

# Computer-Assisted Assessment in Open-Ended Activities through the Analysis of Traces: A Proof of Concept in Statistics with R Commander

Miquel Calvo <sup>1\*</sup>, Artur Carnicer <sup>2</sup>, Jordi Cuadros <sup>3</sup>, Francesc Martori <sup>3</sup>, Antonio Miñarro <sup>1</sup>,  
Vanessa Serrano <sup>3</sup>

<sup>1</sup> Section of Statistics, Faculty of Biology, Universitat de Barcelona, SPAIN

<sup>2</sup> Department of Applied Physics, Universitat de Barcelona, SPAIN

<sup>3</sup> Quantitative Methods Department, IQS Universitat Ramon Llull, Barcelona, SPAIN

Received 31 July 2018 • Revised 14 December 2018 • Accepted 7 February 2019

## ABSTRACT

Open-ended tasks are common in Science, Technology, Engineering and Mathematics (STEM) education. However, as far as we know, no tools have been developed to assist in the assessment of the solution process of open-ended questions. In this paper, we propose the use of analysis of traces as a tool to address this need. To illustrate this approach, we developed a modified version of R Commander that collects traces of students' actions and described a way to analyze them by using regular expressions. We used this tool in an undergraduate introductory statistics course. The traces were analyzed by comparing them to predefined problem-solving steps, arranged by the instructor. The analyses provide information about the time students spent on the activity, their work intensity and the choices they made when solving open-ended questions. This automated assessment tool provides grades highly correlated to those obtained by a traditional test and traditional grading scheme.

**Keywords:** computer-assisted assessment, learning analytics, open-ended activity, R commander, statistics education

## INTRODUCTION

Problem-solving, or, more generally, working in open-ended tasks, is commonly thought to be a fundamental activity in learning STEM, either as an educational goal in itself or as a way to develop scientific and mathematical skills (Bahar & Maker, 2015; Cai & Lester, 2010; English & Sriraman, 2010; Hardin, 2003). However, instructors face two main difficulties: assessment is time-consuming, and the steps followed by the students to obtain a solution is hardly available even when detailed reports are requested.

A common option to reduce assessment time is using objective tests in Virtual Learning Environments (VLEs) and Online Homework Systems (OHSs). Their data are used in the Educational Data Mining (EDM) community to provide feedback to instructors and predict student performance (Romero & Ventura, 2010). However, some authors (Conijn, Snijders, Kleingeld, & Matzat, 2017) argue that this information is of limited value to predict student performance unless relevant assessment data are included. Unfortunately, neither VLEs nor OHSs provide tools to assess the solution process of open-ended tasks.

To obtain relevant content-related assessment data, Intelligent Tutoring System (ITS) applications are designed to capture user interactions by tracing or logging user data. This provides a powerful source of information for both psychological and educational research, as well as for generating useful feedback for users and instructors. Thus, the analysis of information stored in log files facilitates a large variety of studies about the student learning process – see for instance Romero and Ventura (2010) and references therein –. Throughout the past decade, a large

### Contribution of this paper to the literature

- While commonly used Virtual Learning Environments (VLEs) offer tools for assessment, they do not allow for assessing the solution of open-ended activities.
- A modified version of R-Commander, a well-known tool for teaching introductory statistics, was developed to enable collecting traces of student actions when solving open-ended activities.
- These traces were analyzed to assess the work of the students by automatically comparing the collected traces to predefined solution paths. This automatic assessment provides grades comparable to those obtained by a traditional test and traditional grading scheme.

number of papers addressed educational applications of information stored in logged files. In this paper we refer to this data as *traces*. Some of the applications mentioned above may be highlighted: (i) classification of students according to a specific criterion (Stevens, Beal, & Sprang, 2009), (ii) identification of plagiarism (Rosales et al., 2008), (iii) adaptive behavior detection – when a student learns how to manipulate the application in order to get higher marks – (Baker, Corbett, & Koedinger, 2004) or (iv) disengagement detection, lack of learner motivation, (Cocea & Weibelzahl, 2009).

Traces are analyzed using the information produced by the application (Romero, Ventura, & Garcia, 2008). Since the information contained in such files and their internal structure influences how the data are accessed and retrieved, some research teams have produced specifications for generating data files containing information that is relevant to the educational task (Mostow & Beck, 2009; Song & Ma, 2008; VanLehn et al., 2009). The Pittsburgh Science of Learning Center (PSLC) has designed a file format adapted to the problems they study (Ritter & Blessing, 1998; Ritter & Koedinger, 1996). The PSLC data format is well documented in a comprehensive manual which entitles other groups to generate PSLC-compatible data from their own applications (PSLC, 2013). However, their model is primarily intended to capture interactions in educational tutoring applications, where tasks or skills are assessed multiple times. This orientation is less suitable when dealing with rich user interfaces and problem-solving tasks where there is not just one solution path and skills cannot be defined in advance, as is the case in solving open-ended statistical problems.

Efforts are currently being made to face the problem of time-consuming assessment in order to achieve automatic grading in specific open-ended scenarios. For example, De Marsico, Sciarrone, Sterbini, and Temperini (2017) use a Bayesian network to predict students' performance using teacher and peer-evaluation in open-ended questions, and Kinnebrew, Segedy and Biswas (2017) interpret students' open-ended learning and problem-solving behaviors in Betty's Brain environment by using Hidden Markov Models and Differential Sequence Mining. Both approaches are difficult to generalize and provide results that are hard to interpret by an instructor leading a class.

In the field of statistics, multiple online resources for teaching and learning have been devised. For instance, under the section 'Statistics and Probability' within the Multimedia Educational Resource for Learning and Online Teaching website (MERLOT, 2018), over a thousand online teaching resources appear that cover almost all the topics in higher education statistics courses. Moreover, numerous experiences in online statistics learning are reported in the literature (Anderson-Cook & Dorai-Raj, 2003; Basturk, 2005; Carnegie Mellon Open Learning Initiative, 2017; Dinov, Sanchez, & Christou, 2008; Gonzalez & Muñoz, 2006; González, Jover, Cobo & Muñoz, 2010; Lane & Scott, 2000; Larreamendy-Joerns, Leinhardt, & Corredor, 2005; Suanpang, Petocz, & Kalceff, 2004; Symanzik & Vukasinovic, 2003; West & Ogden, 1998). We are not aware that any of them allow for an automated assessment of open-ended tasks using logs.

Nowadays, one of the most used tools in statistics and data analysis is R (Muenchen, 2017; Piatetsky, 2017), an open source programming language and software environment for statistical computing (R Core Team, 2017). To facilitate its use in introductory statistics courses, Fox (2005) developed R Commander, a well-known Graphical User Interface (GUI) for R. It allows the use of R without compromising the learning process, as an R command line interface could represent an obstacle to many students.

