Matute, H., & Vadillo, M. A. (2007). Assessing e-learning in web labs. In L. Gomes & J. García-Zubía (Eds.), Advances on remote laboratories and e-learning experiences (pp. 97-107). Bilbao, Spain: University of Deusto.

Assessing e-learning in web labs

Helena Matute, Miguel A. Vadillo
Psychology Department, University of Deusto. Bilbao, Spain
e-mail: matute@fice.deusto.es – http://www.labpsico.com

Abstract

Remote Labs (also called WebLabs) allow remote access to expensive equipment, and allow for many different people, from different countries and time zones, to share the same facilities at different times, through the Internet. A second aim of web labs is to to develop low-cost variants of web labs for education facilities in developing countries. But one of the main problems that researchers face when trying to develop a web lab, is that of trying to convince institutional and research boards that web labs are useful, efficient, attractive, and, in sum, worth of financial support. There seems to be no doubt that web labs are cheaper than traditional ones; if they were shown to be at least as efficient, research funds will certainly be available for them. Although researchers have often conducted survey studies to assess the users' opinions and attitudes towards web labs, these studies fail to provide scientific and unambiguous support. In this chapter we provide some suggestions on how to use the scientific method to assess the effectiveness of web labs.

1. Teaching vs. research labs

Before we start, let us first clarify that web labs can be understood, and classified, in two main groups: those dedicated to research and those dedicated to e-learning. Nowadays, most engineering web labs are of the second type, but in other disciplines, such as psychology, web labs are equally used for research purposes (see [1] [2]). In the case of research labs, there are obviously many problems in using web labs instead of traditional labs, but there are also many advantages. We have discussed these matters, as applied to psychology web labs, in other publications (see, e.g. [3] [4]; for reviews, see also [5], [6], [7]). Although the differences between engineering and psychology labs are obvious, some of the problems that we encountered and some of the solutions that we proposed could perhaps be of use in the engineering domain as well. As an example, we are finding that using the web to replicate classical, well-known, experimental results in psychology is a critical first step that should be conducted before trusting the web lab as a tool to generate new (and reliable) knowledge. It is possible that readers interested in research engineering web labs may find that some of the ideas developed when implementing a psychological research web lab can be of use to them as well. In the present chapter, however, we will focus on how to assess e-learning in a web lab. That is, we will address the learning issue in the engineering web lab from the point of view of research on the psychology of learning. Indeed, the method used to assess that learning is taking place should not differ as a function of whether the learner (or student or user) is learning grammar or psychology or engineering; The learning process is a general mental process that does not change as a function of what the content of this learning is. It is a mental process that works the same way regardless of whether it is taking place in the web or in a classroom or in the forest. Thus, the important question for our present purposes is simply how to assess that learning has occurred.

Of course, once we are certain that learning has taken place in the web lab, we will also need to perform some additional tests in a real lab in order to make sure that students are able to transfer their competencies to a real lab. Practical lab work is an essential part of engineering education and one argument stressed against web labs is that the students do not learn to use the real instruments, but only web interfaces. Performing some of the tests we will mention below in a real lab will be necessary if we would like to address this concern as well.

2. Measuring the learning process vs. measuring the learning outcomes

Most often in the educational system, satisfaction and attractiveness are being assessed instead of learning. Appreciate, however, that students may find extremely attractive an easy going class in which they do not need to work much, and this, of course, does not necessarily mean that they are learning what they are supposed to learn. At best, subjective evaluations on how much students *believe* they have learned are sometimes used. This can be quite misleading. It is not satisfaction or personal

opinions, but learning itself (i.e., knowledge and competency acquisition) what we are supposed to assess. Moreover, another problem is that when effectiveness rather than attractiveness is assessed, most often learning is assessed only after the training period is over (e.g., end-of-semester exams). In that case, what we assess is the product of learning, but not the way in which it occurs. In other words, we can assess at that time whether learning has taken place, but not whether it proceeded slowly, continuously, gradually, and so on. It is also true, most often this is the best we can do: we usually have no resources to record the students' progress at all times. But when we design experiments to assess learning in a psychology laboratory we usually need to assess, in addition to the final product of learning, the way in which it occurs; or, in other words, the variables that conform the learning curve. That is, we need to assess those variables that allow us to see the progress of learning (how does learning proceed in one task or another). These variables are those that should show improvement as learning progresses. Certainly, one of the main advantages of web labs is that they allow the teacher to automatically record the students' performance as they learn. This provides an extraordinary advantage over traditional systems, as we can construct a learning curve, which can be most informative of the details of the acquisition process.

