How to Model the Design Efficiency of the VLE?

Patrick Wessa

Catholic University of Leuven Association,
Lessius Dept. of Business Studies, BELGIUM
E-mail: patrick@wessa.net

Abstract

This article discusses the use of a predictive decision model about a new type of statistical learning technology which is based on Reproducible Computing. The model predicts discretized exam outcomes based on objectively measured learning activities that are embedded within the pedagogical paradigm of social constructivism. However, the main contribution of this study is based on a quasi-experiment in which the pedagogical efficiency of two competing software design models are compared. In the first system, all learning features are a function of the classical Virtual Learning Environment (VLE). In contrast, the second system is designed from the perspective that learning features are a function of the course's core content (c.q. statistical results). The ceteris paribus effect of the design change (from VLE-based to Content-based) is shown to substantially increase the efficiency of constructivist, computer-assisted learning activities for all cohorts of the student population under investigation. These results may, if confirmed in other circumstances, have important repercussions for the design of future learning environments.

Keywords: Reproducible Computing, VLE, Software Design, Communication

Introduction

Beyond any doubt, there has been a growing interest in Computer Assisted Learning (CAL) in the academic community. Most pedagogical studies however take the system design of the Virtual Learning Environment (VLE) for granted. This is surprising because the efficiency of CAL may be strongly influenced by the VLE's design which is typically beyond the control of the educator.

This study aims to demonstrate that - within the context of undergraduate statistics education - the design effect is measurable and potentially substantial. In order to achieve this goal, a two-year quasi-experiment was setup within the context of an undergraduate statistics course which is embedded in a socially constructivist setting.

The typical, modern VLE integrates a wide variety of general-purpose CAL techniques which are clustered around a course. In this sense the VLE is supposed to be of a generic and course-oriented nature. While there may exist many reasons why such a design is beneficial, there are no guarantees that such VLEs are well-suited to build effective and efficient learning environments in the field of statistics. One of the reasons for this is the fact that most statistics courses involve statistical computing which is not readily available in the VLE. As a consequence, educators often rely on external statistical software products which are often hard - if not impossible - to "seamlessly integrate" into the VLE. It is not surprising that many statisticians have found it necessary to develop new statistical software for the purpose of building a specific-purpose Statistical Learning Environment (SLE).

In this study the VLE design is represented by *Moodle* which is well-known in the academic community, and has been designed within the pedagogical paradigm of social constructivism (Wessa, 2009c). Within the context of this study the design effect that is investigated relates exclusively on socially constructivist learning activities that are supported by the VLE. A design change of the peer review module (and associated communication feature) is the main component that is subjected to change and ex-post analysis. The details of the design change will be explained in the section 2.

The inability of scientists to reproduce empirical research that is published in papers, has received a great deal of attention within the academic community. Several solutions have been proposed but have not been adopted in education because of the inherent impracticalities therein (Wessa, 2009b). For this reason, a new Compendium Platform (CP), which is hosted at http://www.freestatistics.org, was developed and allows us to create constructivist learning environments which are based on reproducible computing (hosted at http://www.wessa.net), and based on the R language) and with several advantages that relate to the monitoring of actual learning processes and quality control (Wessa, 2009b).

Henceforth, the term SLE refers to the computational system that comprises the actual statistical software (R Framework), the Compendium Platform (and associated repository of reproducible computations), and all interfaces that allow users and other software systems to interact with the components that are contained therein.

Design

The investigation was based on an experimental, undergraduate statistics course for business students with a strong emphasis on social constructivism. The course contained a wide variety of statistical techniques and methods. For each technique, students had one or several web-based software modules available which are based on the R Framework. In order to implement this course within a setting of social constructivism for large student populations, it was necessary to impose a strict assignment-review mechanism. This is illustrated in figure 1 which shows a series of weekly events (lectures, assignments, reviews) during a thirteen-week semester. The semester ended with a final (open book) examination about a series of objective multiple choice questions. The examination was intended to test understanding of statistical concepts rather than rote memorization.