Despite commonly used VLEs offer tools for assessment, they do not allow for assessing the solution of open-ended activities. This paper presents the development of a modified version of R Commander able to trace students' actions and how they can be analyzed to evaluate an educational activity and assess student work even for open-ended activities. We also show that assessment based on traces can provide a grade for student work which correlates with more traditional assessment methods. In the following sections, we detail the methodology of this study and highlight the results and conclusions obtained in the first implementation of this assessment model.

## METHODOLOGY

The methodological design of this research consists of three main aspects: (1) modifying R Commander to record traces of users' actions, (2) analyzing log files, and (3) creating and running a pilot activity in a Biostatistics course.

### Modified version of R Commander

As mentioned earlier, R Commander (Fox, 2005) is an open-source GUI for R. As such, its code can be reused and modified to fit different needs. In our case, we modified R Commander to obtain an activity log of the students' work. This modified package can be downloaded from <http://asistembe2.iqs.edu/rcmdrtr/>. A portable version including R and our modified version of R Commander is also available there.

When the user interacts with this modified version of R Commander, either by clicking on a menu or by entering the instructions in the R Script panel, the system records both the user command and the results shown in response. The system also records any editing of the dataset in use. Each record includes a timestamp, an anonymized random code for the user and a session identification.

This user activity is logged as a sequence of XML entities, that are saved locally in the working directory.

### Analysis of the Log Files

The information collected in these log files allows analyzing many aspects of users' work, e.g. time spent on the activity, amount of work done, work intensity, solving strategies and paths taken by the user, and correctness of a specific resolution. We detail below the methods we used to estimate some of the aspects of the log files. We developed a library of R functions, which automates the log analysis. These functions can be downloaded as a *Shiny* application (Chang, Cheng, Allaire, Xie, & McPherson, 2018) from <http://asistembe2.iqs.edu/rcmdrtr/>.

**Time spent on the activity** is obtained by adding up the time spans between consecutive user actions. If the time span is longer than a predefined threshold, this lapse can be excluded from the working time for the user.

**Amount of work** can be understood as the total number of actions recorded in the log files for a specific user.

We visualize **work intensity** by plotting the actions of each user in their time-on-task timeline, i.e. cumulative time spent by the user up to the current action. Closer actions denote more intensity.

In order to analyze the **strategies** followed by a student and the **correctness** of those strategies, the instructor may design some milestones, which can be understood as steps in the solution process. They identify important achievements or potential errors when resolving the activity. When a milestone may be accessed in different ways, we call it an *assessment milestone* and each possible way is an *observation item*. Assessment milestones are defined as logical rules based on the observation items built to assess students' work. A trace-based grade for each user can be obtained by considering the number of successfully reached assessment milestones.

### Pilot Activity in Teaching Biostatistics

To evaluate the feasibility and usefulness of this approach, a pilot study was run in a Biostatistics course for the degrees in Biomedical Sciences, Biotechnology and Biochemistry at a large public research university in Spain.

In this class, students were asked to carry out several activities online. Some of them adopted the format of a practical case, presented to the students at the beginning of the activity, followed by a series of questions and tasks related to the experiment. Those activities were developed in R with package *exams* (Grün & Zeileis, 2009; Zeileis, Umlauf, & Leisch, 2014), which facilitates the creation and maintenance of large-scale exams in statistics courses in PDF, HTML or e-learning questionnaires in Moodle or other VLEs. Those activities were randomly generated and automatically assessed in Moodle. We used one of those activities to test the traced version of R Commander.

The specific activity we used to test the system of collecting students' traces, focused on probability distributions. The practical case study was about a factory spreading several pollutants in the atmosphere (**Figure 1**). We were able to measure the concentration of these air pollutants and checked whether the law was violated. Students had to work with probability density functions, distribution functions and several topics related to probability. The types of questions students had to answer during the activity included questions that expect a numerical answer and also multiple choice questions.

**CASE 1: CO emissions**

As stated before, the emission of carbon monoxide is continually measured throughout the day. There is an alarm system that is triggered if the emission exceeds a certain value (60,000  $\mu\text{g}/\text{m}^3$ ). The alarm causes the production process to slow down until emissions are again acceptable. This process is very fast and never exceeds ten minutes.

Quality controls done last two years conclude that when the production process works correctly, the variable  $N$  = "Number of controls per day where CO emission exceed the acceptable value" follows a Poisson distribution with mean 1.5.

What is the probability that in a day you do not have to slow down the production for excess of CO?

1. Probability that in one day the emission of CO never exceeds the permitted level

The production plan of the factory allows this self-regulation of CO emissions if the number of times in a day when production ( $N$ ) must be slowed down is less than 4. When  $N$  is equal to or greater than 4 times in a day, it is recommended to completely stop production and investigate and solve the problem.

When the production is working correctly, what is the probability that the number of times that CO emissions exceed the allowed value in one day is less than 4?

2. Probability of exceeding CO value allowed less than 4 times

When analyzing the data of last week, the technician in charge finds that the data of Monday were lost. However, it is certain that at some point in that day production was slowed down at least once by excess of CO emissions. It is also sure that the 4 times were not reached because no investigation or complete control of production control was carried out.

By not having more information, the value 3 was written in the incident record book, that is, declaring 3 alarms due to excess emissions of CO. With all these informations, compute the probability that the value 3 is correct.

3. Probability that the value 3 is correct?

**Figure 1.** Screenshot of the activity. Students are required to calculate some probabilities. The students use R-Commander as a statistical calculator and their results are introduced in their corresponding numerical fields

