Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. IV (2009), No. 2, pp. 185-197

Quality Control of Statistical Learning Environments and Prediction of Learning Outcomes through 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 www.freestatistics.org) and computing framework (hosted at 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, Learning Environment, Quality Control, Statistics Education, Psychometrics

1 Introduction

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 [22]; computer emotions and knowledge [17]; learner experiences and satisfaction [34]; etc... The importance of such measurements has been highlighted by many authors from various perspectives ([7], [15], [12]) especially from the perspective of the constructivist pedagogical paradigm ([35], [30], [11], [24]).

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 open up a wide array of measurement opportunities which lead to new areas of research. An excellent example is the use of data mining tools in the open source e-learning environment called Moodle [28].

Even though it seems to be very difficult to measure and empirically prove [25], there is no doubt in my mind that the introduction of computers in homes and classrooms has led to an improvement of 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 system, which is not limited to the Virtual Learning Environment but extends to the statistical software, databases, and learning repositories (Statistical Learning Environment).

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 ([9], [26], [29], [14], [13], [18], [10]). The most prominent citation about the problem of irreproducible research is called Claerbout's principle ([9]):

> *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...*

Several solutions have been proposed ([5], [10], [19]) 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 Statistical Learning Environment (SLE) 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

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 [36]. 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.

There are several important reasons why the R Framework helps in controlling the quality of the statistical learning processes that are supported by the computer:

- The R modules are web applications with an advanced session management which includes all aspects of the computations that are executed. In addition, the session manager uses attributes that identify the student and the course in which (s)he is enrolled. Therefore all computations that are performed within the context of a statistics course can be associated with an individual student - to implement this feature, the educator only needs to use certain HTML tags in the hyperlink that is inserted in the virtual learning environment.

- Every R module is uniquely described by an expandable set of meta data (incl. the actual statistical code) which can be stored and transmitted. This implies that every computation that is executed can be uniquely defined by the R module's meta data and additional information about the data and the parameters that have been specified by the user. As a consequence, every computation can be uniquely described and archived with meta data.

- The R Framework allows other servers (under certain conditions) to send meta data through an ordinary HTTP request which allows it to rebuild and execute the R module with the specified data and parameters in real time. Therefore it is possible to remotely store computational objects and send them back to the R Framework such that the original computation can be reproduced and reused.

- All the processes that are associated with the above items are automatically stored in a so-called process measurement database. This implies that all computer-assisted learning activities are objectively measured and stored for the purpose of analysis.

2.2 Compendium Platform

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 [37]: *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 [37]. 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.

2.3 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 [38].

3 Objective Measurements versus Reported Data

In a recent paper [37] it is illustrated how the Compendium Platform's repository supports "technical" 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 analyze 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. Moreover, the measurement of learning activities and experiences is a conditio sine qua non for controlling the "overall" quality of the SLE. This will be illustrated, based on the data that have been collected from 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 ([12], [8]) and the IBM computer system usability survey [20] which was adapted and extended [27]. 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 ([1], [2], [16]) 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 hyperlinks below Figures 1, 2, and 3 to view the archived computation in the repository. Readers of the printed version of this document, have to manually enter the respective URLs into their internet browser to view the statistical computations that have been stored (at www.freestatistics.org).

Figure 1 displays the bivariate kernel density [21] 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).

