Flexible Generation of E-Learning Exams in R:
Moodle Quizzes, OLAT Assessments, and Beyond

Achim Zeileis, Nikolaus Umlauf, Friedrich Leisch

http://eeecon.uibk.ac.at/~zeileis/

Overview

- Motivation and challenges
- R package **exams**
- Exercises
- Exams
  - Combination of exercises
  - PDF output
  - HTML output
  - XML for **Moodle** or **OLAT**
- Discussion

Motivation and challenges

Motivation:

- Introductory statistics and mathematics courses for business and economics students at WU Wien and Universität Innsbruck.
- Courses are attended by more than 1,000 students per semester.
- Several lecturers teach lectures and tutorials in parallel.
- Need for integrated teaching materials: Presentation slides, collections of exercises, exams, etc.

Challenges:

- *Scalable exams:* Automatic generation of a large number of different exams, both written and online.
- *Associated self-study materials:* Collections of exercises and solutions from the same pool of examples.
- *Joint development:* Development and maintenance of a large pool of exercises in a multi-author and cross-platform setting.

R package exams

Tools chosen: R (for random data generation and computations) and LaTeX (for mathematical notation) $\Rightarrow$ `Sweave`.

Design principles of package exams:

- Each exercise template (also called "exercise" for short) is a single `Sweave` file (`.Rnw`) interweaving R code for data generation and LaTeX code for describing question and solution.
- Exams can be generated by randomly drawing different versions of exercises from a pool of such `Sweave` exercise templates. The resulting exams can be rendered into various formats including PDF, HTML, **Moodle** XML, or QTI 1.2 (for **OLAT** or **OpenOLAT**).
- Solutions for exercises can be multiple/single-choice answers, numeric values, short text answers, or a combination thereof (cloze).

Exercises

Exercise templates: `Sweave` files composed of

- R code chunks (within `<<>>=` and `@`) for random data generation.
- Question and solution descriptions contained in LaTeX environments of corresponding names. Both can contain R code chunks again or include data via `\Sexpr{}`.
- Metainformation about type (numeric, multiple choice, ...), correct solution etc. In LaTeX style but actually commented out.

Simple geometric example:

- Computation of the distance between two points p and q in a Cartesian coordinate system (via the Pythagorean formula).
- Template `dist.Rnw` contained in **exams** package.

```
R> library("exams")
R> exams2pdf("dist.Rnw")
```

Exercises: dist.Rnw

```
<<echo=FALSE, results=hide>>=
p <- c(sample(1:3, 1), sample(1:5, 1))
q <- c(sample(4:5, 1), sample(1:5, 1))
sol <- sqrt(sum((p - q)^2))
@

\begin{question}
What is the distance between the two points
$p = (\Sexpr{p[1]}, \Sexpr{p[2]})$ and $q = (\Sexpr{q[1]}, \Sexpr{q[2]})$
in a Cartesian coordinate system?
\end{question}

\begin{solution}
The distance $d$ of $p$ and $q$ is given by
$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).

Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} =
  \sqrt{(\Sexpr{p[1]} - \Sexpr{q[1]})^2 + (\Sexpr{p[2]} - \Sexpr{q[2]})^2}
   = \Sexpr{round(sol, digits = 3)}$.
[...]
\end{solution}

%% \extype{num}
%% \exsolution{\Sexpr{round(sol, digits = 3)}}
%% \exname{Euclidean distance}
%% \extol{0.01}
```

Exercises: `dist.Rnw`

```
<<echo=FALSE, results=hide>>=
p <- c(sample(1:3, 1), sample(1:5, 1))
q <- c(sample(4:5, 1), sample(1:5, 1))
sol <- sqrt(sum((p - q)^2))
@

\begin{question}
What is the distance between the two points
$p = (\Sexpr{p[1]}, \Sexpr{p[2]})$ and $q = (\Sexpr{q[1]}, \Sexpr{q[2]})$
in a Cartesian coordinate system?
\end{question}

\begin{solution}
The distance $d$ of $p$ and $q$ is given by
$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).

Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} =
  \sqrt{(\Sexpr{p[1]} - \Sexpr{q[1]})^2 + (\Sexpr{p[2]} - \Sexpr{q[2]})^2}
   = \Sexpr{round(sol, digits = 3)}$.
[...]
\end{solution}

%% \extype{num}
%% \exsolution{\Sexpr{round(sol, digits = 3)}}
%% \exname{Euclidean distance}
%% \extol{0.01}
```

