. (2007). Student evaluation of teaching (SET) in web-based classes: Preliminary findings and a call for further research. *The Journal of Educators Online*, 4(2).

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International journal of educational research*, 11(3):253.

Sadoski, M. and Sanders, C. W. (2007). Student course evaluations: Common themes accross courses and years. *Journal of Medical Education Online*, 12(2).

Wachtel, H. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, 23(2):191.

Wild, F., Stahl, C., Stermsek, G., and Neumann, G. (2005). Parameters driving effectiveness of automated essay scoring with lsa. In *Proceedings of the 9th Computer Assisted Assessment Conference*.

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education*, 44(19):A12–A14.

Wright, P. S., Horn, S. P., and Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1).

Notes

[1] `http://www.it.uc3m.es/labas/syllabus_en.html`
[2] `http://www.virtualbox.org`