A new Learning Environment based on Reproducible Ubiquitous Computing: Experiences and Prospects

Patrick Wessa, Leuven Institute for Research on Information Systems (LIRIS), Catholic University of Leuven, Belgium

Ian E. Holliday, Neurosciences Research Institute, School of Life and Health Sciences, Aston University, UK

Peter Reddy, Centre for Learning Innovation and Professional Practice and Health and Human Development Research Group, Aston University, UK

Abstract

In this chapter we explain why Reproducible Computing is important and how it can be used as a socially constructivist learning technology with characteristics of ubiquitous computing. Based on reported experiences and objective measurements from two different statistics courses, we present strong evidence that student perceptions and experiences are very positive towards this new type of technology and educational approach. The most remarkable observation is that the approach was well-received notwithstanding the fact that students initially perceived the workload to be very heavy. The few criticisms that were formulated by students have no solid ground and may be solved through better and more extensive communication.

1. Introduction

Within the context of our newly developed technology for Reproducible Computing (henceforth RC) we developed a learning environment which supports a socially constructivist and (potentially) ubiquitous approach to statistics education. The implementation of this learning environment has been, without a doubt, an interesting and challenging experience for developers, instructors, and students alike. Our reflections about these experiences and the underlying design principles of the RC-based courses might therefore provide useful information for anyone with an interest in ubiquitous learning environments which are based on computer-assisted learning and which have been embedded in the pedagogical paradigm of social constructivism.

In this chapter we outline some of the ubiquitous characteristics of RC technology and explain why they are important within the context of socially constructivist learning. The main focus however, is on the learning activities and experiences of students which have been collected through a series of surveys, several focus panels, and objective measurements as derived from log files or databases. In spite of the fact the learning process is perceived to involve a heavy workload, it is clearly demonstrated that the overall learning experiences are very positive.

Our final thoughts focus on some of the future directions of RC technology which allow us to integrate reproducibility in applications that are not exclusively focused on statistics education. For instance, it is described how this novel technological innovation can be used to obtain an increase in productivity when supervising Master/Ph.D. dissertations or when RC is used within the context of educational games.

2. Background

Within the context of computer-assisted and mathematical education, the pedagogical community has shown great interest in the role and importance of social and individual constructivism (Von Glasersfeld, 1987; Smith, 1999; Eggen and Kauchak, 2001) and its implementation in statistics education in particular (Mvududu, 2003).

While the relevance of the constructivist pedagogical paradigm is well documented there seems to be no direct or obvious relationship with the problem of irreproducible research. Nevertheless, the problem of our inability to reproduce statistical computations that are presented in papers has received quite a bit of attention within the statistical computing community. The most prominent citation about the problem of irreproducible research is Claerbout's principle: "An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and that complete set of instructions that generated the figures." (de Leeuw, 2001, p. 1).

Statistics education is often based on course materials and articles that suffer from the same irreproducibility problems. From a pedagogical point of view, we simply cannot expect students to learn from (statistical) results in course texts, if they are barely able to reproduce – let alone verify – the results from peers, educators, or researchers. Even though, the pedagogical paradigm of social constructivism has often been propagated as a promising approach to achieve true learning, it is unclear how students should be empowered to engage in non-rote learning through interaction (with the subject under study), experimentation, critical thinking/reflection (through peer review), and other social activities that help them to construct knowledge. All of these aspects are impossible to achieve in a traditional course environment because the data, the software, the parameters, and the meta-information about the assumptions that underlie the statistical analysis cannot be communicated in a printed course text or in a traditional e-learning environment.

As one can imagine, the problem of irreproducible empirical research has caused a lot of concern within the statistical computing community. The fact that scientific findings (based on empirical analysis) are difficult – if not impossible - to reproduce and verify, has raised many questions about the status of science. Some of the leading arguments can be found in Peng, Dominici, and Zeger (2006); Schwab, Karrenbach, and Claerbout (2000); Green (2003); Gentleman (2005); Koenker and Zeileis (2007); Donoho and Huo (2004). Several approaches to solve the problem have been suggested and implemented. Some of the more promising attempts are based on the concept of an electronic "Compendium" and have been described in Buckheit and Donoho (1995); Donoho and Huo (2004); Leisch (2003). The proposed solutions are elegant (from a technological point of view) but very hard to implement in an educational setting due to a variety of reasons (Wessa 2008).