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

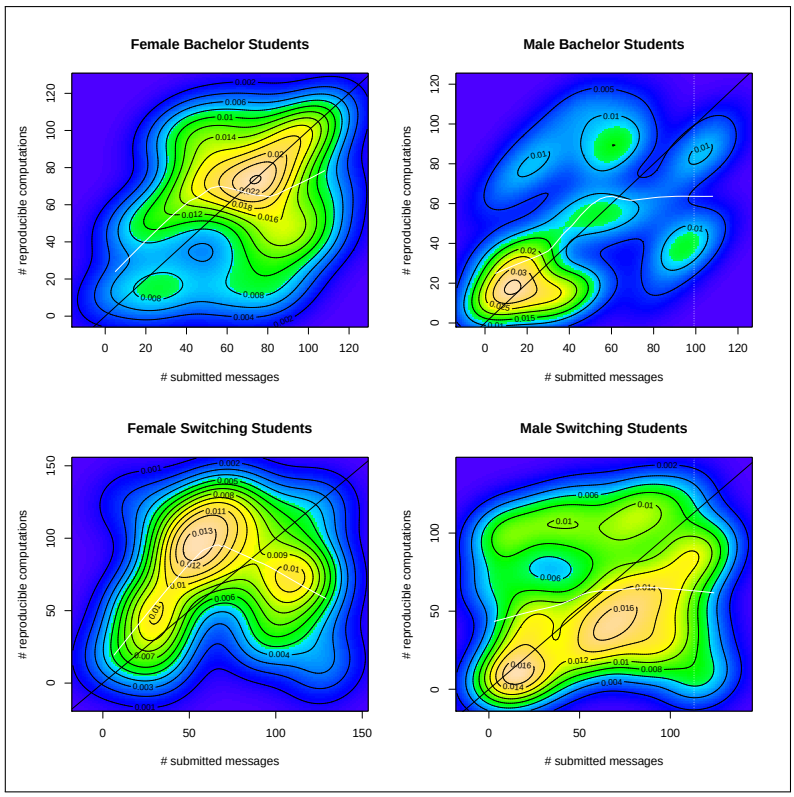

Figure 1: Submitted Feedback versus Reproducible Computations
www.freestatistics.org/blog/date/2008/Jun/30/t1214840420q0fyankop4x9ebf.htm

female bachelor and switching students underestimate themselves. The group of male switching students is heterogeneous.

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.

4 Quality Control

In order to be able to control (and improve) the quality of the SLE, it is necessary to estimate the impact of key-aspects of the learning processes that are associated with the SLE. The methodology that allows us to do this is based on a mathematical model which is described in [40] and relates the learning outcomes to objectively measured activities and reported experiences.

Typically, models that predict learning outcomes based on exogenous variables that are related to the learning (and computing) environment have an extremely low percentage of variance explained. In a recent and extensive study [25], six models were discussed that predicted the Statistics subtest scores of the Massachusetts Comprehensive Assessment System - the variance explained ranged between 4% and 7%.

It is obvious that any model that is used to control the quality of an SLE should perform much better. There are three important requirements to build high-quality models:

1. high-quality exogenous variables (preferably based on objective measurements) [39];

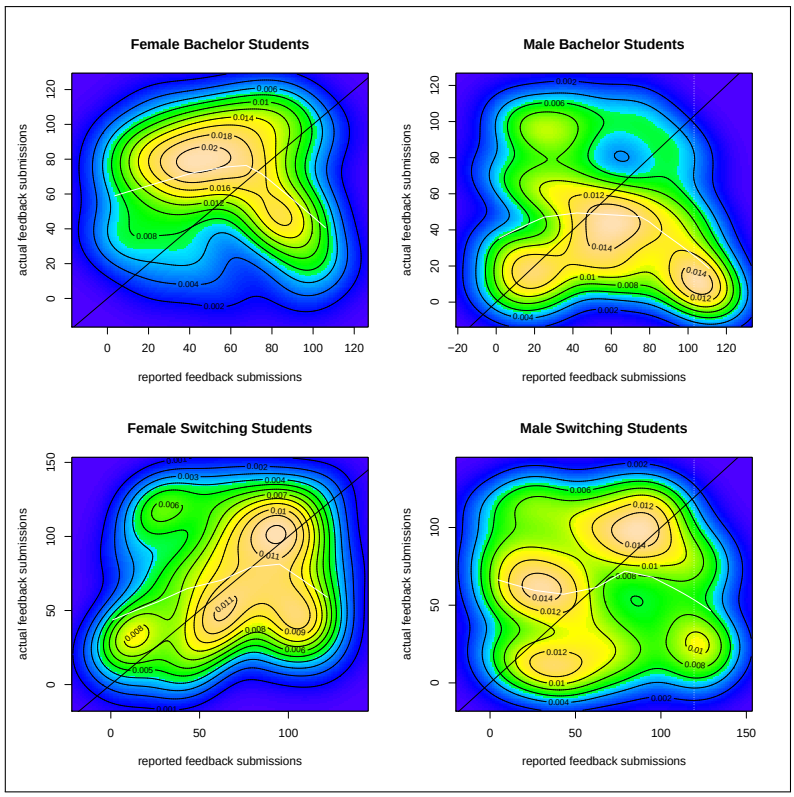

Figure 2: Reported versus Actually Submitted Feedback
http://www.freestatistics.org/blog/date/2008/Jun/30/t12148409608o0dnj2k4s04jil.htm

2. high-quality endogenous variable (c.q. test scores) based on optimal weights of the individual items (section 4.1, [40]);

3. homogeneous sample for which the model is computed.

