How Reproducible Research Leads to Non-Rote Learning Within a Socially Constructivist E-Learning Environment

Patrick Wessa
K.U.Leuven Association, Lessius Dept. of Business Studies, Belgium
patrick@wessa.net

Abstract: This paper discusses the implementation of a new e-learning environment that supports non-rote learning of exploratory and inductive statistics within the paradigm of social constructivism. The e-learning system is based on a new computational framework that allows us to create an electronic research environment where students are empowered to interact with reproducible computations from peers and the educator. The underlying technology effectively supports social interaction (communication), knowledge construction, collaboration, and scientific experimentation even if the student population is very large. In addition, the system allows us to measure important aspects of the actual learning process which are otherwise unobservable. With this new information it is possible to explore (and investigate) the effectiveness of e-based learning, the impact of software usability, and the importance of knowledge construction through various feedback and communication mechanisms.

Based on a preliminary empirical analysis from two courses (with large student populations) it is shown that there are strong relationships between actual constructivist learning activities and scores on objective examinations, in which the questions assess conceptual understanding. It is also explained that non-rote learning is supported by the fact that the system allows users to reproduce results and reuse them in derived research that can be easily communicated.

Keywords: statistics education, reproducible research, social constructivism, non-rote learning

1. Introduction

Within the context of ICT-based and math-related education, the pedagogical community has shown great interest in the role and importance of social and individual constructivism (Von Glasersfeld (1987), Smith (1999), Eggen et al. (2001)) and its implementation in statistics education in particular (Nyaradzo Mvududu (2003)). The following citation may clearly summarize the importance and the great interest of educational researchers in constructivism (Miller 2002):
Constructivism is a philosophy that supports student construction of knowledge. Since students uniquely construct their knowledge, instructional strategies that support constructivist philosophies naturally advocate student understanding. Instructional trends in the mathematics and statistics education communities support the active-learning orientation of constructivist philosophy. I posit that, while not the only philosophy of teaching and learning, constructivism is one of the best such philosophies.

While the relevance of a constructivist pedagogical approach to statistics education 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. (source: de Leeuw, 2001). The importance of the irreproducibility problem has been highlighted by many authors and is related to science, the dissemination of science, and academic education. 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 have been described in Buckheit and Donoho (1995); Donoho and Huo (2004); Leisch (2003).

If academic statisticians find it hard (if not impossible) to verify or review the results in empirical papers, how could we possibly expect students to learn from statistical results without the proper tools to easily review, verify, or challenge them? The solution that I propose within the context of this paper is new and differs from previously developed solutions in the sense that it can be used by anyone and without the need to understand the technicalities of scientific word processing (LaTex) or statistical programming (R code). Such a novel approach is obviously needed when one hopes to support students in their quest to learn and understand important statistical concepts.

The research presented in this article bridges two seemingly separate worlds and describes the implementation of a new e-learning environment that effectively supports statistics education through reproducible research within the constructivist pedagogical paradigm. The outline of the paper is straightforward. Section 2 clearly defines the major conceptual aspects and the infrastructure of the proposed approach while section 3 discusses the integration of the various ICT components. Section 4 provides the preliminary empirical evidence that clearly indicates that the proposed approach is effective and that a thorough investigation promises to yield interesting results in future research.

2. A new e-learning approach

There are several reasons why the constructivist approach may lead to non-rote learning. Such explanations however cannot be empirically tested if they are not defined in a precise and measurable form. Likewise, there is no way to provide empirical evidence to sustain the claim of this article's title without clear descriptions that can be easily implemented and measured. Therefore, I introduce operational descriptions of the concepts that are needed to construct testable hypotheses.

2.1 E-learning environment

The open source software called *Moodle* (which is freely available at http://www.moodle.org/) was used as the Virtual Learning Environment. The are several reasons why this software was chosen:

- it is designed to support social constructivism featuring various tools for communication, collaboration, assessment, interaction, etc...;
- it is well-written and has an open, modular design which allows us to seamlessly integrate other software components into the learning environment;
- it has a well-structured database design which allows researchers to easily retrieve data for research purposes.