The main sections of the statistics course were built around a series of research-based workshops (WS1, WS2, ...) that require students to reflect and communicate about a variety of statistical problems, at various levels of difficulty. The workshops have been carefully designed and tested over a period of six years. Each workshop contained questions about "common datasets" and questions about individual data series - this dual structure of the workshops promoted both, collaboration between students, and individual work. The top (blue) puzzle pieces in figure 1 represent threaded communication (between students) about each workshop.

Each week there was a lecture (L1, L2, ...) which was held in a large lecture hall that was equipped with computer screen projection and internet facilities. During each week, students were required to work on their workshop assignment and - at the same time - perform peer reviews (Rev1, Rev2, ...) about six assignments that were submitted by peers. Each review was based on a rubric of a minimum of three criteria and involved students to submit a workshop score and an extended feedback message (yellow puzzle pieces). The grades that were generated by the peer review process did not count towards the final score of students. Instead, the educator graded the quality of the verbal feedback messages that were submitted to other students. The grading was performed based on a semi-random sampling technique which allowed the educator to grade the quality of a relatively small - but fairly representative - number of submitted feedback messages from each student.

This feedback-oriented process is similar to the peer review procedure of an article that is submitted to a scientific journal. The key idea behind this constructivist environment is that students are empowered to interact with reproducible computations from peers and the educator. Students are required to play the role of an active scientist who investigates problems, presents solutions, and reviews the work of peers. Obviously, Reproducible Computing is a *conditio sine qua non* that allows students to engage in such peer review activities.

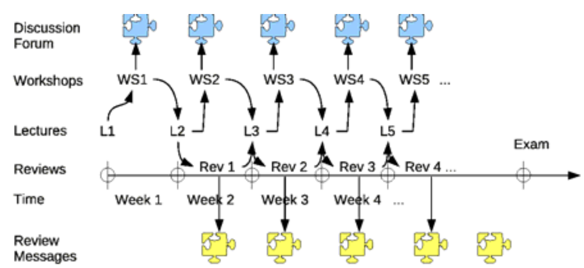

Figure 1 Schedule of learning activities - Year 0

Original System Design - year 0

Figure 2 displays the VLE and SLE as it was used in year 0 (fall semester of 2007). It is clearly seen that this design contained two core objects: the course (yellow) and the computation (blue) which is represented by its snapshot. The course is the core object of the VLE which implies that all features that allow students to engage in collaboration or communication are bound to the course in which they reside. Several forums and instant messaging facilities were available to ask questions or to collaborate in various ways. In addition the Peer Review & Assessment procedure was available from within the VLE.

There are however, several pedagogical problems with this type of design because students were unable to:

- engage in review activities when they view the meta information about a computation - instead they need to login to the VLE and invoke the features of the Peer Assessment module

- read review messages that are submitted by other students about their own work unless they use the VLE and their own Compendium simultaneously
- compare review messages of computations that preceded the ones that are currently under review
- discuss or review statistical analyses across courses or semesters - as soon as the course is closed, all communications contained therein are lost forever

In addition, the collaborative communications about the workshops (blue puzzle pieces in fig. 1) and the feedback messages of the peer reviews (yellow puzzle pieces) were completely separated which implies that working on assignments and learning through peer review were completely detached activities. Finally, and notwithstanding the fact that sequential workshops were related in various ways, there was no structural information about the dynamics of collaborative and review-based communications across workshops. For instance, if students were required to test a certain statistical assumption in an early workshop that was an essential condition to perform some type of analysis in a subsequent workshop, then there was no link between the communications of both. The only way that could have been used to solve this problem (within the current design) was to repeat previous analyses in all related, subsequent workshops. Unfortunately, such an approach would have been highly inefficient and unfeasible because of many practical limitations.