More importantly, mentioned solutions do not have the ability to support ubiquitous computing or learning because they assume that all components that are required to make the statistical computation “reproducible” are bundled into one object (a so-called Compendium). More precisely, the Compendium is defined as a *collection of Text, Data, and Software that allows the reader to reproduce the research that is presented in the document* (Wessa 2008). According to this definition however, it is impossible to invoke the functions to reproduce a statistical computation from within small (or “ubiquitous”) devices because they are often not capable of storing all the meta-data about a computation – let alone that the device would be able to perform the actual computing. For instance, it is unclear how a traditional Compendium would be used on a cell/smartphone if it is required to have a working LaTeX compiler and R interpreter.

Ubiquitous learning does not only involve (mobile) devices – it is primarily related to the fact that the student is able to perform learning activities within various situations or contexts. In one context, the student may use a netbook computer or a smartphone on campus and during the lectures to verify the statistical computations which are demonstrated by the instructor. In another context, the student may do some homework at a home-based desktop computer while using VLE forums, instant messengers, and e-mail in order to collaborate with multiple friends at the same time. The use of Compendiums can support learning in this type of environment but are limited in application if their design requires local resources that mobile devices cannot support. If the Compendium would only contain references to remotely stored computational objects and if these could be re-computed on remote machines, then there would be (almost) no limitation to the environment or situation in which it could be used or communicated.

Because of the above reasons, it was necessary to design a new and innovative technology, the so-called Compendium Platform (henceforth CP), which was based on a new (revised) definition of the Compendium: *a document with (open-access) references to (remotely) archived computations (including Data, Meta-data, and Software) that allows the reader to reproduce, and re-use the underlying analysis*. Unlike the old one, this definition allows us to specify new (and “ubiquitous”) environments in which reproducibility of research results is guaranteed work because there is no requirement to store or compute anything locally – the only requirement is to have an active internet connection (Wessa 2008).

The CP solves the aforementioned problems because it allows any user (the educator or student) to quickly create electronic documents that contain statistical research results that are fully reproducible and reusable. The reader of a Compendium can simply click on a table’s or graph’s caption (which contains a hyperlink) to view and use the meta information that is connected to the research results in a remote and freely accessible repository which is hosted at http://www.freestatistics.org. The reader can re-compute the analysis in real-time (at http://www.wessa.net) and change any aspect of the underlying analysis. The data, the parameters, and even the underlying source code can be changed at any time (Wessa 2009a).

3. Ubiquitous Learning

3.1. Peer Review and Social Constructivism

Even though our RC learning environment has been tested & implemented in more than a dozen courses, we limit the description in the remainder of this chapter exclusively to two cases of particular interest. The first case is an advanced statistics course for undergraduate psychology students (+130) at Aston University, UK. The second case is a statistics course for undergraduate business students (131) and graduate students (154) who are required to complete a preparatory year before they are allowed to enter the Master programme at one of the Business Schools of the K.U.Leuven Association.

Both courses covered a variety of statistical techniques and had a strong emphasis on social constructivism. For each type of analysis, students had one or several web-based software modules available.

In order to implement these courses within the setting of social constructivism for large student populations, it was necessary to impose a strict assignment-review mechanism. This is illustrated in Figure 1 which shows a series of weekly events (lectures, assignments, and reviews) during the course period (at least one semester). The course ended with a final examination about a series of objective multiple choice questions in an attempt to test understanding of statistical concepts rather than rote memorization.

The main sections of the statistics course were built around a series of research-based workshops that require students to reflect and communicate about a variety of statistical problems, at various levels of difficulty. Each workshop contained questions about datasets that are of interest to the student population. The business students were also required to perform various types of analyses about individual data series. In both cases, the workshops involved social interaction (in the form of Peer Review) and a substantial amount of individual work.

Each week there was a lecture which was held in a large lecture hall that was equipped with projection and internet facilities. During the week, students were required to work on their workshop assignment and – at the same time – perform peer reviews about a pre-specified number of randomly assigned submissions from peers. Each review was based on a rubric of a minimum of four criteria and required students to submit rubric scores and feedback messages.

The scores that were generated by the peer review process did not count towards the final grade of students whose work was peer reviewed because there should be no penalty for experimentation and creativity. Instead, the educator graded the quality of the verbal feedback messages that were submitted to other students. The grading was performed based on a sampling technique which allowed the educator to grade the quality of a relatively small – but fairly representative – number of submitted feedback messages from each student (Wessa and De Rycker 2010).