The core section of the courses involved various activities (workshops) that require a lot of research and reflection about a variety of problems at various levels of difficulty. The workshops have been carefully designed and cannot be solved without additional information that is provided within the *Moodle* course or by the tutor. It is for this reason that these problem-oriented workshops and their subsequent lectures are of a “reflective” nature.

The courses that were offered contained a wide variety of statistical techniques and methods. The following topics were covered: probability, descriptive statistics, explorative data analysis, hypothesis testing (about the mean, the variance, and proportions), multiple linear regression, and introductory time series analysis. One could argue that it is rather unusual to treat so many topics in an introductory course. It is however very important for students to learn that statistical problems can be analysed in different ways – based on different techniques. For this reason I introduced a total of 73 different types of techniques with a variety of model parameters which yield a very large number of combinations.

For each technique students had one or several web-based software modules available. The modules are based on the *R Framework* and are available free of charge at http://www.wessa.net/. 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, 2008).

One of the pedagogical advantages of using the *R Framework* is that there is no need for students to understand the underlying statistical code while the computation is still transparent and flexible because the R code can be viewed and even edited by any knowledgeable user. In addition, there is no requirement to download or install anything on the student's computer because all computations are performed within a network of dedicated servers. In other words, anyone with an internet connection can use the computational system for the purpose of research and education. The output that is generated by the statistical software consists of tabular text and charts.

Each technique is described in a series of learning resources that were made available to students in a *Moodle* course. More than 4300 A4 sized pages were made available in electronic form to the students. Several search mechanisms were available to find relevant information which was always presented in modular form (without the requirement to read preceding chapters). One example of such a learning resource is the e-Handbook of Statistical Methods which is freely available from NIST/ SEMATECH (2006) at http://www.itl.nist.gov/div898/handbook/. Another example is the website http://www.xycoon.com/ that contains formal information about a large number of descriptive statistics, hypothesis testing techniques, econometric methods, and tools for time series analysis. The learning resources contain examples, case studies, mathematical proofs, formal properties, and verbal descriptions about the techniques that are available in the statistical software. Most importantly, the underlying assumptions of each technique are described in detail and can be quickly found through simple searches.

2.2 Dynamics of social constructivism

The proposed system was thoroughly tested in two different student populations: 111 Bachelor students, and 129 "Switching" students who already have a professional bachelor degree and registered for a (mandatory) preparation programme before switching to an academic master. The programme of study for both populations involves applied economics and business courses. Statistics is treated as an important and compulsory subject because students are required to engage in empirical research in later years (Bachelor thesis and Master thesis).

All students had to submit their workshop assignments at weekly intervals. During the lectures I illustrated frequently made mistakes based on sample submissions, and explained new methodological issues that might be helpful to solve the problems that students encountered. At the end of each lecture, I provided an introduction into the next workshop assignment. Students had the opportunity to ask questions during the lectures, or through the on-line forum that was supported by *Moodle*.

After each lecture, students worked on their next assignment and provided a well-motivated assessment of the submissions from the previous week (double-blind peer assessment). Even though students had to assess the submitted workshops and give them a score, the peer review was not intended as an evaluation method (it did not count towards their final score). On the other hand, it enabled students to provide feedback, learn from mistakes made by others, communicate solutions about a variety of problems, and provide an incentive in the form of encouragement to fellow students. This feedback-oriented process is similar to the peer review procedure of an article that is submitted to a scientific journal. The process of (anonymous) assessment by peers is an intrinsic part of scientific endeavour, and may help students in nurturing their scientific attitudes (through peer review experiences) and non-rote learning (through construction of knowledge).

Peer assessments have been performed for each workshop and by students from both populations. Switching students had to complete a series of 12 workshops of which the second half was completed by the Bachelor students too. A total of 1907 workshops were completed and subjected to peer review. Every submission was sent to a group of 5-7 students and every review involved between 3 and 6 assessment criteria (questions) that students had to grade. For every graded question students had the ability to provide verbal feedback to the other student.