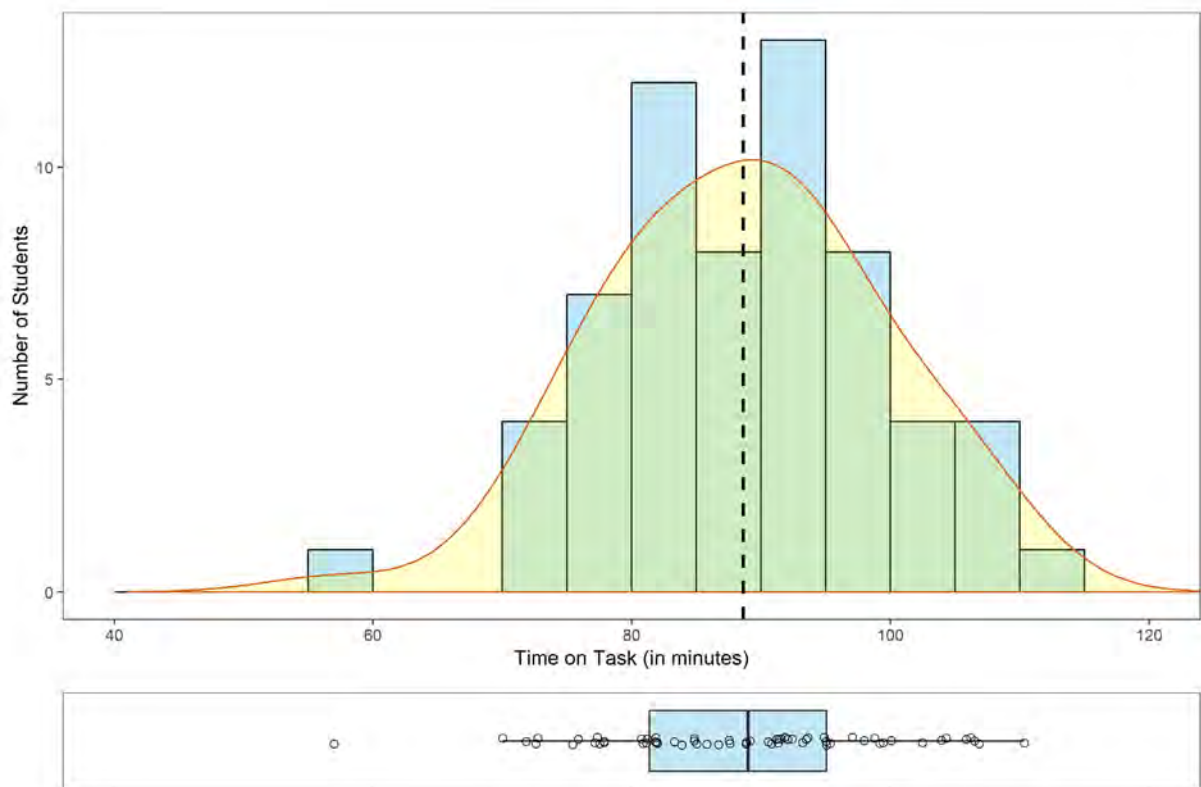
The activity was divided into three parts. In the first section, which consisted of six questions related to the concentration of sulphur dioxide ( $\text{SO}_2$ ), students worked with normal and binomial distributions. The second section, including eight more questions, continued with normal and binomial distributions but also introduced the Poisson distribution and some operations on random variables. The third part, which contained nine additional questions, mainly focused on Poisson random variables.

In order to assess the students' work from trace files, the expected solution path can be modeled as a set of milestones to be reached while each student follows the computer-based learning activity. **Table 1** shows some milestones students have to reach. For instance, in section 2, we are concerned with the concentration of nitrogen dioxide ( $\text{NO}_2$ ). The text states that nitrogen dioxide ( $\text{NO}_2$ ) concentrations follow a normal distribution with a given mean and standard deviation. After computing a 96% confidence interval and taking into account that we take nine measurements a day, i.e. 63 measurements a week, question 5 asks the following: "What is the exact probability (calculating the exact distribution, not the approximate distribution) that in the span of one week, two or more measurements fall outside the 96% confidence interval determined at the beginning?" Milestones O14 and O15 are two possible calculations to obtain the desired answer. As can be seen in **Table 1**, milestones are mainly composed of R functions (such as `pbinom`, `ppois`, and `pnorm`) and some XML tags.

In the spring of 2015, approximately 75 students (of the three degrees cited above) used this activity. We collected valid traces from the work of 62 of them. The activity took place in one session and the total time allotted was two hours. The instructor led the first part of the activity and the students were in charge of parts two and three.

**Table 1.** Milestones students have to reach to correctly answer certain questions of the activity. The milestones in the same bold line box are alternative ways to obtain the answer. Milestones O21A and O21B form a unity and both are necessary to reach the solution. The numbers in red are variable parameters that can differ from one student to another

MILESTONE	TRACED COMMAND IN RCMDR SYNTAX	QUESTION NUMBER IN SECTION	SECTION
O14	<PARAM name='Command' value='pbinom(c(2),%20size=63,%20prob=0.02,%20lower.tail=FALSE)' />	5	2
O15	<PARAM name='Command' value='pbinom(c(2),%20size=63,%20prob=0.02,%20lower.tail=TRUE)' />	5	2
O20	<PARAM name='Command' value='pnorm(c(150,120),%20mean=132.3,%20sd=20.412,%20lower.tail=TRUE)' />	8	2
O21A	<PARAM name='Command' value='pnorm(c(150),%20mean=132.3,%20sd=20.412,%20lower.tail=TRUE)' />	8	2
O21B	<PARAM name='Command' value='pnorm(c(120),%20mean=132.3,%20sd=20.412,%20lower.tail=TRUE)' />	8	2
O22	<PARAM name='Command' value='local(%7b%0a%20%20.Table%20%3c-%20data.frame(Probability=dpois(0:8,%20lambda=2.1))%20%0a%20%20print(.Table)%0a%7d)' />	1	3
O23	<PARAM name='Command' value='ppois(c(0),%20lambda=2.1,%20lower.tail=TRUE)' />	1	3
O24	<PARAM name='Command' value='ppois(c(0),%20lambda=2.1,%20lower.tail=FALSE)' />	1	3