Exercises: dist.Rnw

```
<<echo=FALSE, results=hide>>=
p <- c(sample(1:3, 1), sample(1:5, 1))
q <- c(sample(4:5, 1), sample(1:5, 1))
sol <- sqrt(sum((p - q)^2))
@

\begin{question}
What is the distance between the two points
$p = (\Sexpr{p[1]}, \Sexpr{p[2]})$ and $q = (\Sexpr{q[1]}, \Sexpr{q[2]})$
in a Cartesian coordinate system?
\end{question}

\begin{solution}
The distance $d$ of $p$ and $q$ is given by
$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).

Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} =
  \sqrt{(\Sexpr{p[1]} - \Sexpr{q[1]})^2 + (\Sexpr{p[2]} - \Sexpr{q[2]})^2}
   = \Sexpr{round(sol, digits = 3)}$.
[...]
\end{solution}

%% \extype{num}
%% \exsolution{\Sexpr{round(sol, digits = 3)}}
%% \exname{Euclidean distance}
%% \extol{0.01}
```

Exercises: dist.Rnw

```
<<echo=FALSE, results=hide>>=
p <- c(sample(1:3, 1), sample(1:5, 1))
q <- c(sample(4:5, 1), sample(1:5, 1))
sol <- sqrt(sum((p - q)^2))
@

\begin{question}
What is the distance between the two points
$p = (\Sexpr{p[1]}, \Sexpr{p[2]})$ and $q = (\Sexpr{q[1]}, \Sexpr{q[2]})$
in a Cartesian coordinate system?
\end{question}

\begin{solution}
The distance $d$ of $p$ and $q$ is given by
$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).

Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} =
  \sqrt{(\Sexpr{p[1]} - \Sexpr{q[1]})^2 + (\Sexpr{p[2]} - \Sexpr{q[2]})^2}
   = \Sexpr{round(sol, digits = 3)}$.
[...]
\end{solution}

%% \extype{num}
%% \exsolution{\Sexpr{round(sol, digits = 3)}}
%% \exname{Euclidean distance}
%% \extol{0.01}
```

Exercises: `dist.Rnw`

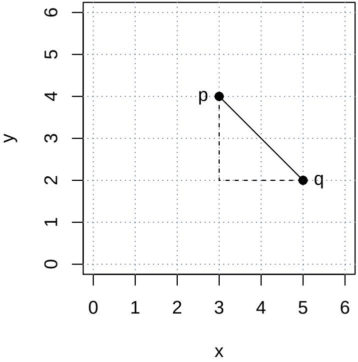
```
<<echo=FALSE, results=hide>>=
p <- c(sample(1:3, 1), sample(1:5, 1))
q <- c(sample(4:5, 1), sample(1:5, 1))
sol <- sqrt(sum((p - q)^2))
@

\begin{question}
What is the distance between the two points
$p = (\Sexpr{p[1]}, \Sexpr{p[2]})$ and $q = (\Sexpr{q[1]}, \Sexpr{q[2]})$
in a Cartesian coordinate system?
\end{question}

\begin{solution}
The distance $d$ of $p$ and $q$ is given by
$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).

Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} =
  \sqrt{(\Sexpr{p[1]} - \Sexpr{q[1]})^2 + (\Sexpr{p[2]} - \Sexpr{q[2]})^2}
   = \Sexpr{round(sol, digits = 3)}$.
[...]
\end{solution}

%% \extype{num}
%% \exsolution{\Sexpr{round(sol, digits = 3)}}
%% \exname{Euclidean distance}
%% \extol{0.01}
```

Exercises: LaTeX output of Sweave("dist.Rnw")

```
\begin{question}
What is the distance between the two points
$p = (3, 4)$ and $q = (5, 2)$
in a Cartesian coordinate system?
\end{question}

\begin{solution}
The distance $d$ of $p$ and $q$ is given by
$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).

Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} =
  \sqrt{(3 - 5)^2 + (4 - 2)^2}
   = 2.828$.

\includegraphics{dist-002}
\end{solution}

%% \extype{num}
%% \exsolution{2.828}
%% \exname{Euclidean distance}
%% \extol{0.01}
```

Exercises: PDF output of `exams2pdf("dist.Rnw")`

- **Problem**
  What is the distance between the two points $p = (3, 4)$ and $q = (5, 2)$ in a Cartesian coordinate system?
  Solution
  The distance d of p and q is given by $d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2$ (Pythagorean formula).
  Hence $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} = \sqrt{(3 - 5)^2 + (4 - 2)^2} = 2.828$.