As a consequence, a total of 41960 grades and 34438 verbal feedback communications were received by students. This implies that, on average, 22 grades and 18 verbal feedback messages were generated for each submitted workshop without any intervention by me. The administration of the peer assessment procedure was automated and fully supported by the *Moodle* software. The grades that were generated by the peer review process did not count towards the final score of students. Instead, I graded the quality of the verbal feedback messages that were submitted to other students based on semi-random sampling techniques.

The semi-random sampling technique is based on various statistics that are automatically produced by the *Moodle* software about submitted reviews. Each review is accompanied by a score which can be easily compared to the scores that were given by other students. For instance, if five (out of a total of six) reviewers submit a grade which is "excellent" and only one students rates the work under review with a "poor" grade then this discrepancy can be immediately detected in the overview screen which is created by *Moodle*. In such a case I would grade the quality of the feedback that accompanies the “poor” grade and two random feedback messages that correspond to “excellent” grades. For reasons of fairness, I make sure that every student's feedback is reviewed (by me) a sufficient number of times.

It is important to emphasize that this grading process was a powerful incentive for students to take the review process seriously. Moreover, the process of verbalisation was an important learning activity that required students to thoroughly investigate the research that was presented by peers.

For obvious reasons, this educational approach (in which students play the role of an active scientist) is only possible if students are empowered with all the necessary tools to exactly reproduce computational results and reuse them in derived work. Hence, a solution for the irreproducible research problem is a *conditio sine qua non* for the creation of an effective learning environment based on the review of submitted research results.

2.3 Reproducible research

Truly reproducible research has to be presented in such a way that any reader is able to confirm the results by recomputing the underlying statistical analysis. This is only possible if the author of research results includes all the meta information (data, parameters, and statistical software) that is necessary to reproduce the analysis into the document that is used for dissemination. Obviously this involves a lot of work for any author (student or scientist). Therefore it was necessary to build an automated procedure that keeps track of all the meta data that is needed to ensure reproducibility so that it can be instantly packaged, transmitted, and stored.

Within the context of the proposed e-learning environment I define a *Compendium* as 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. These objects are archived in a repository (*Compendium Platform*) that is available free of charge at http://www.freestatistics.org/ and which is funded by the OOF 2007/13 project of the K.U.Leuven Association.

There are some unique features of the *Compendium Platform* that are of particular importance in the e-learning environment that is proposed:

- any computation that is created within the *R Framework* can be easily archived in the repository – 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 browser (and an active connection) is required to use the repository;
- educators and researchers are able to retrieve data for research purposes.

With the *Compendium Platform* the process of reproducing computations has become easy and transparent at the same time. This allows students and educators to focus on the interpretation of computational results instead of the underlying technicalities. At the same time, this does not imply any limitation towards advanced students: they are still able to observe and reuse the R code that was used.

2.4 Non-rote learning

The final examinations that were employed in the courses measured analytical skills and conceptual understanding of statistical methodologies rather than the ability to reproduce theoretical aspects, use mechanical rules, or apply cookbook recipes that were memorized. The following three learning goals were specified to define true (non-rote) learning within the context of these introductory, undergraduate statistics courses:

- the ability to *select* one or several *appropriate technique(s)* to analyse a statistical problem;
- the ability to *read computational output* (of software) and correctly interpret it in terms of the problem to be solved;
- the ability to *check* the *underlying assumptions* of the employed technique(s).

Shortly before the final examination, students received a *Compendium* containing raw, non-chronological computer output about the analysis of a dataset that was never before discussed in class. Students were allowed to study the computer output, make notes, and bring all types of documents, text books, and “unconnected” laptops to the exam which had a duration of two hours.

The actual exam consisted of 18 multiple choice questions about the raw computer output in the *Compendium*. All questions had an unambiguous right/wrong answer but students were allowed to write an explanation if there was any doubt about the exact interpretation of the question. In addition, students were allowed to skip questions in order to avoid the guessing penalty: the exam scores were obtained by subtracting the number of wrong answers from the number of right answers. Most questions required students to examine multiple computations (based on different techniques) and careful interpretation to come to the correct (and unique) solution. Two questions were extremely difficult to solve – therefore, any student with an exam score that is equal or greater than 8 is considered to have passed the test.