The third condition refers to the fact that student populations may consist of different types of students with specific learning behaviors. In the formentioned statistics course there were 4 groups with distinct characteristics. This is clearly illustrated in section 3 and in Figures 1, 2, and 3.

Instead of computing separate models (for each of the sub populations) section 4.2 presents a comprehensive model with all combinations of interaction effects (male/female and Bachelor/Switching). This greatly improves the interpretation of the prediction model and allows us to perform differential quality control of the SLE.

4.1 Model

First, a classical regression approach is used to predict the learning outcomes (c.q. exam scores) as a linear function of $(K-1) \in \mathbb{N}_0$ exogenous variables of interest. Let $\vec{y}$ represent an $N \times 1$ vector for all $N \in \mathbb{N}$ students (with $N > K$), containing the weighted sum of G item scores (c.q. scores on individual exam questions): $\vec{y} \equiv \sum_{j=1}^{G} \omega_j \vec{y}_j$ with initial unit weights $\omega_j \equiv 1$. In addition, define an $N \times K$ matrix X that represents all exogenous variables (including a one-valued column which represents the constant), and a $K \times 1$ parameter vector $\vec{b}$ that represents the weights of the linear combination of all columns in X that is used to describe $\vec{y}$. The complete model is denoted M_1 and is defined by $\vec{y} = X\vec{b} + \vec{e}$ where $\vec{e} \leftarrow$ iid $N(\vec{0}, \sigma_e^2)$ represents the prediction error.

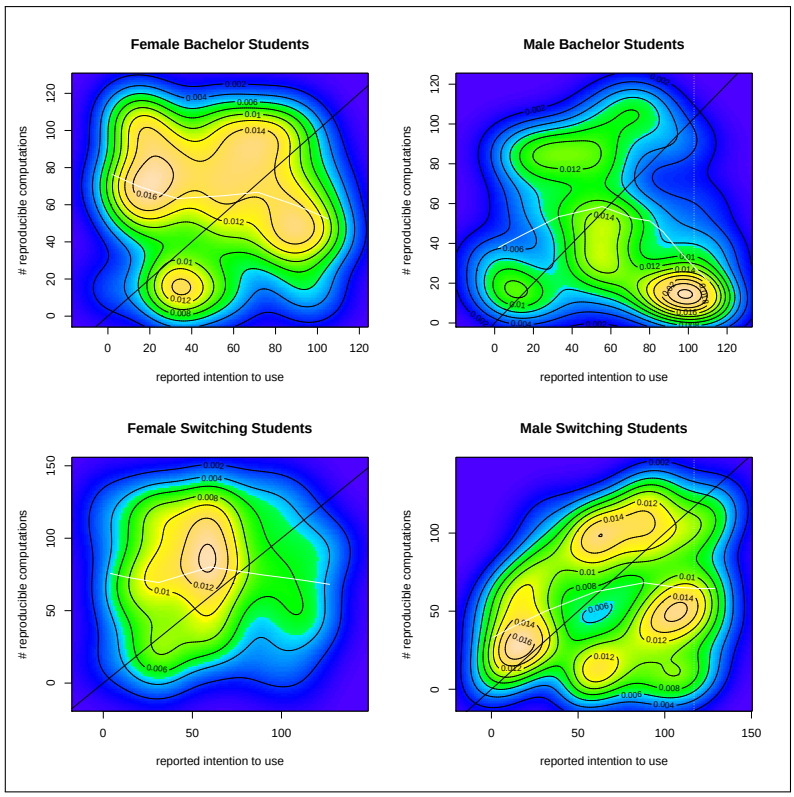

Figure 3: Reported versus Actual Reproducible Computing
http://www.freestatistics.org/blog/date/2008/Jun/30/t1214841152sn6jlyhgseclgqm.htm

In the second model M_2, the prediction of the first model is specified by a linear combination of the individual items (questions) that made up the total exam score. Let Y represent the $N \times G$ matrix that contains all G item scores, then it is possible to define the model $\hat{\vec{y}} = Y\vec{c} + \vec{a}$ where $\vec{a} \leftarrow \text{iid } N(\vec{0}, \sigma_a^2)$. Note that there is no constant term in this model.

