

Explorative Data Mining of Constructivist Learning Experiences and Activities with Multiple Dimensions

Patrick Wessa and Bart Baesens

Abstract—This paper discusses the use of explorative data mining tools that allow the educator to explore new relationships between reported learning experiences and actual activities, even if there are multiple dimensions with a large number of measured items. The underlying technology is based on the so-called Compendium Platform for Reproducible Computing (<http://www.freestatics.org>) which was built on top the computational R Framework (<http://www.wessa.net>).

Keywords—Reproducible computing, data mining, explorative data analysis, compendium technology, computer assisted education

I. INTRODUCTION

THE inability of scientists to reproduce empirical research that is published in papers, has received a great deal of attention within the academic community [1], [2], [3], [4], [5], [6], [7]. Several solutions have been proposed ([5], [7], [8]) but have not been adopted in education because of the inherent impracticalities therein [9]. For this reason, a new Compendium Platform (CP), which is hosted at <http://www.freestatics.org> [10], was developed [9] and allows us to create very effective constructivist learning environments [11] which are based on reproducible computing (hosted at <http://www.wessa.net/> [12], [13] and based on the R language [14]) and with several advantages that relate to the monitoring of actual learning processes and quality control [15].

From a pedagogical point of view, the effectiveness of several competing paradigms are under critical investigation and debate. Arguably, the most prominent paradigm is that of individual and social constructivism which is described in [16], [17], and [18]. More recently, the application of constructivism within the context of computer-assisted science education [19], and the study of worked-out examples versus problem-solving tasks in web-based learning environments [20], has been the subject of thorough investigation. These studies are closely related to the newly developed CP because its underlying technology provides us with a novel opportunity to investigate the effectiveness of different learning styles based on actual (c.q. objectively measured) learning activities, rather than reported perception of learning experiences which can be shown to be highly misleading in unexpected ways [15].

This paper discusses the use of visual and explorative data mining tools that allow researchers and educators to discover

interesting relationships between reported experiences and actual learning activities. Based on an experimental statistics course with a large student population, it is illustrated that such tools enable us to highlight the relationships of interest quickly and without the need of sophisticated statistical techniques. This illustration leads to the (rather surprising) conclusion that learning styles need not be mutually exclusive (as is implicitly suggested in recent literature [20]).

II. COMPENDIUM PLATFORM

The R Framework (hosted at <http://www.wessa.net>) allows educators and scientists to develop new, tailor-made statistical software (based on the R language [14]) 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 [13]. The statistical software consists of a wide variety of so-called R modules which allow students to execute statistical computations through a web-based interface and distributed computing. This allows students to focus primarily on the interpretation of the analysis instead of the underlying technicalities.

Any computation that is generated with R modules can be communicated to other students/scientists through the use of a simple, transparent mechanism which allows one to permanently store the computation in a repository of computational objects that can be easily retrieved, recomputed, and reused. This repository was recently created within the OOF 2007/13 project of the K.U.Leuven Association and is called the Compendium Platform (CP). The main reason for creating the R Framework and the CP, is that it allows anyone to create and use Compendia of reproducible research. A Compendium is defined as [9]: any document with (open-access) references to (remotely) archived Computations (including Data, Metadata, and Software) that allow us to reproduce, and reuse the underlying analysis. Such documents can be easily created by students and permit any reader to (exactly) recompute the statistical results that are presented therein, or to reuse them with different parameters, data, or R code. Two examples of Compendia (that can be fully reproduced, reviewed, and reused) can be found in [21] and [22].

III. COURSE DESIGN

The main sections of the statistics course consist of a series of research-based workshops that require students to reflect

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

Bart Baesens: K.U.Leuven, Dept. of Business and Economics
This research is funded by the OOF2007/13 project.

and communicate about a variety of statistical problems, at various levels of difficulty. The workshops have been carefully designed and cannot be solved without additional information that is provided within the Virtual Learning Environment or by the tutor (during the lectures). This implies that students are required to reflect and think critically about the problems at hand. The pedagogical paradigm of (social) constructivism supports non-rote learning through activities that involve critical thinking, experimentation, and socially responsible collaboration [11].