Figures 1 and 2 show two examples of learning curves. Figure 1 shows the mean time that students require to complete Task A as a function of the number of training sessions. As they learn more, the time they require to complete the task is reduced. Figure 2 shows the same students performing the same fictitious task, but the dependent variable in this case is the number of problems solved per session, again as a function of the number of training sessions. Note that, although opposite in shape, both the slope and the height of these two curves are highly informative of the learning process and can certainly be used to compare how students learn in one or another situation. Of course, the easiness with which a learning curve can be recorded in a web lab, as compared to more traditional labs or classes is a key advantage that can certainly be used as a strong argument in favor of web labs. Traditional teaching techniques provide very few opportunities to assess the learners' performance while training takes place, whereas in web labs each student's activities can be recorded automatically. However, this is not enough. Even if we can argue that in this way we can assess the whole learning process as it occurs, which is great, we will still need to offer a comparison showing that, at the end, our web-lab students learn more. The next section focuses on this point.

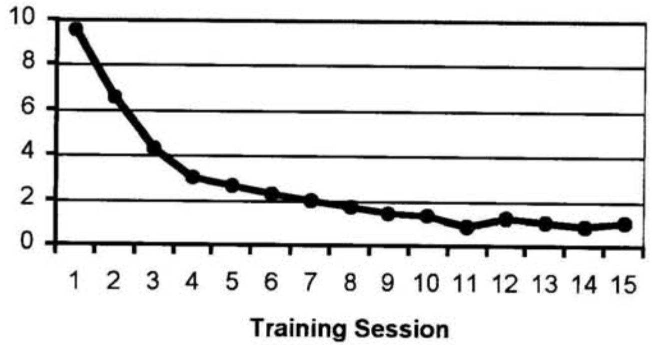

Figure 1

Time to complete a given task is one of the most widely used dependent variables to assess that learning is taking place. As learning proceeds time decreases

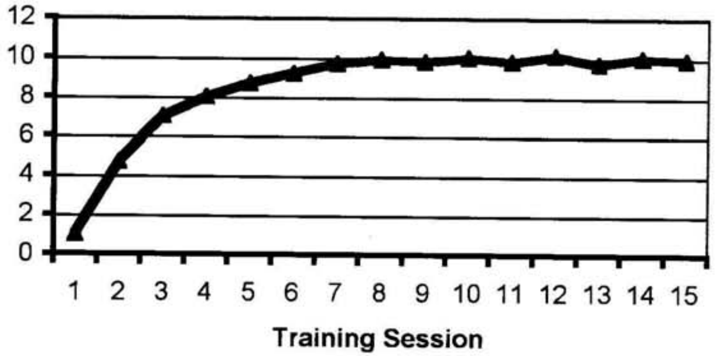

Figure 2

Another dependent variable that is widely used to assess the process of learning is the number of problems that the student is able to solve per session (or per unit of time). In this case the curve increases as learning proceeds

3. How to design an experimental study

Below we suggest some additional ideas also borrowed from the way in which we usually design learning experiments in the field of the psychology of learning. We believe that this can probably be helpful when assessing e-learning in engineering web labs as well. In brief, the problem of assessing the effectiveness of any specific learning (or e-learning) technique is not different from that of using the scientific method to test whether the probability of a given outcome (in this case, learning) is higher or lower when we use that technique as compared to when we do not use it (all other things being equal). As a very first and very general step, any textbook on scientific methods should be of much help in suggesting many ideas and experiments in addition to (or instead of) the ones we propose here (e.g., see [8]). Nevertheless, it might be helpful to highlight some of the most critical points.