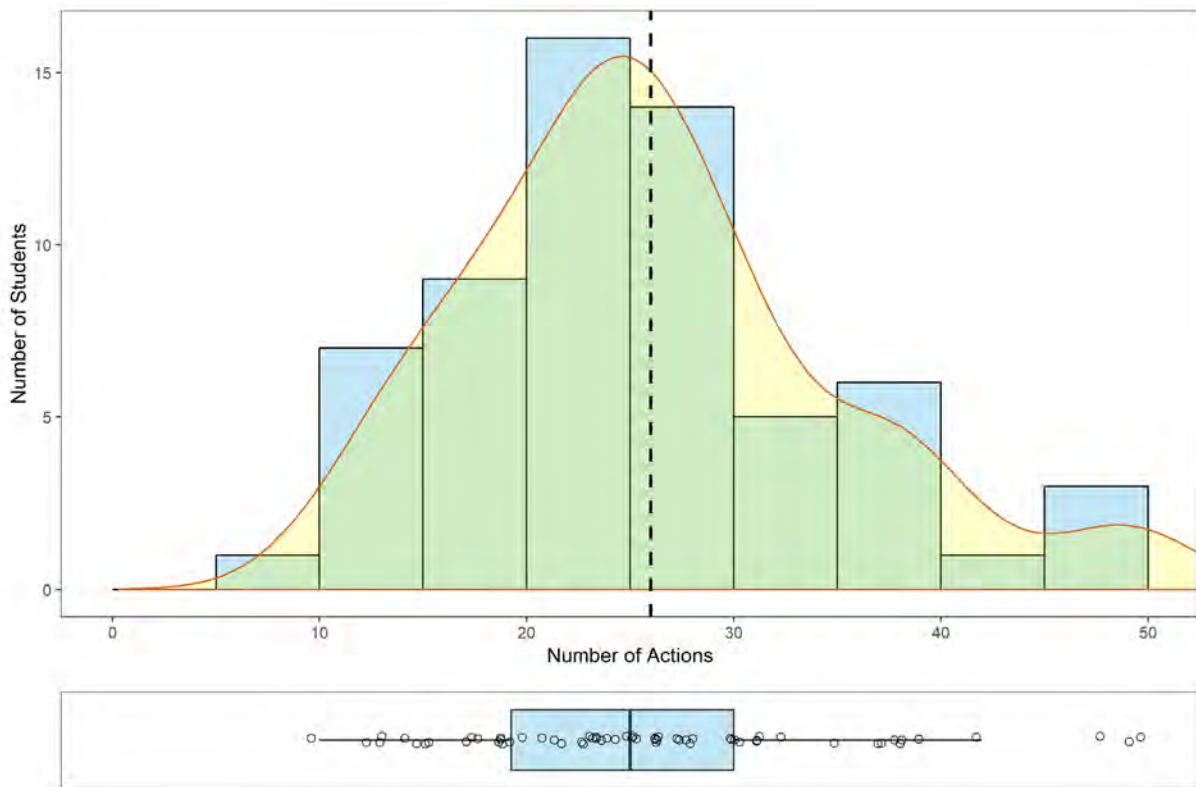


**Figure 2.** Time on task - no idle time discounted. The figure includes a histogram, a density curve, a boxplot, and the raw data plotted using a scatter plot with some jittering

## RESULTS

Traces can be analyzed at different levels. The discussion that follows summarizes the observations at group level (including time and number of actions) and the analysis of the students' work from a set of milestones (for both observation item and assessment milestone) defined for the activity described above.

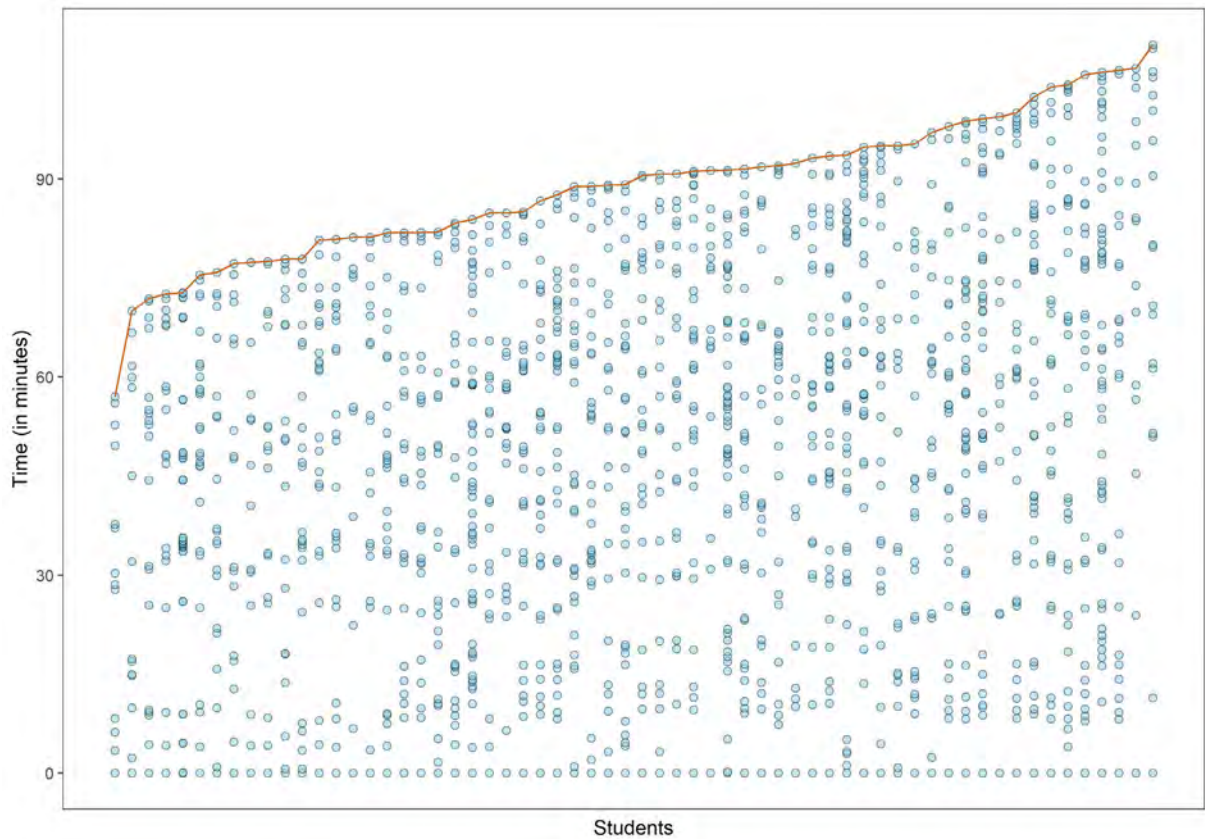
At group level, the distribution of the students' time solving the activity (time on task) is shown in **Figure 2**. This distribution is almost symmetrical, and its mean is approximately 90 minutes. Most students were actively involved in the activity. Only one of them spent less than 60 minutes solving the problem.



**Figure 3.** Number of actions. The figure includes a histogram, a density curve, a boxplot, and the raw data plotted using a scatter plot with some jittering

The number of actions performed by the students ranges from 10 (lines in the log file) to a maximum of approximately 50 (Figure 3). The distribution seems to be unimodal according to a Gaussian distribution, as one would expect. When looking at the work intensity, no signs of tiredness or quitting are found, as the number of actions is homogenous throughout the solution process (Figure 4). However, some blank gaps, are observed. These intervals could indicate the beginning of a new section and the associated time needed to read its heading and to go through the theoretical questions included in the activity. Figure 4 allows identifying students with uncommon action densities: for instance, the two last subjects in the plot display a lower action density compared to the whole group.





**Figure 4.** Actions over time (each column = 1 student, each dot = 1 action). The columns are ordered by the amount of time needed to complete the activity, from lowest to highest

A second level of analysis considers the observation items. For instance, O15 is less frequent than O14 when both lead to the same answer as can be seen in [Figure 5](#).

This Figure depicts the relation between assessment milestones and observation items. The bars represent the assessment milestones with their corresponding success rate, while the circles situated above represent the corresponding alternative observation items in the assessment milestone. It is a direct visualization of how students answer each question. It is important to note that the sum of the percentages of the observation items achieved can be greater than the corresponding figure of the assessment milestone, since students can try different or redundant ways to reach the assessment milestone and the software is able to trace those actions. Milestones with a low frequency are located at the end of sections 2 and 3 where the difficulty is higher. For instance, O10 and O22 are clearly the most often used ways to solve specific questions of the activity, identified as the assessment milestones A04 and A09, despite the students were not led by the instructor in this part of the activity.