Based on reported information from students and extrapolations based on web server log files, it can be estimated that each workshop involves about 9 hours of work per student, per week. In addition, students were required to perform detailed peer reviews of about 5-7 submissions from other students. The underlying idea was to create a learning environment that resembles the real world of empirical scientific research where peer review has been accepted as an intrinsic part of scientific endeavor and an invaluable tool for quality control. 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. The process of (anonymous) assessment by peers is of a "constructivist" nature because it is believed to help students in nurturing their scientific attitudes (through peer review experiences) and non-rote learning (through construction of knowledge).

A group of 240 undergraduate business students participated in the course and completed a total of 1907 workshops which were subjected to peer review. Every submission was assessed with respect to 3-6 criteria. For every graded criterion 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 (per workshop, per student). Fortunately, the otherwise time-consuming administration of the Peer Assessment procedure was automatically performed by the use of the Virtual Learning Environment called Moodle [23] which is freely available. One of the main reasons why Moodle features the administration of Peer Assessment is the fact that it has been designed with a constructivist, pedagogical philosophy in mind.

IV. DATA

The survey data that is used in this illustration was obtained from three well-known questionnaires (ATTLES, COLLES, and CSUQ). All questions are based on a 5-point Likert scale with high values corresponding to favorable ratings (1 = poor, 3 = neutral, 5 = excellent).

The first survey (called ATTLES) is available in Moodle (as a standard questionnaire) [23] and aims to measure student's attitudes towards thinking and learning [24]. The first ten questions relate to "connected" (empathic) ways

of learning whereas the ten last questions are associated with "separate" (critical, detached) ways of knowing: <http://www.freeststatistics.org/moodle/mod/survey/view.php?id=36>. The non-response rate was 8% and the survey was submitted during the first week of the semester.

Student perception of the online learning experience during the semester was measured with the Constructivist On-Line Learning Environment Survey (COLLES) as implemented in Moodle [23]. This survey can be interpreted in terms of student satisfaction about a variety of important aspects: relevance, reflection, interaction, educator, peers, and understanding (for a complete list see: <http://www.freeststatistics.org/moodle/mod/survey/view.php?id=37>). For every aspect there are eight questions, four of which are related to the actually perceived experience. The remaining four questions have identical phrases but are related to the degree of what students prefer. The survey was submitted at the end of the semester, before the scores of the multiple choice test were made available. The non-response rate was 15%.

The third survey is based on IBM's Computer System Usability Survey (called CSUQ) [25] with additional questions that were specifically related to the relationship between software usability and statistics learning. The extended version of this survey (CSUQ+) is described in [26]. The questions were made available within a "Quiz" module in Moodle at the end of the semester and can be examined at: <http://www.freeststatistics.org/moodle/mod/quiz/view.php?id=410>. The non-response rate was 17%.

The objectively measured data about actual student activities was obtained through the use of the CP [9]. For each student, a selected number of measures are used:

- Bcount: the number of statistical computations that was archived in the CP's repository.
- Buniques: the number of unique statistical computations that was archived in the CP's repository.
- nnzfg: the number of non-empty feedback communications that were submitted in the context of peer review.
- sfg: the total length of all submitted feedback communications about peer review (in number of characters).
- mflg: the median length of all submitted feedback communications about peer review (in number of characters).
- aflg: the arithmetic average length of all submitted feedback communications about peer review (in number of characters).
- iqrfng: the inter-quartile range of all submitted feedback communications about peer review (in number of characters).