The third model (M_3) simply combines M_1 and M_2 by relating $\hat{\hat{\vec{y}}}$ to X in the regression model $\hat{\hat{\vec{y}}} = X\vec{f} + \vec{u}$. The estimator for $\vec{f}$ can be shown to be $\hat{\vec{f}} = (X'X)^{-1}X'\hat{\hat{\vec{y}}} = (X'X)^{-1}X'Y(Y'Y)^{-1}Y'X(X'X)^{-1}X'\vec{y}$ ([40]). M_3 is likely to yield different results from M_1 unless the estimated parameters M_2 are (nearly) equal to the original weights $\hat{\vec{c}} = (\hat{c}_1, \hat{c}_2, \hat{c}_3, ..., \hat{c}_G)' \simeq (\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3, ..., \hat{\omega}_G)'$.

From a statistical point of view it is not possible to test the improvement that is induced by the objective exam score transformations. The reason for this is that the traditional F-test assumes that the endogenous variables in two models (M_1 and M_3) to be compared are identical. Therefore it is necessary to use an auxiliary model (M_3^*) which is based on M_3 and includes $\vec{y}$ as an explanatory variable. This extended model $\hat{\hat{\vec{y}}} = X\vec{f} + \vec{y}g + \vec{u}$ can be shown to be equivalent to $\left(Y(Y'Y)^{-1}Y'X(X'X)^{-1}X' - gI_N\right)\vec{y} = X\vec{f} + \vec{u}$ such that it can be concluded that M_3^* is equal to M_1 with a transformed endogenous variable. The interesting aspect about this auxiliary regression is the limiting case when $g \rightarrow 0$ and $Y(Y'Y)^{-1}Y'X(X'X)^{-1}X' \rightarrow I_N$ because it leads to M_1 with $\vec{f} = \vec{b}$ and $\vec{u} = \vec{e}$. This result is important because it is now easy to test if it is necessary to apply the transformation to the endogenous variable. The null hypothesis is simply $H_0 : g = 0$ versus $H_1 : g \neq 0$ which can be tested with the conventional t-test. In other words, if the null hypothesis is rejected then the transformation is necessary and the estimated parameters $\hat{\vec{c}}$ and $\hat{\vec{f}}$ interpretable. The usefulness of this modeling approach is illustrated in the next subsection.

4.2 Empirical Evidence

The data that was collected from the implemented SLE (as described in section 2.3) contained the following exogenous variables:

- Bcount: actual computations
- Gender: 0 = female / 1 = male
- Future: intention to use
- Pop: 0 = Bachelor / 1 = Switching
- nnzfg: actually submitted feedback messages in peer review
- Reflection: reported feedback messages in peer review

Table 1 presents the empirical results of two models (M_1 and M_3). The endogenous variable in M_1 is the sum of all exam questions with unit weights whereas M_3 is based on objective exam score transformations (optimal weights of individual questions).

Table 1: Empirical results

Variable	Estimate M_1	sig.	Estimate M_3	sig.
(Intercept)	6.935987	*	6.333557	***
Bcount	0.033281		0.035939	***
Gender	-2.166419		-1.465320	
Pop	-4.616769		-0.494553	
nnzfg	0.027379	*	0.030161	***
Future	0.625812	.	0.639711	***
Reflection	-0.167591		-0.167980	**
Bcount:Gender	-0.027786		-0.036090	**
Bcount:Pop	0.018510		-0.007901	
Gender:Pop	3.699211		2.116220	
Gender:nnzfg	-0.001449		-0.013074	**
Pop:nnzfg	-0.020088		-0.024768	***
Gender:Future	-0.274359		-0.354713	*
Pop:Future	-0.038318		-0.143013	
Gender:Reflection	0.161042		0.225622	*
Pop:Reflection	0.289011		0.160574	.
Bcount:Gender:Pop	0.019236		0.021735	
Gender:Pop:nnzfg	-0.002538		0.009896	.
Gender:Pop:Future	-0.248325		-0.157158	
Gender:Pop:Reflection	-0.128991		-0.160408	
Residual standard error	3.446		0.9593	
Degrees of freedom	179		179	
Adj. R-squared	0.1607		0.6626	
F-statistic	2.995	***	21.47	***

Signif. codes:
0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

From the results in Table 1 it is clear that M_3 provides - unlike M_1 - a lot of interesting information about the relationship between optimally weighted exam scores and the exogenous variables which are

under the control of the educator. The percentage of variance explained (adjusted R^2) in M_3 is more than 66% which allows us to make much better predictions than what is usually reported in - otherwise excellent - academic articles [25]. As explained before, the traditional F-test cannot be used to test the significance of the improvement. However, the auxiliary regression's null hypothesis $H_0 : g = 0$ is rejected even if an extremely low type I error is chosen (the p-value is 3.23×10^{-11}). This implies that M_3 performs significantly better and the objective exam score transformations are necessary. In addition, several diagnostic tests about the final model (M_3) are shown in Figure 4 - they indicate no statistical inadequacies.