3.1. *Defining the dependent variable*

The first thing we will need to do is to clearly define what we want to teach and what we want to assess. Thus, we need to know exactly what we mean when we refer to the "effectiveness" of a web lab (or any other learning tool). Does "effectiveness" mean that students are able to solve this or the other exam after spending x hours in the lab? Or does effectiveness mean being able to make less than x number of errors in a certain task? Or is it being able to develop a complex and independent project in a certain amount of time? Does effectiveness mean for us attractiveness, so that students will report that they are happy with the lab and the professor when they participate in those opinion surveys that all universities run at the end of each semester?... Whatever our dependent variable means for us, we should be able to define very, very clearly. We should be able to spell it out in the form of: "By the end of this course, students should be able to..." Moreover, we should be able to specify how exactly we are going to assess that the students have met those criteria: What type of exam and/or projects do the students need to pass so that we can agree that they have met the objectives we set for the course? This is not something special for the case of web lab's assessment; it is just something very general that should be specified even if we use a traditional lab (or no lab at all) and would like to know how efficient this technique is. There is nothing new so far.

But there are more things, in addition to this one, that we are supposed to test when we introduce a new, innovative method, such as a web lab. Once we know that our students have learned something, the next questions are: How much did our teaching technique contribute this outcome? Was it better than alternative methods? Thus, once we have clearly defined our dependent variable or variables, we will probably want to decide what our comparison (i.e., control) condition or conditions should be. As an example, we may want to assess the same variables both in the remote lab and in the traditional lab in order to be able to compare their effectiveness;

otherwise we will have no comparison to offer, and convincing institutional boards that our method is better (or at least similar) to the old one will be impossible.

3.2. *Control groups... or how to discard alternative explanations*

One of the very first things we need to keep in mind when assessing a certain technique (our experimental condition) is that we will always need to assess its effectiveness against a control (or comparison) condition. Our control group must be identical to our experimental group in all variables except for the one we are trying to assess (in this case, the teaching method). This is important because otherwise we will not be able to know which variable is responsible for the results that we observe. Ideally, a single class of students from the same country and university and sharing the same teacher can be divided into two homogeneous groups of students (as similar as possible in factors such as I.Q., academic performance, and all other variables that could affect the results); then each of those groups can be exposed to two different teaching strategies: one using a web lab, the other one, the control group, using a different teaching strategy. Among the many teaching strategies that we could use with the control group, we will quite probably prefer a traditional lab, so that we can provide identical problems for both groups to solve and we can then compare the time they need to solve them, the number of problems they solve per session and so on. In principle, this could do.

There are three possible outcomes for this experiment, and the one that is most frequent and easy to get, like in any other experiment, is the null result. That is, the result that shows that there is no difference between groups. In our case, designing an experiment in which we just test Lab A against Lab B and we observe that there is no difference between them will be just too easy. The problem with showing that there is no difference (i.e., that both types of labs are equally effective) is that this result can be attributed to too many different variables. It can be due, for instance, to a lack of sensitivity in the variables we chose as dependent variables. It can also be due to lack of statistical power in that we used a sample that was too small for the type of effect that we were expecting to show. Or there could be a ceiling effect. In sum, null results are not informative. The best thing to do would be to add a few other groups to this study so that we are sure we will avoid the null result. But of course, there is always the possibility that even if we decide to start by running this simple, two-groups, study, we do not get a null result and we can, instead, show that our web lab is better (not equal) to our traditional lab. Again, this could do, at least in principle.

But we should be aware that even if we are able to show that our web lab is more efficient that our traditional lab our critics will still have some arguments against our study. Perhaps the only thing we are showing when we show that our web lab works better than our local lab is that we have a very poor local lab. This design does not provide evidence that our local lab is effective. The possibility exists that our traditional lab is so bad that students learn nothing in it. In sum, a third group will almost always be needed in which students not exposed to any type of lab (but equivalent in all other respects and receiving all other classes) are being assessed on the same problems. This would allow us to test whether the use of any type of lab in our university (whether remote or local) has any beneficial effect on our students and also whether one type of lab training is better or worse than the other.

Finally, there is at least one more alternative explanation that we will possibly need to discard. No matter how nice the results of the above experiment turn out to be, there is always the possibility that providing students with a computer program in which the experiments can be simulated results as effective (and certainly cheaper) than the web lab. We will not be able to discard this critique with the above study alone. If we also want to discard the idea that simulations are equally effective, a fourth group of students using the best possible existing virtual laboratory (simulation) software will be needed in our study... It might happen that learning through simulations is slower, or it might prove to be equally effective. However, the results of real world experiments (remote or local) differ sometimes very much from the results of simulation experiments (whether local or web-based). As some have argued, for engineers, which use both in their daily work, it is necessary to know these differences. Thus, even if they were equally effective, a web lab could be shown to be quite efficient if it proved superior than a local lab in making students understand the differences between simulations and real world experiments.