The variables Bcount and Buniques represent learning activities that can be associated with "learning by problem-solving tasks" because they are directly related to the workshops that students had to complete and submit. High values of Bcount and Buniques correspond to an intensive use of the problem-oriented learning style. All the other variables can be associated with the "learning by example and experimentation" learning style which is quite similar to the approach of worked-out examples. In fact, the submission of peer reviews (including grades and feedback communications) requires the

XY correlation

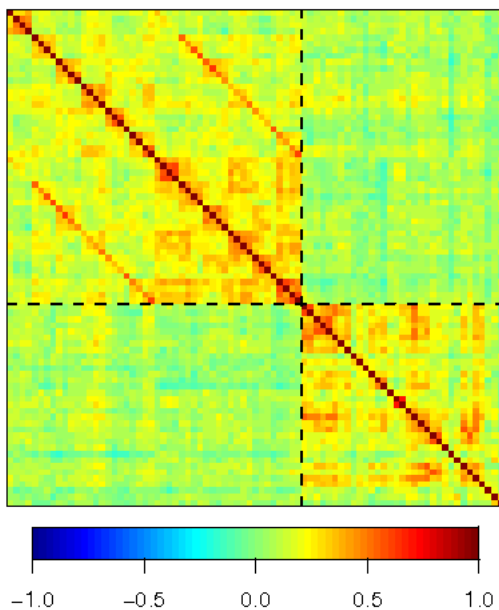


Fig. 1. Association between Satisfaction and Usability

students to thoroughly investigate the worked-out documents from peers and subject them to critical review.

V. ANALYSIS

The data mining approach that is employed in this investigation, is partially based on the CCA R package [27]. The original code was rewritten and expanded with several new graphical tools that explicitly take into account the fact that large numbers of correlations are computed. This expansion is important because it effectively prevents the investigator to find “significant” relationships between the variables of interest when they can be attributed to the type I error. In addition the new code uses Kendall’s tau rank correlations (instead of Pearson correlations) because a large part of the data under investigation is measured on an ordinal scale (c.q. the Likert scale in the surveys). Furthermore, Kendall’s tau is preferred over Pearson and Spearman correlations because they are more “conservative” and have - in general - more desirable properties [28], [29].

Figure 1 illustrates how all items that are related to student satisfaction and perceived usability can be displayed at once. The picture represents the values of Kendall’s tau rank correlations between all two-by-two combinations of the reported items that are contained in the questionnaires (COLLES and CSUQ+). The left top panel (TL) contains the COLLES items whereas the bottom right panel (BR) represents the CSUQ+ items. In the top right (TR) and bottom left (BL) one can see the rank correlations between satisfaction and usability items. The value of each correlation is represented by a color (the legend is displayed at the bottom of the picture).

The most remarkable features of the graphical analysis in Figure 1 can be summarized as follows:

- In panel TL there are several clusters of high correlation along the diagonal (running from the top left to the bottom right). This can be easily explained by the fact that the COLLES questions are grouped in logical aspects that relate to student satisfaction: relevance, reflection, interaction, educator, peers, and understanding.
- In panel TL there are two lines of high correlation which are parallel to the main diagonal. The reason for this is that the items have been arranged such that the actually perceived satisfaction (c.q. the uneven questions in the original COLLES) come first, followed by the preferred items (c.q. even questions in COLLES). Hence, the two parallel lines correspond to the fact that preferred and actual satisfaction (about each item) are highly correlated. Also, it is obvious that preferred items are more highly correlated (among each other) than the actually perceived ones.
- In panel BR it is clearly seen that there are several pockets of high correlation along horizontal and vertical lines. Again, this can be explained by the ordering of items in CSUQ+. Typically, items that correspond to the same (or similar) concepts are highly correlated.
- Panels TR and BL are mirror images and relate to the associations between reported student satisfaction and software usability. The correlations are all positive and few spots of medium-order correlation can be detected.

The analysis in Figure 1 is explorative and neutral with respect to the scientific discipline in which the investigation is situated. Both, psychologists and technology researchers can use this visual correlation matrix as a diagnostic tool to quickly identify correlational structure in the survey data. While this approach might seem appealing, there are some drawbacks that should be mentioned:

- There is no indication of significance of the correlations.
- If one is primarily concerned with the details of one aspect (either IT usability or psychological satisfaction) then there is no way to determine if the items of interest are related to the items of the other survey as a whole.
- Highly correlated items are not readily identifiable. The researcher needs to zoom in on the picture and count the rows/columns before the exact meaning of the items can be identified.