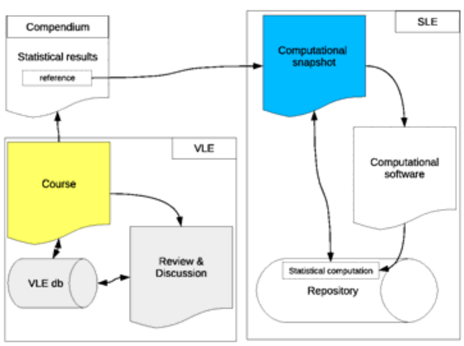

Figure 2 ***VLE/SLE Design - Year 0***

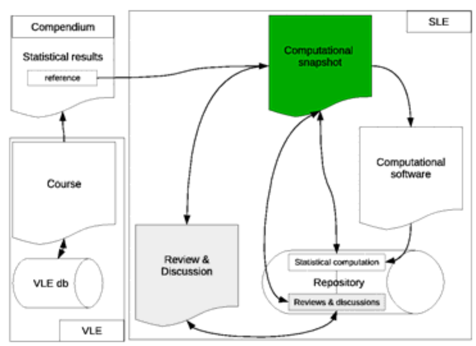

Figure 3 ***VLE/SLE Design - Year 1***

Alternative System Design - year 1

Figure 3 displays the alternative design that was implemented in year 1 (fall semester of 2008). The most important design changes are as follows:

- there is only one core object: the computational snapshot
- all (threaded) collaborative communications about the workshops are available within the computational snapshot (which becomes a dynamic webpage)
- all review messages are associated with the computational snapshot

The consequences of this design change had important consequences for the students because all collaborative and review-related communications were available from within the same source (the computation) which clearly highlights how they are related - as is shown in figure 4, the blue and yellow puzzle pieces within each computation are connected. This is not only true for a single computation - it also applies to discussion/review communications that relate to different computations, irrespective of the time frame, course, or workshop in which they originated. The reason for this is the fact that the Compendium Platform automatically stores and maintains the parent-child relationships that exist between computations. For instance, if the educator creates a Compendium with a worked example that is based on an original computation C1 (see figure 4) then a student may reuse this computations (with changed parameters or data) for the purpose of working on an assignment task (C2). At a later stage, the same (or any other) student may reproduce C2 (and create C3) in order to check the assumptions of a statistical analysis that is embedded in a subsequent workshop. Other students (across courses and years) may reuse C2 for similar purposes (computations C4-C6).

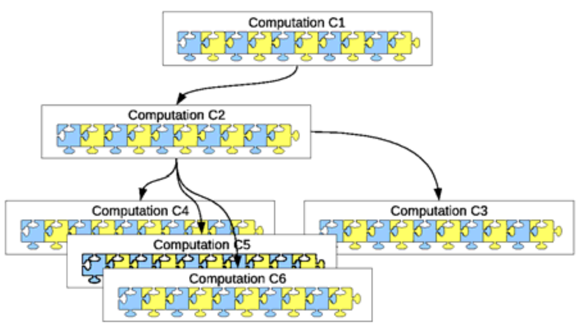

Figure 4 Hierarchical structure of computations - Year 1

The bottom line is that everyone who looks at C2 will have all the information that is available about computations C1-C6, including the hierarchical dependencies of computations and communications associated with them. This design change should increase the efficiency by which users can gain an understanding of statistical concepts and the dynamics of how computations evolve (and improve) over time. Unlike in the traditional setting (year 0) no information is ever lost after the semester because the communications are independent of the courses.

In general words, the fundamental principle that is applied in this VLE/SLE design is that the educational system is subject-oriented instead of course-oriented. In statistics education, it is the statistical computation that is subjected to study - the course is entirely irrelevant. The traditional VLE is an educator-centered system that allows the educator to manage students, and resources that belong to the course. The new SLE design is more student-centered because it is focused on the learning content which implies that all learning features (including communication, peer review, etc...) depend on the (subject-oriented) core object.

Methodology

Measurements

The empirical data was collected through an experimental undergraduate statistics course which was provided during two consecutive years. In each year, the conditions that are under the control of the educator (and the institution) were kept equal except for the system design. The (quasi) experiment is not under perfect control but given the fact that the characteristics of the student population did not change, it is fair to assume that conditions were equal in both years. Therefore it is fair to attribute any changes in learning efficiency (*ceteris paribus*) to the change in system design.