3.3. *Generality*

It should also be taken into account that assessing the effectiveness of a particular web lab against a particular traditional lab and against a particular group of students exposed to no lab condition but in the same university might be something too specific and with little ecological validity (poorly generalizable). If we wanted to use the results of the study to impress research committees and to advance the future of web labs, then several experiments from several different laboratories should converge on the same results. Alternatively, several researchers from different universities and countries could agree on a common experiment to jointly test the effectiveness of web labs in general. This ideal experimental study should include a large sample of web labs, local labs, no lab conditions, and virtual (simulated) labs (though in this later case, and assuming there is enough consensus among researchers as to which one is the best possible simulated lab software, it might make sense to use just the best one, instead of a large collection of simulation programs This would certainly be a critical piece of research for the future of web labs. Figure 3 shows a fictitious outcome of this ideal experiment.

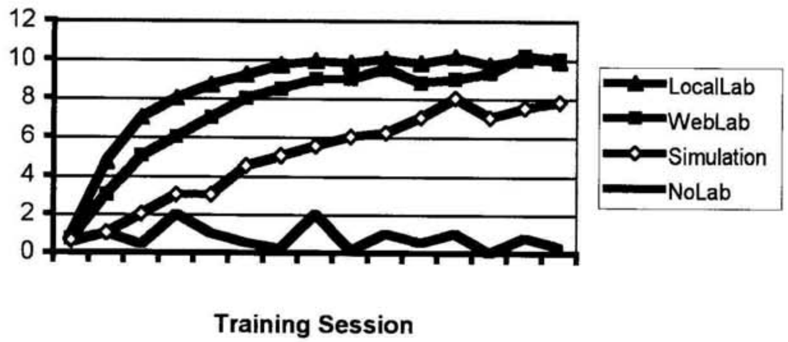

Figure 3

Fictitious outcome of an ideal experiment in which the number of problems solved per session in each of ten participating web labs is averaged and compared to the mean number of problems solved in the corresponding control groups (local lab, no lab, and simulated lab condition) in each of the participating universities. In this example, learning in the averaged web lab is slightly slower than learning in the averaged local lab, but equally efficient after a certain number of training sessions. Moreover, statistical analyses should show that the average web lab is significantly better than the averaged no lab condition and the averaged simulated lab condition.

4. Ethics

There is a matter of ethical concern which we have not yet mentioned but which is closely related to the type of experimental design we have proposed in this paper to assess learning. It is not a trivial matter, as it is something often rose by reviewers of grant proposals. Imagine, for example, that we have developed a new drug that improves people's memory. In order to test the effectiveness of the drug we need to conduct an experiment, giving the drug only to half of the participants and then looking for differences between the memory of those who took the drug and the memory of those who did not. The problem is that some would regard this experiment unethical because it implies giving access to our drug only to half of the volunteers of the study, instead of directly giving it to everybody, so that they all can benefit from the potentially good effects of the drug. Similarly, if we have a web lab that we suspect that will provide enormous educational advantages over older methods, wouldn't it be much more ethical to offer access to it to all our students? Many reviewers will say our proposal is unacceptable!

Our response to this criticism, however, should be clear. If you decide not to run the experiment just because you are sure that the new method *has to* work, then it is trial-and-error what you are doing with your subjects. Every new generation of students (or patients) will suffer the consequences of potential errors in your method. Testing your innovative methods in just one part of your population in order to know, by the end of the experiment, which method is best, is certainly a lot more ethical practice than testing it in the whole population and not even caring about scientifically demonstrating its effectiveness. This being said, however, applying it to the whole population is what you should pursue by all means once you finish your study… *if* you have been able to demonstrate that your method is better than the one at use. You are ethically obliged to do this once you know your method works. Behaving this way is certainly much more ethical that applying the untested method to all possible participants without even caring to run a controlled, experimental test of its consequences.