One could argue that this approach could be easily improved by displaying colors that correspond to the p-values of the rank correlations instead of the tau values. While this is a tempting solution, it introduces another - rather dangerous - statistical problem called “pretesting” which relates to the fact that one abuses significance tests by fishing for significance in a large number of computations. In other words, looking for meaningful information based on p-values inevitably leads to type I errors which should be avoided at all costs.

Therefore, it is interesting to look at two alternatives to the visual correlation matrix as proposed in the CCA package [27] which - at the same time - avoid the fishing for type I errors. Both alternatives are based on the concept that type I

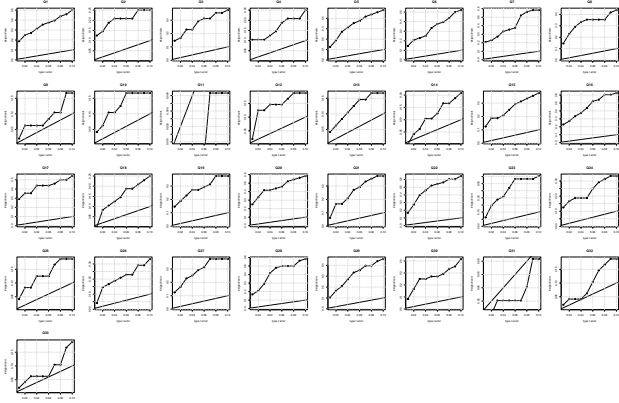


Fig. 2. Usability by Student Satisfaction

errors can only be assessed if all statistical hypothesis tests that have been computed are presented within a meta analysis (c.q. an overall analysis that consolidates all individual hypothesis tests).

Under the assumption that 100 correlations are tested (against a zero-null hypothesis) between independent variables, it can be expected that (on average) the number of false positives equals the prespecified type I error. For instance, if the type I error is 5% then it is expected that about 5 correlations (out of 100) are “significant” while there is no association between the variables. Therefore, it is interesting to count the number of significant correlations (for a series of type I errors) and relate that number to the total number of correlations computed.

In addition, it is also important to present the analysis in such a way that the researcher is able to easily interpret the results in terms of the discipline in which the research is situated. In other words, the psychologist is likely to be interested in the association between each of the COLLES items against all CSUQ+ items (as a whole). However, the technology scientist probably focuses on usability items and how they relate to customer (c.q. student) satisfaction as measured by COLLES. Instead of summing the scores of the survey for which the individual items are not of interest, it is better to compute all correlations and compute the percentual number of significant correlations between an item of interest versus all items of the other survey. In order to determine if an item is meaningful, one just has to compare this relative indicator with the corresponding type I error.

Figure 2 shows how the technology scientist looks at the correlational data of Figure 1. For each item in CSUQ+ a small chart is displayed which is numbered from 1 to 33 and displayed in sequential order (left to right, and top to bottom). For each item the theoretical type I error (ranging from 1% to 10% with an increase of 1% for each step) is displayed on the x-axis and the computed percentage of significant correlations between the CSUQ+ item and all COLLES items is shown on the y-axis. It can be clearly seen that the plotted lines (which represent the percentage of significant correlations) lie above

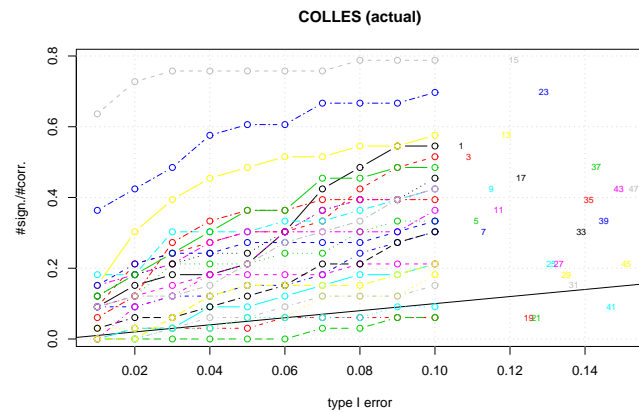


Fig. 3. Student Satisfaction by Usability

the diagonal (straight) line for each CSUQ+ item, except for items 11, and 31:

- item 11: “The information (such as online help, on-page messages, and other documentation) provided with this website is clear”
- item 31: “It is easy to move from one page to another”

All other usability items are clearly related to (overall) student satisfaction. This implies that there is an important relationship between psychological and technical student perceptions. In addition, it is clear that navigation and help files are not related to learning satisfaction.