The measurements were obtained from a Business Studies department in Belgium during two consecutive years (labeled “year 0” and “year 1”). In each year there were two cohorts: bachelor students, and students from the preparatory program which allows graduates from a professional bachelor program to switch to an academic master. In general, bachelor students have better prior understanding of mathematical concepts than prep-students. However, prep-students tend to have a higher degree of maturity and self-motivation than bachelor students.

Table 1 Student Population

	Year 0		Year 1	
	Female	Male	Female	Male
Bachelor	58	53	41	42
Prep.	53	76	45	74
Total	240		202	

In order to be able to compare the dependencies of exam scores from exogenous variables that are based on objective measurements of (constructivist) learning activities, it is necessary to apply optimal exam score transformations for both years. The methodology that allows us to do this is based on a mathematical model which is described in (Wessa, 2009) and has been shown to yield models that improve the predictability of learning outcomes substantially.

After the objective exam score transformation has been applied, it is possible to proceed to the next step which involves the creation of predictive models (c.q. regression trees) that allow us to discover the rules that determine whether students will pass or not. In this study, the degree of predictability is maximized (through the transformation methodology) but is otherwise irrelevant to answer the main research hypothesis: “does the changed VLE/SLE design improve the efficiency of learning activities (such as peer review) in the undergraduate statistics course?” In other words, we are mainly interested in the (efficiency-related) parameters of the decision rules, not the original (untransformed) exam scores (which are incomparable), nor the overall degree of predictability.

Regression Trees

For the purpose of computing a rule-based regression tree, the endogenous variable must be discretized. Therefore, three categories are defined which are called “guess”, “fail”, and “pass” respectively. The “guess” category represents the lowest exam scores which can be attributed to chance (or guessing). Exam scores in the “fail” category are lower than what is needed to pass the exam but higher than what can be (reasonably) explained by chance. The “pass” category contains scores that are sufficiently high to be considered satisfactory even if the numerical value is below 50% of the maximum attainable score. The reason for this is the fact that the exam questions had varying degrees of difficulty and were (overall) designed to be much more difficult than what could be reasonably expected from undergraduate students in business studies.

Introducing a high degree of difficulty in the exam questions is necessary in order to ensure that:

- rote learners are not likely to pass the exam
- we are able to identify the maximum level of understanding
- students are unable to quickly find answers in printed resources that are allowed during the exam

The exam in the second year was slightly more difficult than in the first year (the transformed exam scores in year 1 were slightly lower than in year 0). Therefore it is not possible to simply use identical threshold values for the categories in the transformed exam scores from both years - an objective benchmark is need to generate fair and comparable categories.

The threshold values that define the categories are not arbitrarily chosen but depend on exam score statistics of the previous four years (with exams of similar difficulty). On average the proportion of lowest scores (which fall in the "guess" category) was little less than 10%. The proportion of "guess and fail" scores was approximately one third of all exam scores. These proportions had been quite stable over the time frame of those four years. Therefore it is fair to assume that they represent appropriate, "unconditional" probabilities to pass or fail the exam. As a consequence the threshold values that define the three categories (for each year) are computed as the 1/10 and 1/3 quantiles of the (optimally weighted) exam scores in year 0 and 1.

Even if we wouldn't believe that the threshold values are adequate there is another justification of using the same quantiles (rather than identical exam scores) to determine the categories. The rationale is simply that we want to predict if students fall in the "high", "low", or "extremely low" proportion of all students in the same year (who took the same exam). The parameters in the rule-based regression trees quantify the amount of learning efforts (number of peer review messages, and number of computations) that are required to achieve an exam score that falls within the top 2/3 of all scores.

The rule-based regression trees were computed with the statistical engine called *Weka* which is available from within the R Framework through the *RWeka* interface (Hornik et al., 2009).

Empirical Results

Table 2 Nomenclature in rule-based regression trees

Variable name	Description
nnzfg	# of non-empty, meaningful feedback messages that were submitted
nnzfr	# of non-empty, meaningful feedback massages that were received
Bcount	# of reproducible computations that were generated
Gender	binary gender variable (0 = females, 1 = males)
Pop	binary cohort variable (0 = bachelor students, 1 = prep. students)

Table 2 shows the exogenous variables that were chosen to create rule-based regression trees. The first three variables are positive, numeric integers. The last two variables are binaries that indicate to which cohort the student belongs. Note that the same exogenous variables were used in the objective exam score transformations based on the three-stage regression approach and with all possible interaction effects included.