3. Integrating the e-learning components

The three major components (*R Framework*, *Compendium Platform*, and *Moodle*) can be operated independently or in combination. A series of automated communication mechanisms allows each component to transmit information to the other component. Therefore each component is able to perform tasks in a student-friendly manner and at the same time it provides valuable data for the purpose of educational research. Table 1 provides an overview of how the communication interfaces have been implemented. For each component a brief discussion of the technological implementation is provided.

Table 1: Communication mechanisms between the three components of the new e-learning environment

How do the row-components communicate with the column-components?			
	Moodle	*R Framework*	*Compendium Platform*
Moodle	Moodle Session id	UserID and CourseID are trasmitted through HTTP GET request	UserID and CourseID are transmitted through HTTP GET request
R Framework	Stored Moodle session id is used in HTTP GET	R Framework Session id	User Session Data (incl. software, data, parameters) through a HTTP GET callback mechanism
Compendium Platform	Stored Moodle session id is used in HTTP POST	Stored User Session Data (incl. Software, data, parameters) is submitted through a HTTP POST request	Repository Session id

Let us now have a look at a brief example that illustrates how the three components communicate: a selected sample of the employed learning resources was made freely available and can be consulted in a *Moodle* course at http://www.freestatistics.org/moodle/ (click on “Open Course Materials” and login as guest user). Suppose a student wants to review the solution to exercise 1.13 (available under section 2 of the on-line course). For this purpose I created a tailor-made R module which solves this particular problem and allows students to experiment with various parameter settings. If a student clicks on the hyperlink (called “The Babies Calculator”) in the *Moodle* course then the respective R module (based on the *R Framework*) is shown in a separate window which contains an URL that contains two tags:
http://www.wessa.net/rwasp_babies.wasp?protag=Open+Course+Materials&utag=Guest+User
These tags identify the user (“Guest User”) and the course (“Open Course Materials”). Both tags are stored in server-side sessions on the wessa.net web server and allow us to attribute subsequent computational actions to the actual user who submitted the requests. This clearly illustrates, as indicated in Table 1, that *Moodle* communicates with the *R Framework* through a simple HTTP GET request where the UserID and CourseID is contained.

Now, suppose that the student clicks on the Compute button in the R module. The *R Framework* receives the submitted request and instantly creates pre-processed R code which is stored in the web server's local cache. Now a special load-balancing software is invoked which selects the remote machine that has to execute the computation from a list of dedicated R servers. The wessa.net web server downloads the computational result from the R server and creates a nicely formatted result page based on a template and sends it back to the student. This process has very favourable properties in terms of performance, scalability, and security (Wessa, 2008). In addition, all computational results (including the UserID and CourseID) are stored in a session database of the *R Framework*.

Suppose that the student wants to include the computational results in a paper in such a way that anyone can verify, reproduce and reuse them. The student clicks on the hyperlink “Click here to blog (archive) this computation (opens new window)” and fills out a simple submission form. When the student clicks the submit button the *R Framework* will retrieve the stored information from the session database and create a package that can be safely transmitted. It then calls a remote procedure at the *Compendium Platform* which downloads the package through an HTTP GET callback (see Table 1). The *Compendium Platform* stores the packaged computation in its repository and creates records about important meta data and keywords that allow for various types of queries to be executed.

If all goes well, the student will see a result page with a hyperlink to the archived computation. The student can visit this link and view the html page that provides a summary of the computed analysis. In this example the system generated the following reference that can be inserted into any document (Statistical Computations at FreeStatistics.org, 2008):
http://www.freestatistics.org/blog/date/2008/Jun/06/t12127572549onpj8u7m2ygvcq.htm/

The fact that this link has been inserted into this article makes it (by definition) a *Compendium*. Now any reader is able to reproduce the simulation experiment that was originally conducted (just click the link to try). Note that the analysis is based on simulation techniques: the obvious implication is that the reproduced computations may slightly differ from the archived result.