The same data can be viewed through the eyes of the psychologist. A similar chart as in Figure 2 can be produced for all COLLES items (and their respective rank correlations with all CSUQ+ items). In this case, the psychologist might be primarily interested in the actually perceived satisfaction items: these are odd numbered questions in COLLES. Figure 3 displays the computed percentages (of significant correlations) which are displayed on the y-axis which is shared for all COLLES items (the lines are overlaid). In addition, the lines are displayed with a series of different line styles and colors. In order to be able to easily identify the psychological items that are strongly related to usability, it is necessary to display the item numbers on the right side of the chart. The height of each printed item number corresponds to the percentage of significant correlations for a type I error of 10%. In addition, the numbers are displayed with increasing coordinates on the x-axis such that overlaps are avoided.

Figure 3 shows that items 13, 15 and 23 have the highest percentage of significant rank correlations with usability:

- item 13: “I think critically about other students’ ideas.”
- item 15: “I think critically about ideas in the readings.”
- item 23: “other students respond to my ideas.”

This finding is of great importance because it indicates that the relationship between critical thinking/communication and software usability is very strongly supported by the empirical survey data. The pedagogical consequences may be fundamental: students who engage in critical (peer review-

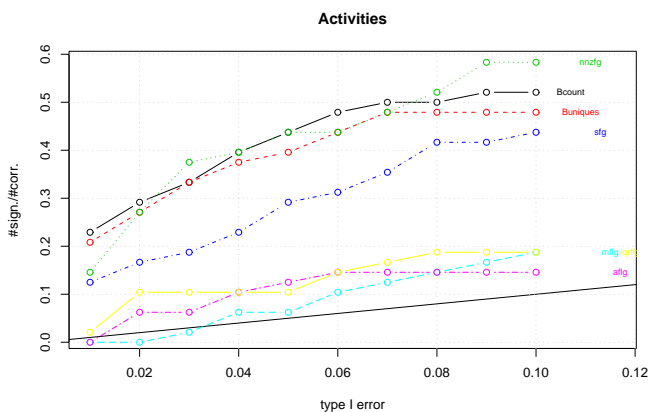


Fig. 4. Student Activities by Satisfaction

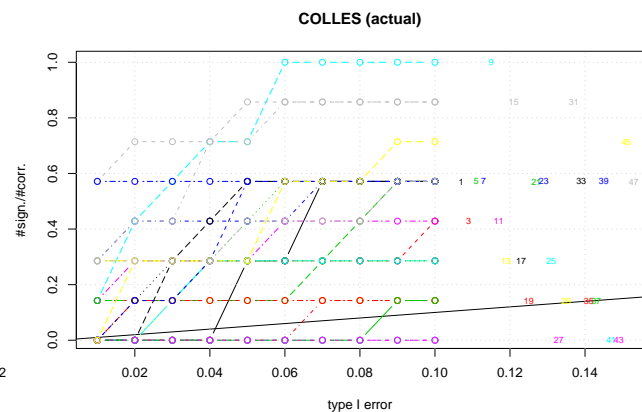


Fig. 5. Student Satisfaction by Activities

related) thinking and communication also report a high degree of satisfaction in terms of computer/software usability. Figure 3 clearly illustrates the fact that visualisation techniques can be used to quickly highlight important relationships that are of interest to the scientist without falling into the type I error trap.