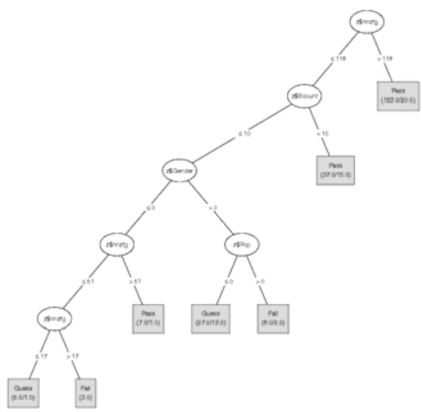

Figure 5 ***Regression Tree - Year 0***

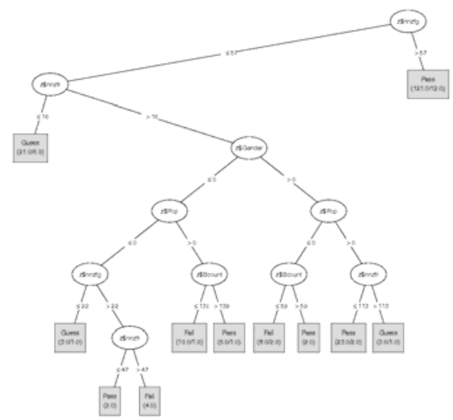

Figure 6 ***Regression Tree - Year 1***

The first rule-based regression tree (fig. 5) displays the situation for year 0 in which the traditional VLE design is used. The most important rule that determines success (c.q. falling into the top 2/3 proportion of all students in year 0) is the number of submitted feedback messages (related to peer review). It can be clearly seen that students pass if *nnzfg* > 118 which means that they need to submit more than 118 feedback messages in order to pass the exam. The other students (with *nnzfg* ≤ 118) fall into two categories, depending on the number of reproducible computations they generated. Students with *nnzfg* ≤ 118 and *Bcount* > 10 are predicted to pass the exam - in other words, students who did not engage sufficiently in feedback activities could compensate this by reproducing more than 10 archived computations. However, the accuracy of this particular prediction is not very good because there where only 37 cases correctly attributed to the "pass" category whereas 15 cases were incorrectly predicted (the number of in/correctly classified cases can be seen in the gray boxes).

There are two specific rules in the regression tree that cause concern. The first one, is the rule that states that male students who did not spend a sufficient amount of effort in terms of feedback and reproducing computations (*nnzfg* ≤ 118 and *Bcount* ≤ 10 and *Gender* = 1) either fall into the "guess" or "fail" category (depending on the *Pop* cohort they belong to). The second rule that causes concern is the one that states that female students may pass the exam, even if they have only between 52 and 118 submitted feedback messages (*nnzfg* ≤ 118 and *Bcount* ≤ 10 and *Gender* = 0 and *nnzfg* > 51).

The bottom line is that both rules imply that the VLE/SLE system in year 0 favors female students and discriminates against males. This may be surprising because it is often believed that male students have "better" attitudes towards computing than females.

In this situation however, it is shown that female students are better able to cope with the detached structure between collaborative and review-based communication on the one hand, and reproducible computing on the other hand. This phenomenon may have psychological causes that are related to the fact that there are gender differences in how students use communication in learning. Within the context of this study, such an explanation remains speculative and unanswered. However, and more importantly, it is clear that the design of the VLE and SLE is not optimal - at least for an important part of the student population.

Figure 6 shows the rule-based regression tree for year 1 (in which the new VLE/SLE design was implemented). It can be easily observed that the structure is fundamentally different from the previous situation. By far, the most important property of this regression tree is the root rule which states that students pass if they submit more than 57 meaningful feedback messages. This is less than half the amount that was necessary with the previous system design and demonstrates a spectacular increase in review-based learning efficiency. More importantly, the discrimination effect has completely disappeared which implies that males are now equally well able to make good use of the learning environment. Students who did not submit a sufficient number of feedback messages and only received 16 messages (or less) fall into the "guess" category. This makes a lot of sense because students who don't submit workshop papers, don't get reviews.