This feedback-oriented process is similar to the peer review procedure of an article that is submitted to a scientific journal. The key idea behind this constructivist environment is that students play the role of an active scientist who investigates problems, presents solutions, and reviews the work of peers. For obvious reasons, RC is a *conditio sine qua non* that allows students to engage in such peer review activities.

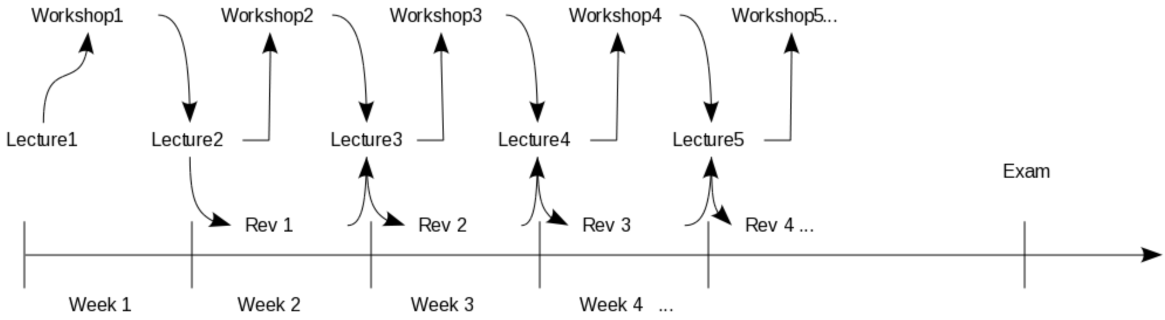

Figure 1. Peer Reviews about Weekly Workshops

From the above description it can be concluded that students are required to use RC in various situations and contexts. As pointed out before, this is only possible if the underlying technology supports ubiquitous computing by design; this is briefly explained in the next section.

3.2. Ubiquitous Technology for Reproducible Computing

The R Framework is a key component of the RC learning environment (see Figure 2) and allows educators and scientists to develop new, tailor-made statistical software which are called "R modules" and are based on the R language (R Development core team, 2010). Educators may browse the collection of available R modules at the R Framework's website (http://www.wessa.net) and use them for the purpose of generating research output that can be used to explain statistical techniques and concepts in their courses. If the educator has a working knowledge of the R language then new R modules can be created and published.

There are some unique features of the Compendium Platform that are of particular importance:

- any computation that is created within the R Framework can be easily "blogged" (i.e. archived) in the repository of the CP – there is no need for students to keep track of the data, the model parameters, or the underlying statistical software code;
- any user who visits the unique URL of an archived computation is able to instantly reproduce the computation or reuse it for further analysis - only an internet connection is required to use the repository;
- educators and researchers are able to retrieve data for research purposes.