The CP allows to measure actual student activities accurately and objectively. The relationships between these activity-based variables and student's psychological satisfaction are displayed in Figure 4. All variables seem to be correlated with student satisfaction - however, four variables emerge as the most important activities: nnzfg, Bcount, Buniques, and sfg. This result is surprising because two variables are related to worked-out examples and the other two correspond to problem-solving. It can be concluded that it is not necessary to make a choice of one or the other learning style - both can be used simultaneously and can be highly correlated with student satisfaction.

Figure 5 displays the psychologist's view of the relationships between student satisfaction and actual learning activities. The highest ranked COLLES items are the following:

- item 9: "I think critically about how I learn."
- item 15: "I think critically about ideas in the readings."
- item 31: "the tutor models critical self-reflection."
- item 45: "I make good sense of the tutor's messages."

This information is interesting because it shows that critical thinking is a key ingredient of student's learning experiences and actual activities. Based on Figures 3 and 5 the following can be concluded by the psychologist:

- a high-level computer system usability is strongly associated with critical thinking and the student's satisfaction of learning.
- the structure and design of actual, computer-assisted learning activities are very important in relation to critical thinking/communication and the student's satisfaction that is derived thereof.

VI. CONCLUSION

This paper demonstrates that it is possible to design simple - but effective - data mining tools for the purpose of the exploration of interesting relationships between reported experiences and actual learning activities. The illustrated computations also show that researchers are able to quickly discover the relationships that are of interest to their discipline, without the need to be concerned about the details of other data (not related to the discipline) or the type I error trap.

The results of particular interest are twofold. The first result is the fact that critical thinking and communication is strongly correlated with computer-assisted tasks and its associated usability perception. A second result is that statistics courses can involve both, worked-out and problem-oriented, learning styles simultaneously without loss of student satisfaction. The underlying CP technology supports different learning styles and at the same time it allows the educator to research the learning processes of students in ways that were not available before.

The future holds some exciting challenges for the team that is involved in the OOF2007/13 project. The most prominent one relates to an open software environment that allows educators to easily perform various types of data mining analysis about student's learning processes. This does not only involve the illustrations that are provided in this paper but extend to a wide range of model-based evaluation and research tools that relate the learning outcomes (in terms of objective exam scores) to reported experiences and objectively measured learning activities.

ACKNOWLEDGMENT

This research was funded by the OOF 2007/13 project of the K.U.Leuven Association. The author would like to thank Ed van Stee for his valuable suggestions and contributions.

REFERENCES

- [1] J. de Leeuw, "Reproducible research: the bottom line," in *Department of Statistics Papers, 2001031101*, Department of Statistics, UCLA., 2001.