4. Preliminary empirical evidence

This section provides preliminary empirical evidence that supports the claim made by the title. The purpose of this analysis however, is not to find definitive answers but to foster discussions about the pedagogical implications and about directions in future research.

4.1 Hypotheses

Based on previously defined concepts and data descriptions it is now possible to formulate two statistical hypotheses that can be tested.

Hypothesis 1. H0: the number of submitted (verbal) feedback messages (about the workshops of peers) is not associated with exam scores.

Hypothesis 2: H0: the number of received (verbal) feedback messages (about the student's workshops) is not associated with exam scores.

If learning occurs through the “active” construction of knowledge then the test should reject the first null hypothesis because the verbal formulation of feedback about workshops requires students to have constructed a sufficient level of understanding. The argument here is that students who don't understand the statistical concepts, allowing them to write meaningful feedback, will just submit a grade with an empty feedback text. As explained in section 2.2 there were 34438 verbal feedback messages out of a total of 41960 grades. This implies that 18% of all grades (7522 grades) were not accompanied by text. Students knew that I would grade the quality of (a sample of) their feedback so they had every reason to make the feedback messages meaningful. I can confirm that almost all feedback messages that I rated were meaningful and (to some degree) intended to provide moral support. It is also important to emphasize the fact that meaningful feedback can only be written if results from peers are reproducible and reusable. Hence, the number of submitted feedback messages is a proxy measure for the ability of the student to construct knowledge based on reproducible research. If this variable is associated with objective exam scores (measuring conceptual understanding instead of rote memorization) then we can reject the null hypothesis and conclude that the title of this article is justified.

If learning occurs through “passive reception” of explanations or feedback then the test should reject the second null hypothesis. Such a rejection would imply that true understanding can be fostered through the reading of feedback. If the second hypothesis is rejected and the first is not rejected, then the *Compendium Platform* should be primarily used to create course materials instead of a simulated research environment where research results are challenged through peer review.

The main difference between active and passive modes of learning is related to responsibility. In active (constructivist) learning the student is responsibly engaged in learning activities because the e-learning environment allows the educator to track, verify, and accurately measure the learning activities and processes. In passive learning the student completes the assignment and then waits for a reply in the form of feedback. Even if the feedback contains valuable information then there is no guarantee that the student actually makes good use of it.

In this sense, there are interesting analogies between statistics learning and scientific research:

- reproducibility of research leads to honesty and responsibility;
- peer review (grading) of reproducible research leads to quality output;
- reviewing the work of peers (and writing meaningful feedback) is very demanding but at the same time potentially edifying.

4.2 Analysis

The exam scores that represent non-rote learning have been cut into three mutually exclusive intervals. The lowest interval]-3,4] represents scores that could be associated with pure guessing. The second interval]4,7] contains scores that are insufficient but unlikely to be attributed to pure guessing. The third interval]7,18] represents scores where students have passed the exam. Note that this exam only accounts for 50% of the final scores that students received. For the purpose of testing both hypotheses however, it is important that we only use the objective exam scores.

The number of submitted and received feedback messages have both been cut into two mutually exclusive intervals ("low" and "medium/high"). In each case the cut-off point was chosen such that minimum frequency requirement (in each cell) was satisfied.

Table 2: Reproducible Computations - Two-dimensional Contingency Table - by population

Are review messages about reproducible research related to exam scores?				
Bachelor	# Submitted Verbal Feedback Messages		# Received Verbal Feedback Messages	
Exam Score	(0,100]	(100,450]	(0,100]	(100,450]
(-3,4]	12	6	10	11
(4,7]	10	14	8	16
(7,18]	14	45	16	44
X-squared	11.58		3.13	
df	2		2	
p value	0.00305		0.20891	
Switching	# Submitted Verbal Feedback Messages		# Received Verbal Feedback Messages	
Exam Score	(0,150]	(150,450]	(0,170]	(170,450]
(-3,4]	11	8	7	12
(4,7]	12	19	13	20
(7,18]	14	59	14	59
X-squared	12.21		5.74	
df	2		2	
p value	0.00223		0.05663	
(click to reproduce this computation)				

Table 2 presents the analysis of two-dimensional contingency tables and Chi-square tests for both hypotheses. Each test was performed for the Bachelor and Switching student population separately.