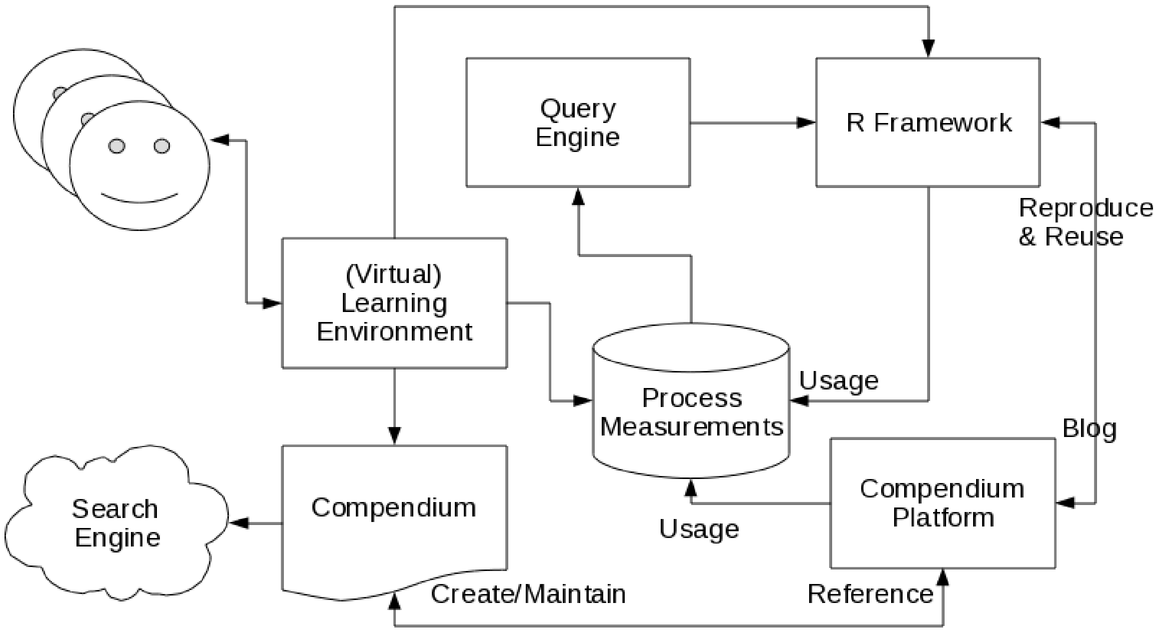

Figure 2. Reproducible Computing & the Compendium

A detailed description of the components in Figure 2 is available in Wessa (2008).

RC has many ubiquitous characteristics but some of the more important ones (from an educational point of view) are the following:

- no downloads are required and all computations are performed on a distributed network of remote servers;
- meta data are stored and maintained in an online, remote repository;
- interactions between computations are preserved/stored and can be used to offer customized information about similar or related computations;
- the user interface is based on html/xml and can be used on small-screen devices with low bandwidth. However, a minimum screen resolution may be required if one wishes to display charts;
- all usage data (creating, editing, reproducing or viewing computations) are available for the purpose of customization of service and academic analysis;
- social interactions (social networks) can be traced and visualized in so-called sociograms (Wessa 2009c). In theory, this allows us to build new applications that take into account the social relationships between users in order to offer customised services;
- peer review messages are linked to the computational repository which allows us to support, monitor, and measure peer review activity based on objective measurements

The bottom line is that there is a complete separation between presentation, storage, and computing. This is a fundamental requirement for ubiquitous computing and learning.

3.3. Reported Student Experiences

The survey data were obtained from the business students in Belgium because they were required to submit the surveys before the end of the semester. The questions were taken from three well-known questionnaires (ATTLES, COLLES, and CSUQ). The first survey (called ATTLES) was available from the start of course and aims to measure student's attitudes towards thinking and learning (Galotti et al., 1999). The first ten questions relate to "connected" (empathic) ways of learning whereas the ten last questions are associated with "separate" (critical, detached) ways of knowing.

Students' perception of their online learning experience during the course was measured with the Constructivist On-Line Learning Environment Survey (COLLES) which focuses on a spectrum of important aspects of the learning experiences: relevance of the subject, reflective/critical thinking, interactivity, educator support, peer support, and interpretation of peer/educator messages. For every aspect there were eight questions, four of which are related to the actually perceived experience. The remaining four questions have identical phrases but are related to the degree of what students prefer.

The third survey is based on IBM's Computer System Usability Survey (called CSUQ) as originally developed by Lewis (1993) with additional questions as proposed by Poelmans et al. (2008) that are specifically focused on the relationship between software usability and statistics learning.

The analysis of the survey responses was performed in such a way that anyone is able to interpret the results. Each question was based on a 5-point Likert scale (5 is excellent, 3 is neutral, and 1 is poor). By subtracting a fixed constant (= 3) we obtained scores that are contained in the interval [−2, 2] where the neutral score is zero valued. This score $S_{i,j}$ represents the transformed reply (for all questions i = 1, ..., Q and for all students j = 1, ..., N) for which the following definitions can be formulated:

- $D_{i,j}^{+} = 1$ if $S_{i,j} > 0$, $D_{i,j}^{+} = 0$ and $S_{i,j} \leq 0$
- $D_{i,j}^{-} = 1$ if $S_{i,j} < 0$, $D_{i,j}^{-} = 0$ and $S_{i,j} \geq 0$
- P_i^s is the sum of all positive scores: $P_i^s = \Sigma_{j=1}^{N} D_{i,j}^{+} S_{i,j}$ for i = 1, ..., Q
- N_i^s is the sum of all absolute values of negative scores: $N_i^s = \Sigma_{j=1}^{N} D_{i,j}^{-} |S_{i,j}|$ for i = 1, ..., Q
- P_i^c is the number of positive scores: $P_i^c = \Sigma_{j=1}^{N} D_{i,j}^{+}$ for i = 1, ..., Q
- N_i^c is the number of negative scores: $N_i^c = \Sigma_{j=1}^{N} D_{i,j}^{-}$ for i = 1, ..., Q

