Measurement and Control of Statistics Learning Processes based on Constructivist Feedback and Reproducible Computing

Patrick Wessa

K.U.Leuven Association, Lessius Dept. of Business Studies, Belgium
E-mail: patrick@wessa.net

Abstract

This article introduces a new approach to statistics education that allows us to accurately measure and control key aspects of the computations and communication processes that are involved in non-rote learning within the pedagogical paradigm of Constructivism. The solution that is presented relies on a newly developed technology (hosted at http://www.freestatistics.org) and computing framework (hosted at http://www.wessa.net) that supports reproducibility and reusability of statistical research results that are presented in a so-called Compendium. Reproducible computing leads to responsible learning behaviour, and a stream of high-quality communications that emerges when students are engaged in peer review activities. More importantly, the proposed solution provides a series of objective measurements of actual learning processes that are otherwise unobservable. A comparison between actual and reported data, demonstrates that reported learning process measurements are highly misleading in unexpected ways. However, reproducible computing and objective measurements of actual learning behaviour, reveal important guidelines that allow us to improve the effectiveness of learning and the e-learning system.

Keywords: Reproducible Computing, Virtual Learning Environment, Communication, Statistics Education, Learning Processes

1 Introduction and Literature

In education-related research it is common practice to investigate learning processes through measurements that are based on questionnaires. Reported measures often reveal interesting information about a wide variety of aspects of computing-assisted learning such as: computer attitudes (Meelissen and Drent, 2008); computer emotions and knowledge (Kay 2008); learner experiences and satisfaction (Sun et al. 2008); etc. The importance of such measurements has been highlighted by many authors from various perspectives (Chen, 2008; Hilton et al., 2004; Galotti et al., 1999) – especially from the perspective of the constructivist pedagogical paradigm (Von Glasersfeld, 1987; Smith, 1999; Eggen et al., 2001; Mvududu, 2003).

These reported measures, while intrinsically interesting, may not always provide us with the information we need to assess and improve systems that support e-learning. Moreover, the implementation of new learning technologies and data analysis tools opens up a wide array of measurement opportunities which leads to new areas of research. An excellent example is the use of data mining tools in the open source e-learning environment called Moodle (Romero et al., 2008).

Even though it seems to be very difficult to measure and empirically prove (O'Dwyer et al., 2008), there is no doubt in my mind that the introduction of computers in homes and classrooms has led to an improvement in overall learning productivity, educational communication mechanisms, social constructivism, and collaboration. However, the use of computers and software in statistics education may – unwillingly – result in several types of adverse effects because the complex processes that are required to learn and (truly) understand statistical concepts are often mystified by technicalities and a variety of practical problems that have nothing to do with mathematics or statistics. It is within this context that I argue that a system for Quality Control should be embedded into the e-learning environment which is not limited to the Virtual Learning Environment but extends to the statistical software, databases, and learning repositories.

There is an important, additional benefit for implementing such a monitoring and control system – it is directly related to the problem of irreproducible research which has received a great deal of attention within the statistical computing community (de Leeuw, 2001; Peng et al., 2006; Schwab et al., 2000; Green, 2003; Gentleman, 2005; Koenker and Zeileis, 2007; Donoho and Huo, 2004). The most prominent citation about the problem of irreproducible research is called 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).

Several solutions have been proposed (Buckheit and Donoho, 1995; Donoho and Huo, 2004; Leisch, 2003) but have not been adopted in statistics education because they require students to understand the technicalities of scientific word processing (LaTex) or statistical programming (R code). Based on a newly developed e-learning environment I propose a solution that is feasible for educational purposes and allows us to monitor, research, and control the learning processes based on the dynamics of between-student communication and collaboration.