In addition to this specific problem concerning the use of control groups, there are also many other ethical issues that need to be considered when designing research with humans. Very strict and clear ethical regulations apply in most countries concerning human research. These codes are quite similar to each other and their only purpose is to protect those people that serve as subjects in our experiments (e.g., see [9], [10]). Moreover, there is an increasing international consensus that, in addition to the rules concerning human research in general, research conducted through the Internet presents some peculiarities and should also be subject to additional ethical recommendations (see [11]). These recommendations include things such as how to treat the data of Internet research subjects, which should be absolutely anonymous and voluntary, and how to make sure that the participant has read and understood the information regarding the purpose and method of the study before accepting to send his or her data to the experimenter. It is important to take this into account because it means that we cannot use any type of software to get personal or other type of data from the students without their consent. This means that some type of data which are often collected just because it is customary to do so (e.g., gender, age, name…)

should simply not be included in the design of our study (unless they are very important with respect to the study we need to perform, and, in this case, we will need to ask the students directly to please provide these data –and be aware that their response is not necessarily true).

Concluding comments

We hope to have shown that there are no fixed rules on how the scientific assessment of web labs should look like. The number of groups that will be needed for the study should be carefully discussed beforehand, and the variables clearly defined. The experiments will always need to be designed as a function of what our question is, what we would like to conclude from the study, and as a function of the type of alternative explanations that we would like to discard. As shown in one of the examples above, if we want to show that both web labs and local labs are effective as compared to no labs, but we want to also be able to discard the critique that a good simulated lab computer program would have done as well as our web and local labs, then we need to include such simulated lab condition in our study. In other words, the most important point when deciding how to design the experiment is to make sure that we are using the right groups as a function of what we want to conclude from the experiment and the critiques that we would like to discard.

The other thing that should have become quite apparent in the preceding paragraphs is that one experiment will probably not be enough. For each study that a scientist might publish to demonstrate that Treatment A is better than Treatment B, critics of this idea will be able to publish several other studies in which, looking at other variables, or including different controls, they could reach different conclusions. Thus, it is important to make sure that our experimental study addresses, beforehand, as many of those possible critiques and alternative explanations as possible. If this requires using so many control groups that the experiment will become unviable, then we should consider the option of running several concatenated, smaller studies. When reported together, they should be able to tell a whole, consistent, solid story about the effectiveness of web labs.

References

[1] Birnbaum, M. H., & Wakcher, S. V. (2002). "Web-based experiments controlled by JavaScript: An example from probability learning". *Behavior Research Methods, Instruments, and Computers*, 34, 189-199.

[2] McGraw, K.O., Tew, M.D., & Williams, J.E. (2000). "The integrity of web-delivered experiments: Can you trust the data?". *Psychological Science,* 11, 502-506.

[3] Matute, H., Vadillo, M. A., & Bárcena, R.(in press). "Web-based experiment control software for research and teaching on human learning". *Behavior Research Methods.*

[4] Vadillo, M. A., Bárcena, R., & Matute, H. (2006). "The internet as a research tool in the study of associative learning: An example from overshadowing". *Behavioural Processes*, 73, 36-40.

[5] Birnbaum, M. H. (Ed.) (2000). *Psychological experiments on the Internet.* San Diego, CA: Academic Press.

[6] Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). "Should we trust web-based studies? A comparative analysis of six preconceptions". *American Psychologist,* 59, 93-104.

[7] Reips, U. -D., (2002). Standards for Internet-based experimenting. *Experimental Psychology,* 49, 243-256.

[8] Myers, A., & Hansen, C. (1997). *Experimental psychology. 4th edition.* Pacific Grove, CA: Brooks/Cole.

[9] American Psychological Association (2002). "Ethical principles and code of conduct". *American Psychologist,* 57, 1060-1073.

[10] Colegio Oficial de Psicólogos (1993). *Código deontológico del psicólogo.* Madrid: Colegio Oficial de Psicólogos. [Available at http://www.cop.es/vernumero.asp?id=7]

[11] Frankel, M.S., & Siang, S. (1999). *Ethical and legal aspects of human subjects research on the internet. A report of a workshop.* Washington, D.C.: American Association for the Advancement of Science. [Available at http://www.aaas.org/spp/dspp/sfrl/projects/intres/main.htm]