It is now possible to define three summary scores for each question:

1. the arithmetic mean: $(1/N)\ \Sigma_{j=1}^{N} S_{i,j}$ for i = 1, ..., Q
2. the difference between positive and (absolute) negative scores, divided by the absolute sum of all scores: $(P_i^s - N_i^s) / (P_i^s + N_i^s)$ for i = 1, ..., Q
3. the difference between the number of positive and negative scores, divided by the number of non-zero scores: $(P_i^c - N_i^c) / (P_i^c + N_i^c)$ for i = 1, ..., Q

The first two measures can only be used if a quasi-interval scale is assumed. The third measure does not make the assumption of a quasi-interval scale because the scores are substituted by frequencies (counts). The drawback of the third measure is that is does not differentiate between extreme answers (±2) and moderate answers (±1). In other words, the third measure has the advantages that are associated with ordinal (rank-based) measures but at a cost of loss of information. The first measure is contained in the interval [−2, 2] and last two measures lie in the interval [−1, 1].

In each survey, and for all questions, a high summary score is associated with a "favourable" situation. A negative score indicates a weak point that may be considered for improvement. Instead of displaying the computed statistics in this book chapter, we include the links to the remotely stored computational objects. Any reader who wishes to examine, reproduce, or verify our findings is encouraged to visit the repository and (if so desired) to re-compute or re-use the analysis on our remote network of computational servers.

The ATTLES survey scores are available online (Wessa, Holliday and Reddy 2010a; http://www.freestatistics.org/blog/date/2010/Apr/05/t1270471667mb02yn344mzcneh.htm/). The conclusion from the analysis is positive for most aspects of the ATTLES survey. Negative summary scores are found for the following two questions:

- Q2: I like playing devil's advocate - arguing the opposite of what someone is saying.
- Q9: I often find myself arguing with the authors of books that I read, trying to logically figure out why they're wrong.

The negative summary scores for these questions may indicate that our students dislike challenging others and don't have a critical attitude towards claims made by (book) authors. Hence, the introduction of new learning technologies that allow students to reproduce or challenge computations from peers is expected to be difficult and lead to negative learning experiences. As a consequence it is rather unlikely that students enjoy assignments that are related to RC and Peer Assessment.

Fortunately, the analysis about COLLES (Wessa, Holliday and Reddy 2010b; http://www.freestatistics.org/blog/date/2010/Apr/05/t1270472835g8j33hyexeibe23.htm/) shows an overwhelming amount of evidence that students perceive their learning experience (at the end of the semester) as very positive. This comes as a surprise because of the fact that the workload is perceived to be high, and the observation that RC goes against student's attitudes towards learning and thinking (as measured in the initial ATTLES survey). All COLLES questions have a positive summary score – some are even close to the maximum value.

Another interesting result can be observed in the CSUQ analysis (Wessa, Holliday and Reddy 2010c; http://www.freestatistics.org/blog/date/2010/Apr/05/t1270473420yuhs2t83pn49hzj.htm/) which clearly shows that the web-based software was highly rated by students. The only exception is related to question 10: "The website gives error messages that clearly tell me how to fix problems." This negative summary score is due to the fact that error messages (produced by the R language) are of a "technical" or "purely statistical" nature. For this reason, students were instructed to archive computational results with error messages and send the link to instructor (by e-mail) or post it in the VLE forum. The Compendium Platform allowed the instructor to quickly reproduce the errors, detect problems, and solve any computational or software-related issue and report back to the student. This method of error handling is not only very efficient - it also provides a lot of valuable insight into the nature of problems that are commonly encountered.

Some results in the CSUQ analysis are of particular interest:

- Q21: Learning Statistics with this website is more effective than with a traditional handbook
- Q22: Overall, the website was helpful in learning statistics
- Q23: I intend to use this website when I need to apply statistics in the future
- Q28: To learn statistics, this website is better than the statistical courses I have had so far

- Q29: Next year, I will probably use The website/software again if I have to do statistical assignments
- Q35: The fact that I can reproduce computations of others is beneficial for my Learning
- Q40: my Learning improves because The website allows me to reproduce and reuse The computations of others
- Q41: I have a deeper understanding (of statistics) because the website's technology allows me to evaluate the work of other students

The scores for these questions are quite large and imply that students believe that the web-based CP helps them to learn statistics. The overall appreciation is very strong and may compensate the fact that the learning process involves a lot of work, and that RC does not agree with their initial attitudes towards thinking and learning.