2 Reproducible Computing

2.1 R Framework and Compendium Platform

The R Framework allows educators and scientists to develop new, tailor-made statistical software (based on the R language) within the context of an open-access business model that allows us to create, disseminate, and maintain software modules efficiently and with a very low cost in terms of computing resources and maintenance efforts (Wessa, 2008a). The so-called R modules empower students to perform statistical analysis through a web-based interface that does not require them to download or install anything on the client machine. This permits students to focus primarily on the interpretation of the analysis – however, the R Framework also allows advanced students and scientists to inspect and change the R code that was coded by the original author. This results in the creation of so-called “derived” R modules that may be better suited for particular purposes.

If a derived R module contains generic improvements or if a computation needs to be communicated to other students/scientists then it is necessary to have a simple, transparent mechanism that allows one to permanently store the computation in a repository of computational objects that can be easily retrieved, recomputed, and reused. Such a repository was recently created within the OOF 2007/13 project of the K.U.Leuven Association and is called the Compendium Platform. The main reason for creating the R Framework and the Compendium Platform, is that it allows anyone to create and use Compendia of reproducible research. A Compendium is defined as (Wessa 2008b): *a research document where each computation is referenced by a unique URL that points to an object that contains all the information that is necessary to recompute it*. Such documents can be easily created (even by students) and permit any reader to (exactly) recompute the statistical results that are presented therein. A few simple clicks are sufficient to have the R Framework reproduce the results and to reuse them in derived work (Wessa 2008b). The practical implications of this technology will become obvious in section 3 because the three figures that are presented can be recomputed and reused through the Compendium Platform.

2.2 Communication, Feedback, and Learning

The concept of Reproducible Computing was implemented in several undergraduate statistics courses in order to thoroughly test the new system and to measure key aspects of the educational activities and experiences. Two different student populations were investigated in detail: a group of (academic) bachelor students, and a group of so-called “switching” students. The second population is of particular interest because it consists of students who obtained a (professional) bachelor degree and decided to make the “switch” to an academic master which requires them to complete a preparatory year.

On the one hand, switching students are highly motivated and more mature than the bachelor students. A priori, one would expect them to prefer practical activities (such as communication and computing) above theory and critical reflection. On the other hand, one might expect the bachelor students to have a more critical (scientific) attitude and better mathematical background than the switching students.

Students from both populations took a similar statistics course which covered topics from introductory statistics, regression analysis, and introductory time series analysis. The main learning activities in both statistics courses were based on a weekly series of workshops where each student was required to investigate practical, empirical problems. At the end of each week, students submitted their papers electronically. During the lecture I proposed a series of solutions and illustrated commonly made mistakes. After the lecture, students had to work on the next assignment and complete a series of peer reviews (assessments) about the work that was submitted the week before. The assessment grades did not count towards the final score – however, each submitted peer review was accompanied by verbal feedback messages. I graded a (quasi random) sample of these messages in order to provide students with a strong incentive to take the review process seriously. There is strong empirical evidence that this approach had beneficial effects on non-rote learning of statistical concepts (Wessa 2008c).

3 Measurement and Control

In Wessa (2008b) it is illustrated how the Compendium Platform's repository supports quality control of the statistical software and accompanying documentation for students. On the one hand, reproducible computing allows students to accurately communicate computational problems and questions without the need to understand the underlying technicalities. On the other hand, it allows the educator (and creator of the computational software) to analyse the reported problem (based on the detailed, raw output of the R engine that executed the request) and to transparently communicate the solutions to the students. In addition, the measurement of learning activities and experiences is a conditio sine qua non for controlling the overall quality of learning systems. This will be illustrated based on the data that was collected for both student groups. At the same time, the importance of objective (as opposed to reported) measurements is illustrated based on a simple, comparative diagnostic tool.