- [2] R. D. Peng, F. Dominici, and S. L. Zeger, "Reproducible epidemiologic research," *American Journal of Epidemiology*, 2006.
- [3] M. Schwab, N. Karrenbach, and J. Claerbout, "Making scientific computations reproducible," *Computing in Science & Engineering*, vol. 2, no. 6, pp. 61–67, 2000.
- [4] P. J. Green, "Diversities of gifts, but the same spirit," *The Statistician*, pp. 423–438, 2003.
- [5] R. Gentleman, "Applying reproducible research in scientific discovery," *BioSilico*, 2005.
- [6] R. Koenker and A. Zeileis, "Reproducible econometric research (a critical review of the state of the art)," in *Research Report Series*, no. 60, Department of Statistics and Mathematics Wirtschaftsuniversität Wien, 2007.
- [7] D. L. Donoho and X. Huo, "Beamlab and reproducible research," *International Journal of Wavelets, Multiresolution and Information Processing*, 2004.
- [8] F. Leisch, "Sweave and beyond: Computations on text documents," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, (Vienna, Austria), 2003.
- [9] P. Wessa, "Learning statistics based on the compendium and reproducible computing," in *Proceedings of the World Congress on Engineering and Computer Science (International Conference on Education and Information Technology)*, (Berkeley, San Francisco, USA), UC Berkeley, San Francisco, USA, 2008.
- [10] P. Wessa and E. van Stee, *Statistical Computations Archive (online software at <http://www.freestatistics.org>)*. K.U.Leuven Association, Belgium, 2008.
- [11] P. Wessa, "How reproducible research leads to non-rote learning within a socially constructivist e-learning environment," in *Proceedings of the 7th European Conference on e-Learning*, (Cyprus), 2008.
- [12] P. Wessa, *Free Statistics Software (online software at <http://www.wessa.net>)*. Office for Research Development and Education, 1.1.23-r2 ed., 2008.
- [13] P. Wessa, "A framework for statistical software development, maintenance, and publishing within an open-access business model," *Computational Statistics*, 2008.
- [14] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
- [15] P. Wessa, "Measurement and control of statistics learning processes based on constructivist feedback and reproducible computing," in *Proceedings of the 3rd International Conference on Virtual Learning*, (Constanta, Romania), 2008.
- [16] E. Von Glasersfeld, "Learning as a constructive activity," in *Problems of Representation in the Teaching and Learning of Mathematics*, pp. 3–17, Hillsdale, NJ: Lawrence Erlbaum Associates, 1987.
- [17] E. Smith, "Social constructivism, individual constructivism and the role of computers in mathematics education," *Journal of mathematical behavior*, vol. 17, no. 4, 1999.
- [18] P. Eggen and D. Kauchak, *Educational Psychology: Windows on Classrooms*. Upper Saddle River, NJ: Prentice Hall, 5th ed. ed., 2001.
- [19] L. Moreno, C. Gonzalez, I. Castilla, E. Gonzalez, and J. Sigut, "Applying a constructivist and collaborative methodological approach in engineering education," *Computers & Education*, vol. 49, pp. 891–915, 2007.
- [20] M. Yaman, C. Nerdel, and H. Bayrhuber, "The effects of instructional support and learner interests when learning using computer simulations," *Computers & Education*, vol. 51, pp. 1784–1794, 2008.
- [21] P. Wessa, "A compendium of reproducible research about descriptive statistics and linear regression," tech. rep., K.U.Leuven Association, Belgium, 2008.
- [22] P. Wessa, "A compendium of reproducible research about time series analysis," tech. rep., K.U.Leuven Association, Belgium, 2008.
- [23] Moodle, "A free, open source course management system for online learning," in <http://www.moodle.org>, 2008.
- [24] K. M. Galotti, B. M. Clinchy, K. Ainsworth, B. Lavin, and A. F. Mansfield, "A new way of assessing ways of knowing: the attitudes towards thinking and learning survey (atls)," *Sex roles*, pp. 745–766, 1999.
- [25] J. R. Lewis, "Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," in *Technical Report 54.786*, IBM Corporation, 1993.
- [26] S. Poelmans, P. Wessa, K. Milis, E. Bloemen, and C. Doom, "Usability and acceptance of e-learning in statistics education, based on the compendium platform," in *Proceedings of the International Conference of Education, Research and Innovation*, International Association of Technology, Education and Development, 2008.
- [27] I. Gonzalez and S. Déjean, *CCA: Canonical Correlation Analysis*, 2008. R package version 1.1.
- [28] S. Arndt, C. Turvey, and N. Andreasen, "Correlating and predicting psychiatric symptom ratings: Spearman's r versus kendall's tau correlation," *Journal of Psychiatric Research*, vol. 33, pp. 97–104, 1999.
- [29] S. Arndt and V. Magnotta, "Generating random series with known values of kendall's tau," *Computer Methods and Programs in Biomedicine*, vol. 65, pp. 17–23, 2001.

Patrick Wessa Patrick Wessa leads the Compendium Platform project at the K.U.Leuven Association, Belgium. His research interests are primarily focused on the intersection between Information Technology, Statistics, and Education. His web applications are increasingly popular among academics and cited in many scholarly articles.

Bart Baesens Bart Baesens is a professor at the Department of Decision Sciences and Information Management, Faculty of Business and Economics, K.U.Leuven (Belgium). He has done extensive research on predictive analytics, data mining, classification and credit risk management. His findings have been published in well-known international journals and presented at international top conferences.