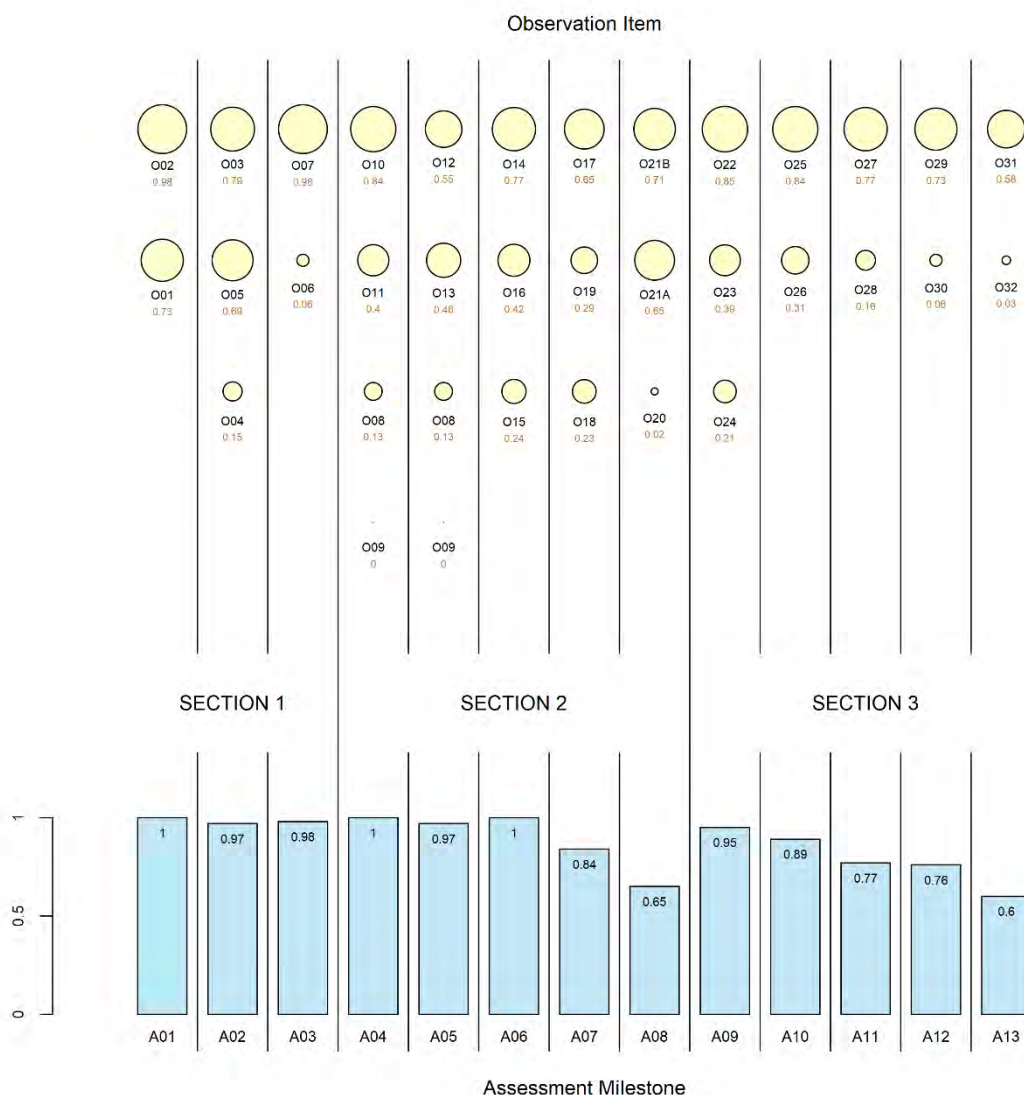
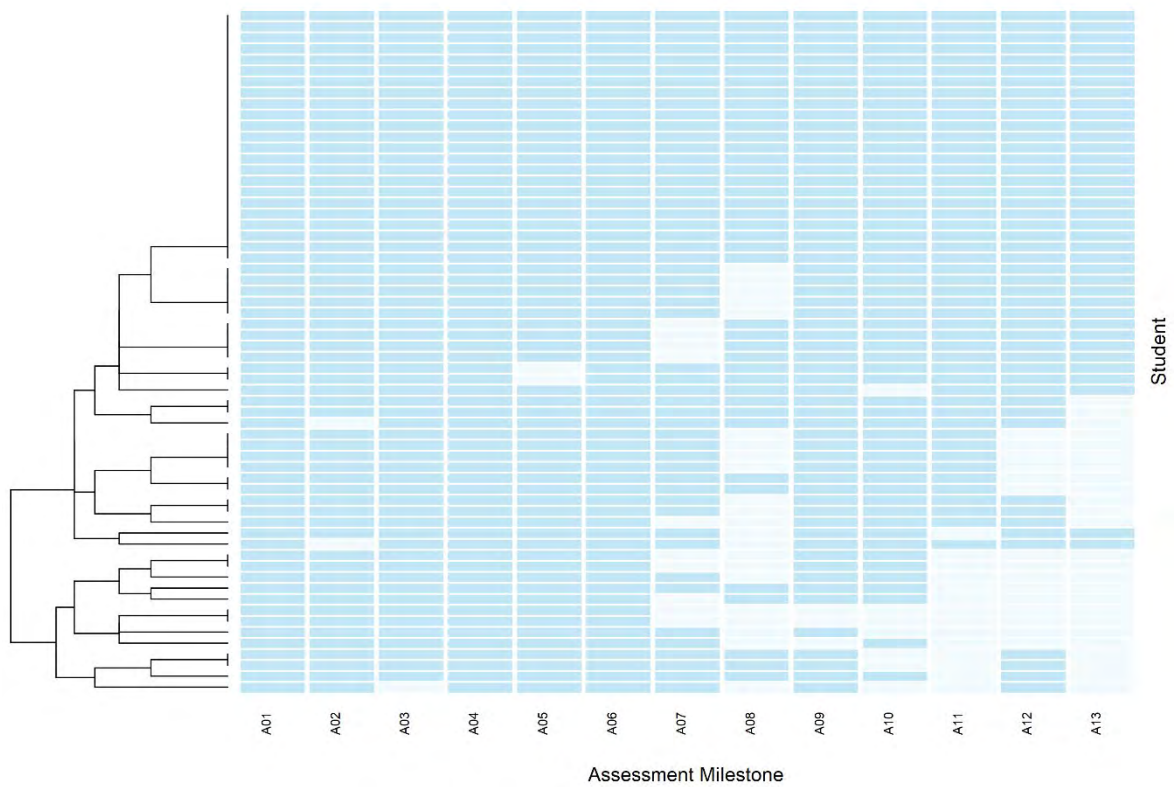