The reported measurements were obtained through questionnaires on a 5-point Likert scale and should consequently be treated as ordinal data. The questions were based on well-known psychological surveys (Galotti et al. 1999; COLLES 2004) and an extended version of the IBM computer system usability survey (Lewis 1993). Useful data was obtained from a total of 111 bachelor students and 129 switching students – the response ratio was very high (between 82.9% and 92% depending on the questionnaire). All observations of actual learning activities were measured on a ratio scale (the number of archived computations and the number of submitted feedback messages). A total number of 34438 meaningful, verbal feedback communications and 6587 archived computations were registered. In order to compare the actual and reported data, all measurements were converted to ordinal rank orders. In addition, the Pearson's rho correlations and Kendall's tau rank correlations (Arndt et al. 1999; Arnd, Magnotta 2001; Hollander et al. 1973) that represent the degree of linear association between the properties under investigation were computed (these can be consulted in the archived computations about the Figures). In electronic versions of this paper, one can simply (ctrl-)click the pictures to view the archived computation in the repository. Readers of the printed version of this document, are referred to the bibliography where three references can be found (including the URLs) about the statistical computations that have been stored at www.freestatistics.org.

Figure 1 displays the bivariate kernel density (Lucy et al. 2002) between the rank order of the number of feedback messages that have been submitted in peer reviews about the workshops (x-axis) and the rank order of the number of (reproducible) computations that have been archived in the repository (y-axis). The rank orders have been computed within the Bachelor population for the top panels, and within the

Switching population for the bottom panels. This implies that the ranks that are attributed to female and male students are expressed on the same axes and can be compared. Figure 1 clearly demonstrates that female bachelor students are much more involved in feedback and computing than their male colleagues. At the same time, female switching students are more computing-oriented whereas the male switching students seem to have a slight preference for feedback communication. This information has important repercussions for controlling the quality of the learning environment and it provides clear guidelines towards actions that should be taken (by me) to improve participatory incentives towards male bachelor students in future courses. Would I have been able to gain this insight based on reported measurements alone? The answer is clearly negative (as is illustrated in Figures 2 and 3).

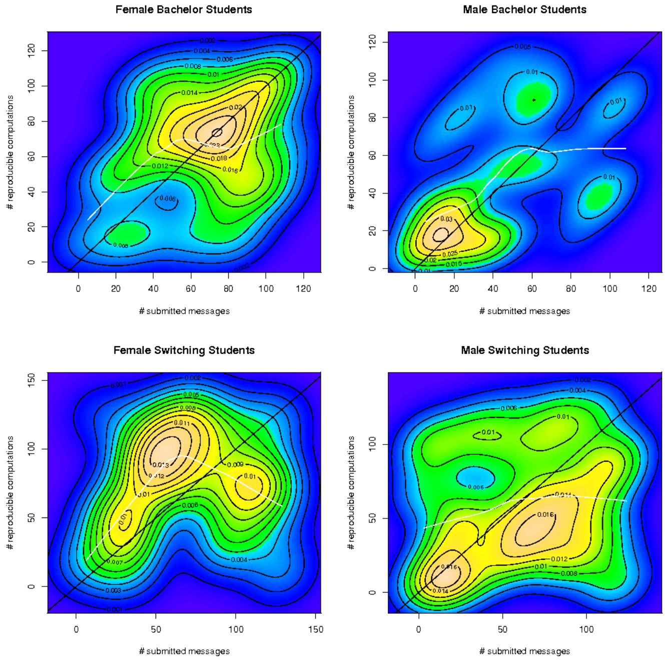

Figure 1. Submitted Feedback versus Reproducible Computations

It is quite obvious that male bachelor students highly over-estimate their performance in terms of feedback submissions (see Figure 2) because the rank orders of reported measures (x-axis) are higher than the ranks of actual feedback submissions (y-axis). Female bachelor students however, underestimate their involvement (relative to their male colleagues) because they are concentrated above the diagonal line. In the male switching student population several clusters of high density can be detected which leads us to conclude that we cannot treat them as one homogeneous group.

In Figure 3 the comparison between reported computing measures (x-axis) and actual computing (y-axis) leads to similar conclusions. Male bachelor students highly exaggerate their efforts, whereas female bachelor and switching students underestimate themselves. The group of male switching students is heterogeneous.