The Survey results were confirmed by several focus group discussions in Belgium which were conducted by independent and experienced people. Most students reported that they initially felt rather anxious due to the uncertainty that is associated when the new technology and pedagogical approach is introduced. However, almost all students became used to the new situation quickly (after about two weeks) and most of them began to appreciate the benefits of RC for the purpose of learning in a socially constructive setting.

Based on the focus groups and individual interviews, the following “most important problems” were identified:

1. A workload which is too high.
2. Uncertainty about the correctness of solutions and/or Peer Reviews.
3. No grades are received for the submitted workshops.
4. Everyone else is using traditional statistical software.

The first problem turns out to be a false one. There is a learning curve involved in acquiring the skills to do Peer Reviews of high quality. If there is proper guidance and support from the educator, this problem should disappear after a few weeks. The second problem can be solved during the lectures by introducing real examples of good & bad types of analysis which can be found in the submitted workshops. In addition, the educator can provide students with an illustrated solution about the previous week’s workshop assignment. The third problem is actually not a real problem but rather a matter of perception. Student should understand that the effort that is invested into their workshops ultimately leads to better understanding and higher quality of Peer Reviews (which in turn are subject to grading by the instructor). In addition, it is possible to grade the quality of the Peer Reviews in an efficient and “fair” manner as is demonstrated by Wessa and De Rycker (2010). The fourth problem is a rather poor argument in favour of traditional software. Almost all students agree that RC provides better support for learning while traditional software involves many problems that are counterproductive due to a variety of issues (such as: installing software, importing/converting data, emphasis on the user interface instead of the subject under study, old or outdated methods, etc...). Most commercial software products for statistical computing are easy to learn (most of them come with a rich Graphical User Interface) once the user truly understands the statistical method to be applied and its underlying assumptions. The RC solution allows the educator to create tailor-made statistical software with the freedom to offer the features and choices in its implementation that are of importance, and to take away the pain of data import and manipulation. The bottom line is that students can always (and easily) learn to use commercial software after they have acquired a true understanding of the underlying statistical concepts.

It is interesting to point out that many students reported that they particularly liked the ubiquitous nature of the learning environment. The fact that they were able to use RC in various contexts (whenever and wherever they liked) was one of the main aspects that contributed to their satisfaction.

3.4. Objective Measurements

Another way to assess the impact of RC on learning is to look at objectively measured statistics about the actual learning activities. Based on web server statistics from the last three months of 2009, we were able to observe that our students:

- viewed +143000 pages (during a total of 26107 visits) about blogged computations on the freestatistics.org website. Most of these pageviews can be associated with the weekly peer review activities. The average time on site per visit was 17 minutes and 33 seconds.
- generated +131000 statistical computations (during a total of 14064 sessions) on the wessa.net website. The average time on site per visit was 15 minutes and 16 seconds.

Based on the measurements of the Peer Review software we found that the students spent an average of 12-15 minutes (depending on the course) between successive reviews. The number of feedback messages per review was 118-139 (depending on the course) and the average length of each message was 259-336 characters (depending on the course and the language). Each student reviewed (on average) 37-43 Compendia during an average of 11-13 unique days. The number of unique (full) hours in which reviews took place was 20-21.

Based on these statistics and making some (mild) assumptions, we estimate that students had an average workload of about 6-8 hours during each week (regular classes not included). An important aspect about this is the fact that the workload occurs in a regular (weekly) pattern throughout the course period. Figure 3 shows the daily number of visits of the freestatistics.org website by our students and clearly illustrates the weekly pattern.

Figure 3. Daily visits of freestatistics.org by our students

The ubiquitous nature of the learning environment is clearly illustrated by Figure 3 because there are still many visits during the off-peak days (even during the one-week break in November). If the web server statistics are investigated in more details, we find that the off-peak period traffic is associated with IP addresses that are different from those which occur during the peaks. Also the types of computers (and devices) that are used are different. For instance, in the business studies course we see a lot more netbook and mobile devices during the weekdays and lecture hours. Another point of interest is the fact that many computations were viewed or reproduced through links that were transmitted by e-mail, forums, and various types of instant messengers. This confirms the information that was reported during focus group discussions.