It is clear that the first hypothesis should be rejected for both student populations (left side of Table 2). The p-values are extremely small which leaves no room for doubt. The results are preliminary and do not provide proof of a causal relationship. However, for the purpose of presenting the new e-learning environment, it represents a very strong indication that the creation of the *Compendium Platform* was a good investment and that a detailed analysis of the database in future research is well worth the effort. On the right side of Table 2 we can see that the second hypothesis should not be rejected unless a high type I significance threshold is employed. Depending on the actual cut-off points that define the categories, the p-value for the Switching students might fall (slightly) below the 5% level. The p-value for the Bachelor students however, never falls below a two-digit percentage.

5. Conclusions and Future Research

The proposed e-learning environment has various unique properties that support statistics learning within a constructivist setting:

- The *R Framework* allows students to perform any type of statistical analysis without the requirement to understand the underlying technicalities and without the need to download/install any executable code on their computer.
- The *Compendium Platform* allows students to archive, reproduce, and reuse computations. In addition, students can easily create/maintain Compendia of reproducible research which support various forms of constructivist learning activities (communication, collaboration, and peer review).
- All computational features have been seamlessly integrated into the *Moodle* learning environment. The three independent systems are perceived as a single e-learning environment by students.

From a pedagogical point of view it was demonstrated that reproducible research allows students to engage in peer review activities which leads to non-rote learning. At the same time the proposed technology presents us with a unique research opportunity to investigate statistics learning based on actual learning activities which are otherwise unobservable.

Taking into account the results from this analysis, I propose that future research should focus on (but not be limited to) the following questions:

- Which other proxy variables could be used instead of the count of submitted feedback messages?
- Could we find a measure for quality of feedback?
- How are these findings related to other data that is available (software usability, computational statistics, learning attitudes, group behaviour, learning experiences)?
- Can we induce causation? Are there any confounding variables that might result in spurious associations? For instance: the median workshop score is an excellent proxy variable that reflects prior knowledge of students.
- What are the best predictors for non-rote learning?

Acknowledgements
The Compendium Platform is funded by the OOF 2007/13 project of the K.U.Leuven Association.
I would like to thank Ed van Stee for his useful comments and suggestions.

References

de Leeuw, Jan (2001), “Reproducible Research: the Bottom Line”, *Department of Statistics, UCLA. Department of Statistics Papers*, nr. 2001031101, http://repositories.cdlib.org/uclastat/papers/2001031101

Donoho, D. L., Huo, X. (2004), “BeamLab and Reproducible Research”, *International Journal of Wavelets, Multiresolution and Information Processing*

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

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

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

Koenker, R., Zeileis A. (2007), “Reproducible Econometric Research (A Critical Review of the State of the Art)”, Department of Statistics and Mathematics Wirtschaftsuniversität Wien, Research Report Series, Report 60

Leisch, F. (2003), “Sweave and beyond: Computations on text documents”, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, ISSN 1609-395X

Miller, Jacqueline B. (2002), “Examining the interplay between constructivism and different learning styles”, Retrieved October 20, 2005 from http://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

NIST/SEMATECH, “e-Handbook of Statistical Methods”, http://www.itl.nist.gov/div898/handbook/, 2006

Peng, R. D., Francesca Dominici, and Scott L. Zeger (2006), “Reproducible Epidemiologic Research”, *American Journal of Epidemiology*

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

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, Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2008/Jun/06/t12127572549onpj8u7m2ygvcq.htm/

Statistical Computations at FreeStatistics.org, Office for Research Development and Education, URL http://www.freestatistics.org/blog/date/2008/Jun/14/t1213478471rqn93e3pbymvbmg.htm/

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. (2008), “A framework for statistical software development, maintenance, and publishing within an open-access business model”, *Computational Statistics*, original publication is available at www.springerlink.com (DOI 10.1007/s00180-008-0107-y)