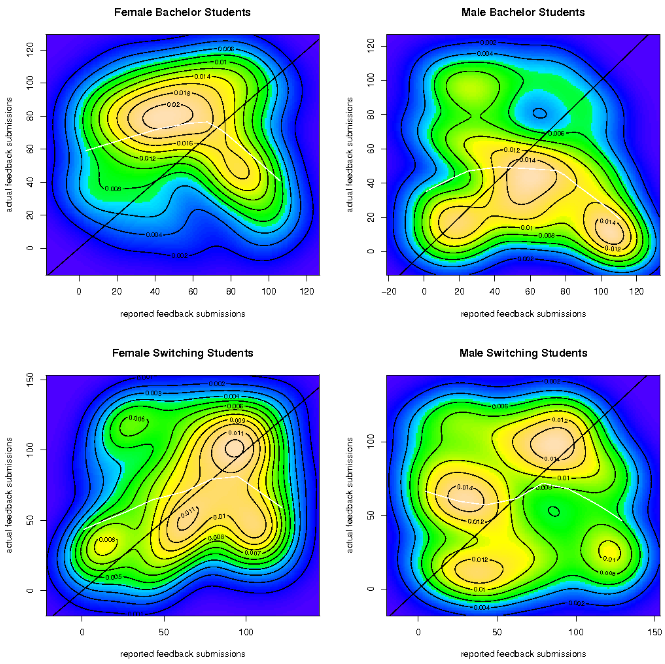

Figure 2. Reported versus Actual Submitted Feedback

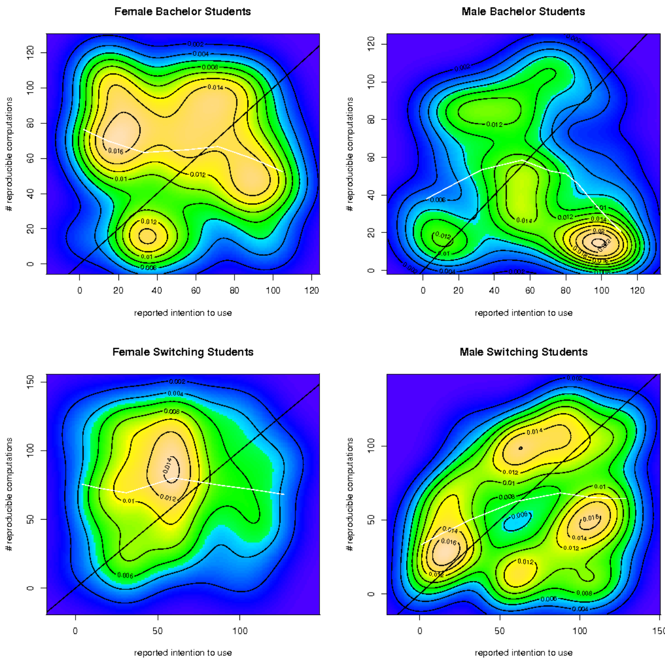

Figure 3. Reported versus Actual Reproducible Computing

Overall, the testimony of students is extremely misleading and poorly correlated with actual observations. If we would have recomputed Figure 1 with reported measures then the conclusions would have been the opposite of what is true. The reader can try out this experiment by simply reproducing the computation of Figure 1 with reported measures on both axes.

The good news is that we now have a tool available to assess actual and reported learning activities for any student population that makes use of the new compendium technology. Ultimately, this allows us to take control and improve the e-learning environment, the statistical software, the course materials, and overall learning experiences of all students.

4 References

Arndt, S., Turvey, C., Andreasen, N. (1999), Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation, *Journal of Psychiatric Research*, 33, 97-104

Arndt, S., Magnotta, V. (2001), Generating random series with known values of Kendall's tau, *Computer Methods and Programs in Biomedicine*, 65, 17-23

Attitudes to Thinking and Learning Survey, (n.d.), Retrieved December 22, 2004, from http://www.moodle.org/

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory study, *Journal of Experimental Education*, 57, 247–261.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983), Graphical Methods for Data Analysis., Wadsworth & Brooks/Cole.