The most interesting aspects of this analysis are the estimated parameters of M_3. With regard to quality control of the SLE the following conclusions can be made:

- There is a positive effect of performing reproducible, statistical computations (Bcount). This effect is significant at the 0.1% type I error level and cannot be measured without optimal weights (M_1). However, this effect is only relevant for female students because the parameter that is associated with Bcount:Gender is also significant and has a negative sign.

- Submitting feedback messages (in peer review) is very beneficial and improves exam scores (p-value $< 0.01\%$). This effect is about twice as large for female students than for males (the Gender:nnzfg parameter partially offsets the effect for male students). In addition, students from the switching population benefit less from feedback submissions.

- The reported "intention to use" (as measured in the usability survey) positively affects exam scores. This effect is strongest for female students. Note that previous research has shown that intention is mainly related to student's perception about the comparative advantage (of the software system) to learn statistics as compared to other alternatives (such as textbooks) [27].

- Females who report a high number of submitted feedback messages have significantly lower exam scores. On the other hand, male students who exaggerate their efforts are not in danger of having lower exam scores. This implies that the female exaggeration bias is small but harmful - the male exaggeration bias is big and harmless.

Based on these empirical results it is now possible to control (improve) the quality of the SLE:

- Female students should be encouraged to generate more reproducible computations.

- Peer review (based on Reproducible Computing) is highly beneficial to learn statistics - especially when it requires students to engage in submitting feedback messages to their peers. Male students need to (at least) double their efforts (compared to females) in order to obtain the same effect. Students from the switching population also need more feedback submissions than bachelor students.

- It is important to explain the SLE to students - emphasizing the comparative advantages of the system and the potentially improved exam scores. However, male students need more (or better) arguments before they accept the new technology and exhibit an increased degree of "intention to use."

- Female students who exaggerate their reported efforts should receive accurate feedback about their real performance which is based on objective measurements. Self assessment and reflection about student's actual efforts (as compared to perceived efforts) should be an integral part of the SLE.

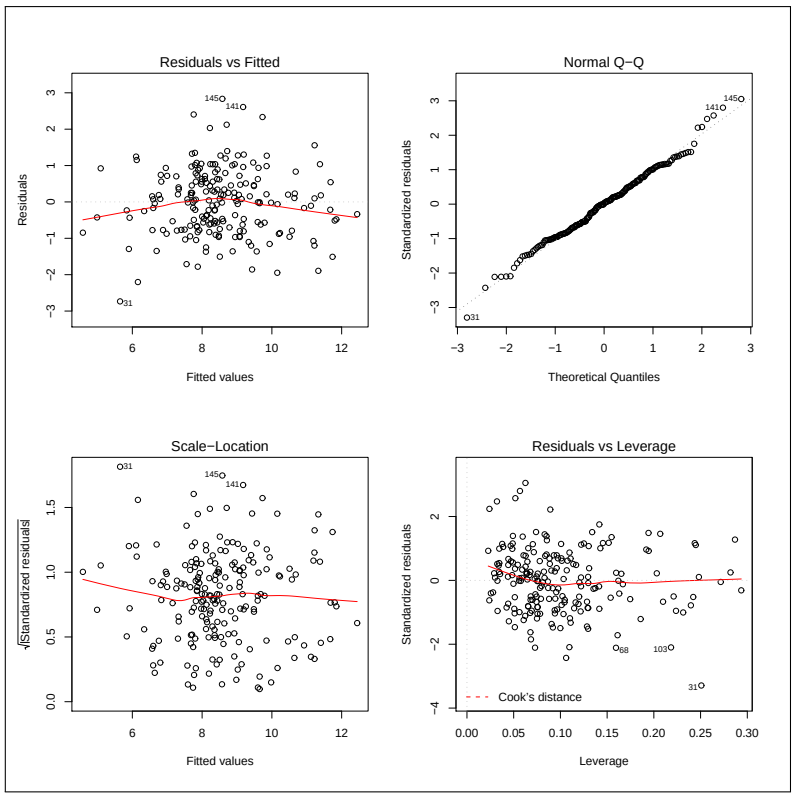

Figure 4: Diagnostics of M_3

5 Summary and Conclusions

The good news is that we now have a technology and methodology 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 SLE which includes the e-learning environment, the statistical software, the course materials, and the overall learning experiences of all students.