4. Future Trends

The future of RC promises to be challenging and exciting. The foundations of RC provide us with plenty of opportunities to pursue new roads of research and to build new applications. Within the educational context there are three new projects that might be of particular interest.

The first project extends the CP to the field of collaborative, academic writing. Even though it may seem strange at first sight, statistical computations have a lot in common with snippets of text that are written by students. Most of the finished documents or papers that we observe are the product of multiple attempts to write and revise text snippets over a longer period of time. Towards the completion of the text, the snippets are glued together in order to obtain a text which makes sense to the reader. Similarly, in a statistical paper we typically have several computations that have some kind of relationship to each other – or relate to the same data set. Moreover, a text snippet can be written/revised by an individual or through collaboration – it is also possible to review texts or text snippets. In this sense, it is possible to extend the database of the online repository of the CP to allow us to store, track, and reproduce entire texts that are produced by students. This is not limited to the snapshots (or versions) of texts (such as in a traditional wiki) but comprises the dynamic interactions between users and their respective contributions.

The second project focuses on scientific publishing and dissertations at the Master or Ph.D. level. The underlying idea is that students should be allowed (or even encouraged) to publish their research in a student journal in which the primary requirement is not the novelty or relevance of the contribution but instead the scientific rigor and reproducibility. Just imagine how such an international journal would improve the productivity for supervisors and students alike. Instead of going through the lengthy effort of training each generation of students to become proficient in writing a dissertation in a particular field of interest we could provide them with a head start if we could guarantee that the published dissertations from the previous generations are fully reproducible.

A third application emphasizes the embedding of RC into new or existing games with an educational purpose. One particular example is an online Stock Market Game in which the students participate and play the role of a professional trader. The stock prices are influenced by the participants (the so-called law of demand and supply) and a computer player which represents the rest of the market and acts according to research-based Artificial Intelligence rules. By introducing RC into the game we empower students to learn about statistical analysis techniques and financial models in an exciting, real-time environment which can be partially controlled through the constraining parameters of the computer player. In addition, every important decision of the participant (which is based on analysis) can be monitored in real time and reproduced for the purpose of post game analysis. By seamlessly integrating RC into the game we provide valuable feedback to the learner and the educator which was never available before. Educational games have the potential of becoming much more interesting (as an educational tool) because of reproducibility of analysis-based decisions.

Best of all is the fact that RC and the CP are made available free of charge for the purpose of (non-commercial) research and education. Anyone can build new applications based on RC or contact the authors to discuss opportunities for scientific or educational collaboration.

5. Conclusion

We have described and discussed RC (and associated CP) as a socially constructivist learning technology with characteristics of ubiquitous computing. The technology builds a bridge between two seemingly unrelated worlds of statistical computing and educational science. At the same time it solves some important problems in a manner which is flexible (tailor-made solutions) and easily accessible at the same time (no technical knowledge is required by the students).

Without the need to employ sophisticated analysis, we provided strong evidence that student perceptions and experiences are very positive towards this new type of technology and educational approach. The most remarkable observation is that the approach was well-received notwithstanding the fact that students complained about a heavy workload in the beginning of the course. The (few) criticisms that were formulated have no solid ground and may be solved through better and more extensive communication.

References

Buckheit, J. and Donoho, D. L. (1995). Wavelets and Statistics, In A. Antoniadis (Ed.) *Wavelab and reproducible research*. Springer-Verlag.

de Leeuw, J. (2001). *Reproducible research: the bottom line*. (Department of Statistics Papers, 2001031101). Department of Statistics, UCLA.

Donoho, D. L. and Huo, X. (2004). Beamlab and reproducible research. *International Journal of Wavelets, Multiresolution and Information Processing.*, 2 (4): 391-414, December

Eggen, P. & Kauchak, D. (2001). *Educational Psychology: Windows on Classrooms*. Upper Saddle River, NJ: Prentice Hall, 5th ed. Edition.

Galotti, K. M., Clinchy, B. M., Ainsworth, K., Lavin, B., and Mansfield, A. F. (1999). A new way of assessing ways of knowing: the attitudes towards thinking and learning survey (ATTLS). *Sex roles*, 44(9/10), p. 745–766.