Chen, Z. (2008), Learning about Learners: System Learning in Virtual Learning Environment, *International Journal of Computers, Communications & Control*, Vol. III (2008), No. 1, pp. 33-40

COLLES (2004), Constructivist On-Line Learning Environment Survey, Retrieved December 22, 2004, from http://www.moodle.org/

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

Galotti, K M, Clinchy B M, Ainsworth K, Lavin, B & Mansfield A F (1999), A new way of assessing ways of knowing: the attitudes towards thinking and learning survey (ATTLS), Sex roles 40(9/10) p745-766

Hilton, S., Schau, C., Olsen, J. (2004), Survey of Attitudes Toward Statistics: Factor Structure Invariance by Gender and by Administration Time, *Structural Equation Modeling*, Volume 11, Number 1

Hollander, Myles, and Douglas A. Wolfe (1973), Nonparametric statistical inference., New York: John Wiley & Sons. Pages 185-194 (Kendall and Spearman tests).

Kay R. H. (2008), Exploring the relationship between emotions and the acquisition of computer knowledge, *Computers & Education*, 50, 1269–1283

Lewis J. R. (1993), IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use, IBM Corporation, Technical Report 54.786

Lucy D., Aykroyd, R.G. & Pollard, A.M.(2002) Non-parametric calibration for age estimation. Applied Statistics 51(2): 183-196

Meelissen M. R.M, Drent M. (2008), Gender differences in computer attitudes: Does the school matter?, *Computers in Human Behavior*, 24, 969–985

Miller, Jacqueline B., Examining the interplay between constructivism and different learning styles, Retrieved October 20, 2005 from www.stat.auckland.ac.nz/~iase/publications/1/8a4_mill.pdf

Mvududu, Nyaradzo (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*, Volume 11, Number 3

O'Dwyer L. M., Russell M., Bebell D., Seeley K. (2008), Examining the Relationship between Students' Mathematics Test Scores and Computer Use at Home and at School, *Journal of Technology, Learning, and Assessment*, 6 (5)

Romero C., Ventura S., GarcÌa E. (2008), Data mining in course management systems: Moodle case study and tutorial, *Computers & Education*, 51, 368–384

Smith, Erick (1999), Social Constructivism, Individual Constructivism and the Role of Computers in Mathematics Education, *Journal of mathematical behavior*, Volume 17, Number 4

Statistical Computations at FreeStatistics.org (2008a), Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2008/Jun/30/t1214840420q0fyankop4x9ebf.htm, Retrieved Mon, 30 Jun 2008 15:40:30 +0000

Statistical Computations at FreeStatistics.org (2008b), Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2008/Jun/30/t12148409608o0dnj2k4s04jil.htm, Retrieved Mon, 30 Jun 2008 15:49:23 +0000

Statistical Computations at FreeStatistics.org (2008c), Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2008/Jun/30/t1214841152sn6jlyhgseclgqm.htm, Retrieved Mon, 30 Jun 2008 15:52:41 +0000

Sun P., Tsai R. J., Finger G., Chen Y.,Yeh D. (2008), What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction, *Computers & Education*, 50, 1183–1202

Von Glasersfeld, E. (1987), "Learning as a Constructive Activity", in Problems of Representation in the Teaching and Learning of Mathematics, Hillsdale, NJ: Lawrence Erlbaum Associates, 3-17.

Wessa P. (2008a), A framework for statistical software development, maintenance, and publishing within an open-access business model, *Computational Statistics*, www.springerlink.com (DOI 10.1007/s00180-008-0107-y)

Wessa P. (2008b), Learning Statistics based on the Compendium and Reproducible Computing, submitted to be presented and published in *Proceedings of the International Conference on Education and Information Technology* (ICEIT'08), San Francisco, USA

Wessa P. (2008c), How Reproducible Research Leads to Non-Rote Learning Within a Socially Constructivist E-Learning Environment, submitted to be presented and published in *Proceedings of the 7th European Conference on e-Learning* (ECEL'08), Cyprus