There is a striking resemblance between female prep-students and male bachelor students (fig. 6): they both pass the exam when a sufficient number of computations have been reproduced. In addition, the female bachelor students and male prep-students are also similar with respect to the number of received feedback messages: if this number is too high, then the student does not pass because it indicates that they are making too many mistakes or are not making good use of inbound messages.

As explained before the overall predictability (of both rule-based regression trees) is not an important aspect which determines if the design effect had any impact on learning efficiency. Nevertheless, an overview of within and out-of-sample prediction performance is provided in table 3 because it is important to show that the models do not suffer from severe "over-fitting" which might invalidate all conclusions made on the basis of the regression tree's parameters.

Table 3 Prediction Performance of Regression Trees

Statistic	Year 0		Year 1	
	Within Sample	Cross Validation	Within Sample	Cross Validation
Correctly Classified	78.3%	72.5%	87.1%	74.8%
Incorrectly Classified	21.7%	27.5%	12.9%	25.2%
Number of Leaves	7		11	
Size of Tree	13		21	
Total Number of Cases	240		202	

The results in table 3 clearly illustrate that the out-of-sample prediction quality is adequate. In case of over-fitting, one would observe high percentages of correctly classified instances within sample and a (very) low percentage out-of-sample. The out-of-sample prediction quality is computed by applying a so-called Cross Validation technique which randomly divides the data set into a large training subset and a testing subset. The parameters are estimated, based on the training sample and the prediction is computed for the testing subset. This procedure is repeated 10 times (10-fold Cross Validation) to obtain an average measure of out-of-sample prediction quality.

Conclusions

The empirical analysis has clearly shown that the change in VLE/SLE design had a very beneficial effect in terms of increasing the learning efficiency of submitting peer review messages. More importantly, the design change has resulted in the elimination of a discrimination effect which was embedded in the original design where communication and computation was separated. In any case, the methodology that was outlined can be used to test for any software-related or content-based aspect as long as it is controllable by the educator or designer of the learning system. However, one should take care to take into account that exam scores are properly treated in order to avoid the pitfalls that are associated with exam questions.

Obviously, this study is limited to the case of our undergraduate statistics course for business students. Also, there was a strong focus on one specific type of constructivist learning activity (peer review) which implies that other pedagogical approaches might have resulted to other conclusions.

Nevertheless, it is interesting to formulate a general conjecture about a fundamental principle of good VLE design. The proposed conjecture states that good VLE design requires the developer to define a single subject-based, core object instead of using the traditional, educator-centered course object. In simple words, it is better to integrate learning features (forums, messaging, peer review, etc...) into the software that treats the subject under study than to build general-purpose VLEs. If this conjecture would turn out to be true, it would have important repercussions for the design of VLEs in general and specific-purpose software (such as: statistical software, wikis, CAD/CAM applications, programming environments, etc...) in particular.

REFERENCES

HORNIK K., ZEILEIS A., HOTHORN T., BUCHTA C. (2009). RWeka: An R Interface to Weka. R package version 0.3-16. URL http://CRAN.R-project.org/package=RWeka

ROMERO C., VENTURA S., GARCIA E. (2008), Data mining in course management systems: Moodle case study and tutorial, *Computers & Education*, 51, 368-384

WESSA, P. (2009a), Quality Control of Statistical Learning Environments and Prediction of Learning Outcomes through Reproducible Computing, *International Journal of Computers, Communications & Control* 4(2)

WESSA, P. (2009b), Reproducible Computing: a new Technology for Statistics Education and Educational Research, *IAENG Transactions on Engineering Technologies*, American Institute of Physics, Eds: Rieger, Burghard, Amouzegar, Mahyar A., and Ao, Sio-Iong

WESSA, P. (2009c), How Reproducible Computing Leads to Non-Rote Learning Within Socially Constructivist Statistics Education, *Electronic Journal of e-Learning* 7(2)