Gentleman, R. (2005). Applying reproducible research in scientific discovery. BioSilico, URL http://gentleman.fhcrc.org/Fld-talks/RGRepRes.pdf

Green, P. J. (2003). Diversities of gifts, but the same spirit. *The Statistician*, 52(4), p. 423–438.

Koenker, R. and Zeileis, A. (2007). *Reproducible econometric research (a critical review of the state of the art)*. (Research Report Series, number 60). Department of Statistics and Mathematics Wirtschaftsuniversität Wien.

Leisch, F. (2003). Sweave and beyond: Computations on text documents. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria.

Lewis, J. R. (1993). *Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use*. (Technical Report 54.786). IBM Corporation.

Mvududu, N. . (2003). A cross-cultural study of the connection between students' attitudes toward statistics and the use of constructivist strategies in the course, *Journal of Statistics Education*, 11(3).

Peng, R. D., Dominici, F., and Zeger, S. L. (2006). Reproducible epidemiologic research, *American Journal of Epidemiology*, 163(9), 783-789

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Schwab, M., Karrenbach, N., and Claerbout, J. (2000). Making scientific computations reproducible. *Computing in Science & Engineering*, 2(6), 61-67

Smith, E. (1999). Social constructivism, individual constructivism and the role of computers in mathematics education, *Journal of mathematical behaviour*, 17(4), 411-425

Von Glasersfeld, E. (1987). Learning as a constructive activity. In *Problems of Representation in the Teaching and Learning of Mathematics*, 3–17. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wessa, P. (2008). Learning statistics based on the compendium and reproducible computing. In *Proceedings of the International Conference on Education and Information Technology*, Berkeley, San Francisco, USA.

Wessa, P. (2009a). A framework for statistical software development, maintenance, and publishing within an openaccess business model, *Computational Statistics*., 24(2), 183-193

Wessa, P. (2009b). How reproducible research leads to non-rote learning within a socially constructivist statistics education. *Electronic Journal of E-Learning*, 7(2), 173-182

Wessa, P. (2009c). Exploring Social Networks in Reproducible Computing and Collaborative Assignments, In *Proceedings of the International Conference on E-Learning 2009*, University of Toronto, Canada

Wessa, P. and De Rycker A. (2010), Reviewing Peer Reviews – a Rule Based Approach, In *Proceedings of the International Conference on E-Learning 2010*, Malaysia, forthcoming

Wessa, P., Holliday, I., Reddy, P. (2010a). Statistical Computations at FreeStatistics.org, Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2010/Apr/05/t1270471667mb02yn344mzcneh.htm/ , Retrieved Mon, 05 Apr 2010 13:37:41 +0000

Wessa, P., Holliday, I., Reddy, P. (2010b). Statistical Computations at FreeStatistics.org, Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2010/Apr/05/t1270472835g8j33hyexeibe23.htm/ , Retrieved Mon, 05 Apr 2010 13:38:05 +0000

Wessa, P., Holliday, I., Reddy, P. (2010c). Statistical Computations at FreeStatistics.org, Office for Research Development and Education, URL

http://www.freestatistics.org/blog/date/2010/Apr/05/t1270473420yuhs2t83pn49hzj.htm/ ,
Retrieved Mon, 05 Apr 2010 13:38:35 +0000

Definitions

Reproducible Computing: *the use of computers and computer software in a manner that allows anyone else to replicate the results that are obtained at a later moment in time. In the traditional literature, researchers have almost exclusively used the term "Reproducible Research" instead of Reproducible Computing – this however may be confusing because the expression exclusively refers to the reproducibility of the computational aspects of research.*

R Framework: *a set of software components which allow the user to compile and maintain web-based software modules (so-called R modules) based on R code and a simple description of the user interface. The R Framework ensures that there is a complete separation between user interface, storage, computing, and session management.*

Compendium (traditional definition): *collection of Text, Data, and Software that allows the reader to reproduce the research that is presented in the document*

Compendium (new definition): *a document with (open-access) references to (remotely) archived computations (including Data, Meta-data, and Software) that allows the reader to reproduce, and re-use the underlying analysis*

Compendium Platform: *a system which facilitates the creation, maintenance, and permanent storage of statistical computation objects that empower authors to publish "Compendia" of reproducible and re-usable research through a series of web services.*