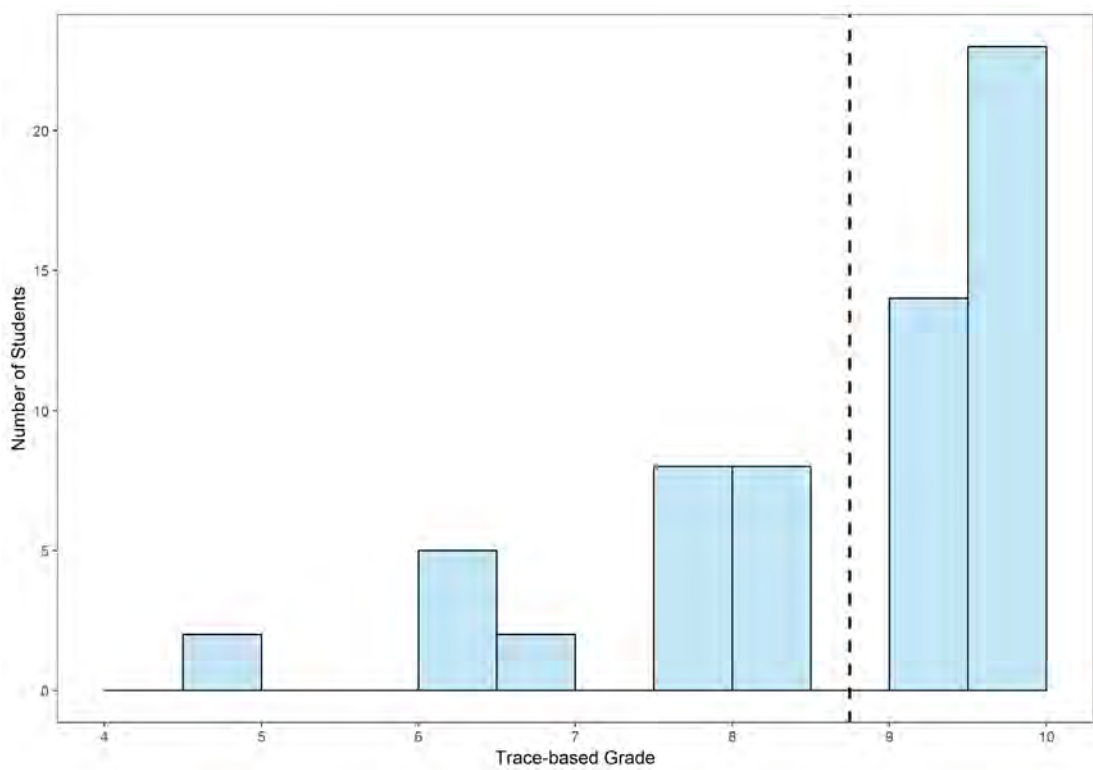
Figure 5. Relation between achievement in observation items and assessment milestones

Figure 6 is a heatmap of the assessment milestones showing clusters of students. This figure highlights the grading process: the grade of a student is proportional to the number of assessment milestones achieved. Additionally, this chart can allow a quick identification of patterns of student’s resolution processes and may provide insights into the learning difficulties of individuals or groups. Figure 7 shows the grade distribution according to this rule.