Bibliography

[1] 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

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

[3] Attitudes to Thinking and Learning Survey, (n.d.), Retrieved December 22, 2004, from www.moodle.org

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

[5] Buckheit J., and Donoho D. L. (1995), *Wavelets and Statistics*, Springer-Verlag, Editor: Antoniadis, A.

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

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

[8] Constructivist On-Line Learning Environment Survey, (n.d.), Retrieved December 22, 2004, from www.moodle.org

[9] de Leeuw J. (2001), Reproducible Research: the Bottom Line, *Department of Statistics Papers*, 2001031101, Department of Statistics, UCLA., URL http://repositories.cdlib.org/uclastat/papers/2001031101

[10] Donoho D. L., and Huo, X. (2005), BeamLab and Reproducible Research, *International Journal of Wavelets, Multiresolution and Information Processing*, 2(4), 391-414

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

[12] 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*, 40(9/10) p745-766

[13] Gentleman R. (2005), Applying Reproducible Research in Scientific Discovery, BioSilico, URL http://gentleman.fhcrc.org/Fld-talks/RGRepRes.pdf

[14] Green P. J. (2003), Diversities of gifts, but the same spirit, *The Statistician*, 52(4), 423-438

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

[16] Hollander M., and Wolfe D. A. (1973), *Nonparametric statistical inference*, New York: John Wiley & Sons., 185-194 (Kendall and Spearman tests).

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

[18] Koenker R., and Zeileis A.(2007), A., Reproducible Econometric Research (A Critical Review of the State of the Art), *Research Report Series*, Department of Statistics and Mathematics Wirtschaftsuniversit Wien

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

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

[21] Lucy D., Aykroyd R. G. and Pollard A. M.(2002), Non-parametric calibration for age estimation, *Applied Statistics* 51(2), 183-196

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

[23] Miller J. B., (n.d.), Examining the interplay between constructivism and different learning styles, Retrieved October 20, 2005 from http://www.stat.auckland.ac.nz/ĩase/publications/1/8a4_mill.pdf

[24] 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)

[25] 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)

[26] Peng R. D., Dominici F., and Zeger S. L. (2006), Reproducible Epidemiologic Research, *American Journal of Epidemiology*, 163(9), 783-789

[27] Poelmans S., Wessa P., Milis K., Bloemen E., and Doom C. (2008), Usability and Acceptance of E-Learning in Statistics Education, based on the Compendium Platform, *Proceedings of the International Conference of Education, Research and Innovation*, International Association of Technology, Education and Development

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

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

[30] Smith E. (1999), Social Constructivism, Individual Constructivism and the Role of Computers in Mathematics Education, *Journal of mathematical behavior*, 17(4)

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

[32] Statistical Computations at FreeStatistics.org (2008b), Office for Research Development and Education, Retrieved Mon, 30 Jun 2008, URL http://www.freestatistics.org/blog/date/2008/Jun/30/t12148409608o0dnj2k4s04jil.htm

[33] Statistical Computations at FreeStatistics.org (2008c), Office for Research Development and Education, Retrieved Mon, 30 Jun 2008, URL http://www.freestatistics.org/blog/date/2008/Jun/30/t1214841152sn6jlyhgseclgqm.htm

[34] 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

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

[36] 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)

[37] Wessa P. (2008b), Learning Statistics based on the Compendium and Reproducible Computing, *Proceedings of the International Conference on Education and Information Technology*, Berkeley, San Francisco, USA

[38] Wessa P. (2008c), How Reproducible Research Leads to Non-Rote Learning Within a Socially Constructivist E-Learning Environment, *Proceedings of the 7th European Conference on e-Learning*, Cyprus

[39] Wessa P. (2008d), Measurement and Control of Statistics Learning Processes based on Constructivist Feedback and Reproducible Computing, *Proceedings of the 3rd International Conference on Virtual Learning*, Constanta, Romania

[40] Wessa P. (2009a), Discovering Computer-Assisted Learning Processes based on Objective Exam Score Transformations, *Proceedings of the World Congress on Educational Sciences*, Cyprus