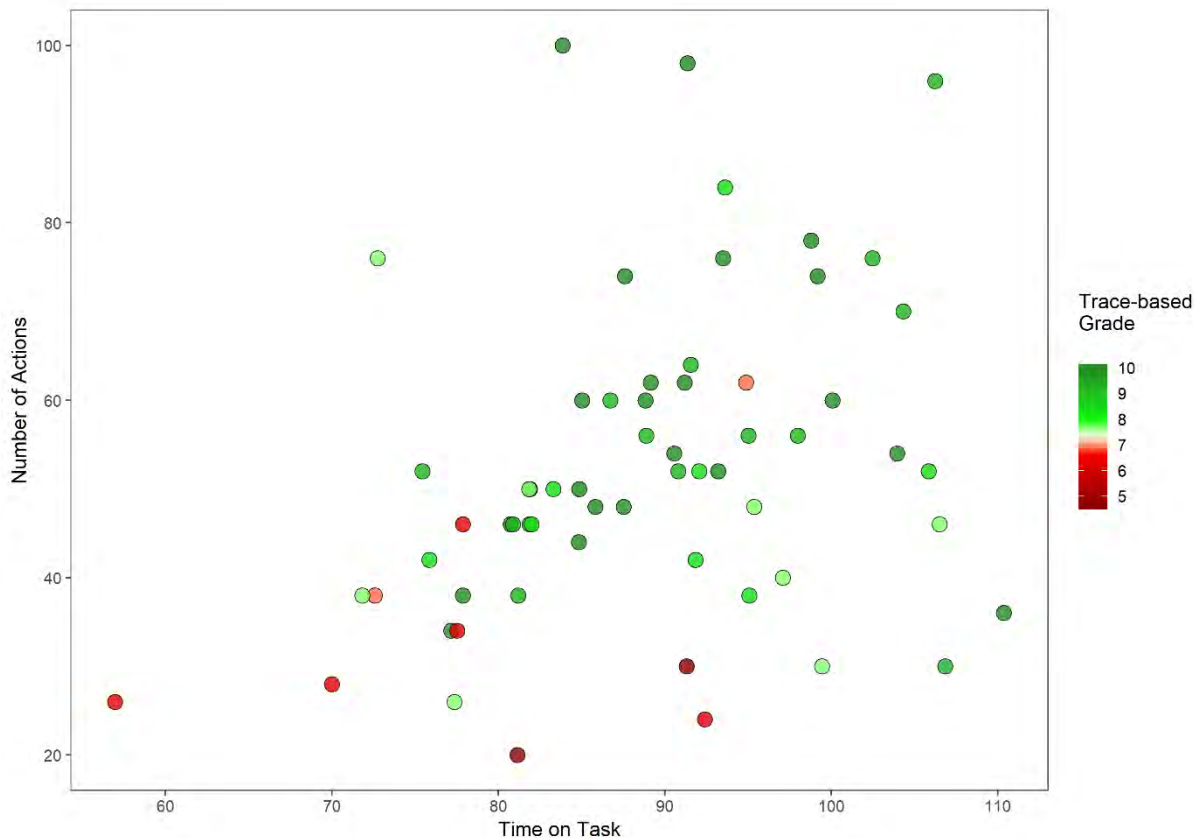




**Figure 6.** Heatmap of assessment milestones (each row = 1 student, each column = 1 milestone)



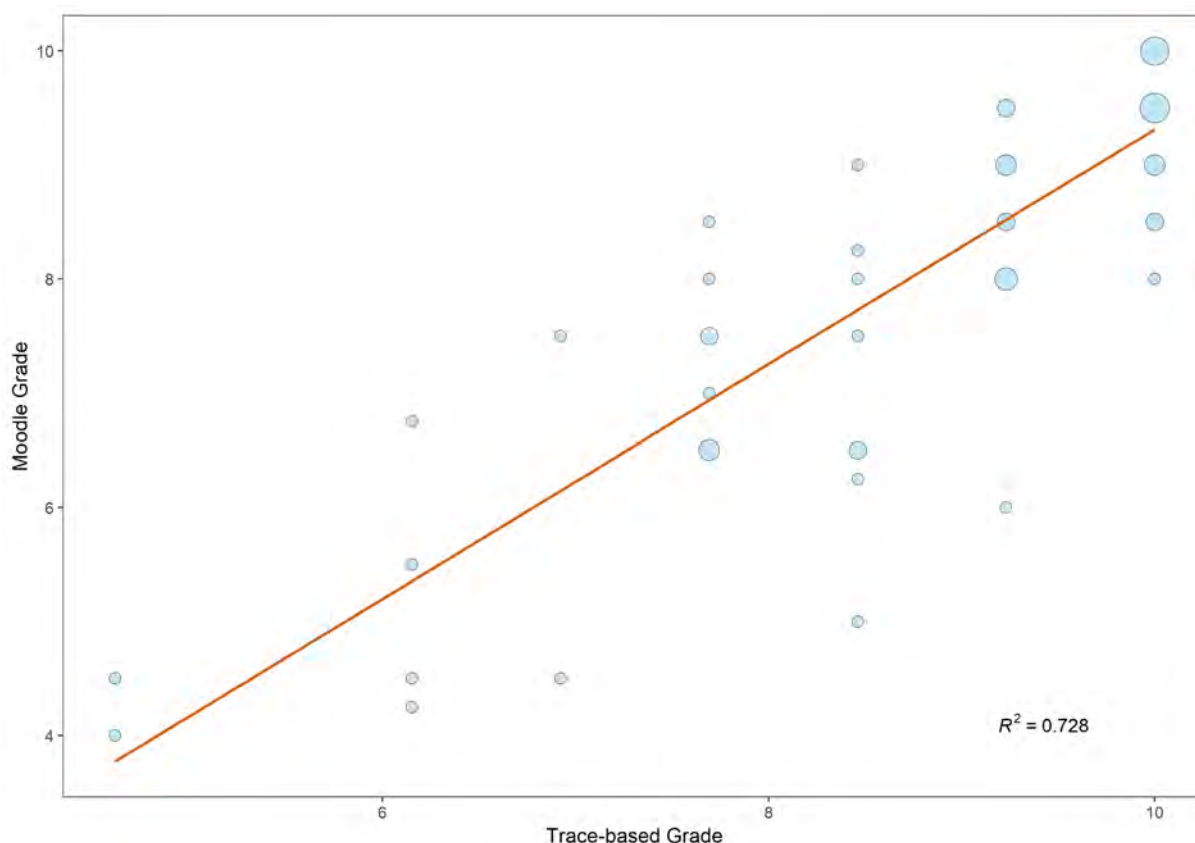
**Figure 7.** Trace-based grade distribution. The grades are computed as the proportion of achieved assessment milestones, scaled from 0 to 10. 10 means all the assessment milestones were achieved. Grades are grouped in half-point width bins, upper bound is included. The dashed line indicates the mean of the grades



**Figure 8.** Actions-Grade-Time Scatterplot (trace-based). The number of actions recorded into the log file are presented in front of time spent solving the activity. A diverging color scale is used to encode the trace-based grade, computed from the proportion of achieved milestones and scaled from 0 to 10. Red colors indicate grades in the lower range and green colors mean grades in the higher range. A neutral color is used for the mid-range values

Figure 8 can be interpreted as a student efficiency plot, i.e., students with a high grade (green) combined with a low number of actions in a short period of time are the most efficient ones.

Figure 9 shows a scatter plot and the regression line between the grades obtained in the Moodle questionnaire and the trace-based grades. Interestingly, the correlation coefficient (0.85) suggests that trace-based assessment and the traditional scheme implemented in Moodle provide similar results. It is worth mentioning that the Moodle forms include some theoretical questions that cannot be picked up by traces.



**Figure 9.** Relation between the Moodle grades, obtained from the questionnaire in Moodle, and the trace-based grades, computed from the assessment milestones. The size of the marker is proportional to the number of students

## CONCLUSIONS

We developed a modified version of R Commander, allowing traces of students' actions to be collected. Its use is transparent for the users and leaves a file behind that can be submitted as a deliverable. This modified R Commander version can be downloaded from <http://asistembe2.iqs.edu/rcmdrtr/>.

We used this tool in an undergraduate introductory statistics course aimed for different student profiles. We designed a methodology to analyze the collected data and to estimate a grade from them. The traces were analyzed by comparing them to predefined solution steps. These analyses informed the instructors about the time the students spent, their work intensity and the choices the students made when solving the activity. Students with uncommon action densities could also be identified.

In the specific setting used in this study, the automated assessment presented here provided grades comparable to those obtained by a traditional test and a traditional grading scheme. The correlation coefficient between the trace-based grades and those obtained through the Moodle questionnaire reached 0.85.

The different analyses described in this work can also be valuable to instructors (in this case, in statistics) by informing them of learning gaps, by identifying students with specific learning needs, or by providing insights to give tailored feedback to individual students. The authors also foresee this approach to be useful to scholars involved in discipline-based education research.

Some of the limitations of this study are: (1) the assessment milestones and the observations items are inherent to each activity and must be specifically defined; (2) consequently, some training may be required for those willing to create new observation items and/or assessment milestones; (3) the huge amount of information provided must be carefully analyzed to obtain insightful conclusions. In this sense, further work needs to be done to explore the possibility of extending this approach. We expect this assessment methodology to open the door to widespread automated grading of open-ended problems.

## ACKNOWLEDGEMENTS

This project has been supported with funds from the REDICE program at Universitat de Barcelona (REDICE12 1743-01, REDICE-14-508) and from the CEI Aristos Campus Mundus 2015 (project ACM2016\_26).

## Authorship Statement

Authors are listed in alphabetical order. M. Calvo, A. Carnicer, J. Cuadros, and A. Miñarro designed the study and developed the software; M. Calvo and A. Miñarro conducted the experiment, M. Calvo, J. Cuadros, F. Martori, A. Miñarro, and V. Serrano contributed to analysis and interpretation of data. All authors contributed to drafting the manuscript and reviewed it.

## REFERENCES

- Anderson-Cook, C., & Dorai-Raj, S. (2003). Making the concepts of power and sample size relevant and accessible to students in introductory statistics courses using applets. *Journal of Statistics Education*, 10. <https://doi.org/10.1080/10691898.2003.11910721>
- Bahar, A., & Maker, C. (2015). Cognitive Backgrounds of Problem Solving: A Comparison of Open-ended vs. Closed Mathematics Problems. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(6). <https://doi.org/10.12973/eurasia.2015.1410a>
- Baker, R., Corbett, A., & Koedinger, K. (2004). Detecting student misuse of intelligent tutoring systems. In: *Intelligent tutoring systems*, Heidelberg: Springer, pp. 54–76. [https://doi.org/10.1007/978-3-540-30139-4\\_50](https://doi.org/10.1007/978-3-540-30139-4_50)
- Basturk, R. (2005). The effectiveness of computer-assisted instruction in teaching introductory statistics. *Educational Technology & Society*, 8, 170–178.
- Cai, J., & Lester, F. (2010). Why is teaching with problem solving important to student learning. *Problem Solving: Research Brief*.
- Carnegie Mellon Open Learning Initiative (2017). *Statistics course*. Retrieved on 24 January 2018 from <http://oli.cmu.edu/courses/all-oli-courses/statistics-course-details/>
- Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. (2018). shiny: Web Application Framework for R. R package version 1.2.0. Retrieved from <https://CRAN.R-project.org/package=shiny>
- Cocea, M., & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-learning environments. *User Modeling and User-Adapted Interaction*, 19, 341–385. <https://doi.org/10.1007/s11257-009-9065-5>
- Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17–29. <https://doi.org/10.1109/TLT.2016.2616312>
- De Marsico, M., Sciarrone, F., Sterbini, A., & Temperini, M. (2017). Supporting Mediated Peer-Evaluation to Grade Answers to Open-Ended Questions. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(4), 1085–1106. <https://doi.org/10.12973/eurasia.2017.00660a>
- Dinov, I., Sanchez, J., & Christou, N. (2008). Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. *Computers & Education*, 50, 284–300. <https://doi.org/10.1016/j.compedu.2006.06.003>
- English, L., & Sriraman, B. (2010). Problem solving for the 21st century. In *Theories of mathematics education* (pp. 263–290). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-00742-2\\_27](https://doi.org/10.1007/978-3-642-00742-2_27)
- Fox, J. (2005). The R Commander: a basic statistics graphical user interface to R. *Journal of Statistical Software*, 14(9). <https://doi.org/10.18637/jss.v014.i09>
- Gonzalez, J., & Muñoz, P. (2006). e-status: An automatic web-based problem generator-applications to statistics. *Computer Applications in Engineering Education*, 14, 151–159. <https://doi.org/10.1002/cae.20071>
- González, J., Jover, L., Cobo, E., & Muñoz, P. (2010). A web-based learning tool improves student performance in statistics: A randomized masked trial. *Computers & Education*, 55, 704–713. <https://doi.org/10.1016/j.compedu.2010.03.003>
- Grün, B., & Zeileis, A. (2009). Automatic generation of exams in R. *Journal of Statistical Software*, 29(10), 1–14. <https://doi.org/10.18637/jss.v029.i10>
- Hardin, L. E. (2003). Problem-solving concepts and theories. *Journal of veterinary medical education*, 30(3), 226–229. <https://doi.org/10.3138/jvme.30.3.226>

- Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2017). Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments. *IEEE Transactions on Learning Technologies*, 10(2), 140-153. <https://doi.org/10.1109/TLT.2015.2513387>
- Lane, D., & Scott, D. (2000). Simulations, case studies, and an online text: a web-based resource for teaching statistics. *Metrika*, 51, 67-90. <https://doi.org/10.1007/s001840000045>
- Larreamendy-Joerns, J., Leinhardt, G., & Corredor, J. (2005). Six online statistics courses: Examination and review. *The American Statistician*, 59, 240-251. <https://doi.org/10.1198/000313005X54162>
- MERLOT (2018). *Multimedia Educational Resource for Learning and Online Teaching*. Retrieved on 24 January 2018 from <http://statistics.merlot.org>
- Mostow, J., & Beck, J. (2009). What, how, and why should tutors log? In: *Proceedings of Educational Data Mining 2009*, 269-278.
- Muenchen, R. A. (2017, June 19). The Popularity of Data Science Software [Web log message]. Retrieved from <http://r4stats.com/articles/popularity/>
- Piatetsky, G. (2017, May 22). New Leader, Trends, and Surprises in Analytics, Data Science, Machine Learning Software Poll [Web log message]. Retrieved from <https://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html>
- PSLC (2013). *Guide to the tutor message format. a standard XML vocabulary for logging student and tutor actions*. Retrieved on 24 January 2018 from <https://pslcdatashop.web.cmu.edu/dtd/guide/index.html>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Ritter, S., & Blessing, S. (1998). Authoring tools for component-based learning environments. *Journal of the Learning Sciences*, 7, 107-132. [https://doi.org/10.1207/s15327809jls0701\\_4](https://doi.org/10.1207/s15327809jls0701_4)
- Ritter, S., & Koedinger, K. (1996). An architecture for plug-in tutor agents. *Journal of Artificial Intelligence in Education*, 7, 315-347.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40, 601-618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51, 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
- Rosales, F., Garcia, A., Rodriguez, S., Pedraza, J., Méndez, R., & Nieto, M. (2008). Detection of plagiarism in programming assignments. *IEEE Transactions on Education*, 51, 174-183. <https://doi.org/10.1109/TE.2007.906778>
- Song, C., & Ma, K. (2008). Applications of data mining in the education resource based on XML. In: *2008 International Conference on Advanced Computer Theory and Engineering*, IEEE, 943-946. <https://doi.org/10.1109/ICACTE.2008.170>
- Stevens, R., Beal, C., & Sprang, M. (2009). Tracking the development of problem solving skills with learning trajectories. In: *Proceedings of the 17th International Conference on Computers in Education*, 99-106.
- Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student attitudes to learning business statistics: Comparison of online and traditional methods. *Educational Technology & Society*, 7, 9-20.
- Symanzik, J., & Vukasinovic, N. (2003). Teaching experiences with a course on web-based statistics. *The American Statistician*, 57, 46-50. <https://doi.org/10.1198/0003130031126>
- VanLehn, K., Koedinger, K., Skogsholm, A., Nwagwe, A., Hausmann, R., Weinstein, A., & Billings, B. (2009). What's in a step? toward general, abstract representations of tutoring system log data. *User Modeling 2007*, 455-459.
- West, R., & Ogden, R. (1998). Interactive demonstrations for statistics education on the world wide web. *Journal of Statistics Education*, 6(3). <https://doi.org/10.1080/10691898.1998.11910624>
- Zeileis, A., Umlauf, N., & Leisch, F. (2014). Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond. *Journal of Statistical Software*, 58(1), 1-36. <https://doi.org/10.18637/jss.v058.i01>