Exams: Combination of exercises

Idea: An exam is simply a list of exercise templates. For example, using statistics exercise templates contained in **exams**.

```
R> myexam <- list(
+    "boxplots",
+    c("confint", "ttest", "tstat"),
+    c("anova", "regression"),
+    "scatterplot",
+    "relfreq"
+  )
```

Draw random exams:

- First randomly select one exercise from each list element.
- Generate random numbers/input for each selected exercise.
- Combine all exercises in output file(s) (PDF, HTML, ...).

Exams: Combination of exercises

Interfaces: Generate multiple exams via `exams2pdf()`, `exams2html()`, `exams2moodle()`, `exams2qti12()`, ...

Workhorse function: Internally, all interfaces call `xexams()` that handles (temporary) files/directories and carries out four steps.

1. *Weave:* Each of the selected exercise `.Rnw` files is weaved into a `.tex` file. Default: The standard `Sweave()` function.
2. *Read:* Each resulting `.tex` file is read into an R list with question, solution, and metainformation. Default: `read_exercise()`.
3. *Transform:* Each of these exercise-wise list objects can be transformed, e.g., by converting LaTeX text to HTML. Default: No transformation.
4. *Write:* The (possibly transformed) lists of exercises, read into R for each exam object, can be written out to one ore more files per exam in an output directory. Default: No files are written.

Exams: PDF output

`exams2pdf()`:

- The *write* step embeds all questions/solutions into (one or more) master LaTeX template(s).
- LaTeX templates control whether solutions are shown, what the title page looks like, etc.
- Compilation of each exam via pdfLaTeX (called from within R).

A single exam is popped up in a PDF viewer:

```
R> exams2pdf(myexam, template = "exam")
```

Multiple exams are written to an output directory:

```
R> odir <- tempfile()
R> set.seed(1090)
R> exams2pdf(myexam, n = 3, dir = odir,
+    template = c("exam", "solution"))
```

Exams: PDF output

R University
Statistics Exam 2013-07-07 **Exam ID 00001**

Name: ______________________________

Student ID: ______________________________

Signature: ______________________________

1. (a) ☐ (b) ☐ (c) ☐ (d) ☐ (e) ☐
2. ☐☐☐☐☐☐ . ☐☐☐
3. ☐☐☐☐☐☐ . ☐☐☐
4. (a) ☐ (b) ☐ (c) ☐ (d) ☐ (e) ☐
5. (a) ☐ (b) ☐ (c) ☐ (d) ☐ (e) ☐

1. In Figure 1 the distributions of a variable given by two samples (A und B) are represented by parallel boxplots. Which of the following statements are correct? *(Comment: The statements are either about correct or clearly wrong.)*

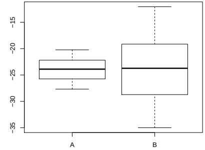


Figure 1: Parallel boxplots.

   (a) The location of both distributions is about the same.
   (b) Both distributions contain no outliers.
   (c) The spread in sample A is clearly bigger than in B.
   (d) The skewness of both samples is similar.
   (e) Distribution A is about symmetric.

2. A machine fills milk into 500ml packages. It is suspected that the machine is not working correctly and that the amount of milk filled differs from the setpoint $\mu_0 = 500$. A sample of 226 packages filled by the machine are collected. The sample mean $\bar{y}$ is equal to 499.7 and the sample variance s^2_{n-1} is equal to 576.1.

   Test the hypothesis that the amount filled corresponds on average to the setpoint. What is the absolute value of the *t* test statistic?

3. For 49 firms the number of employees X and the amount of expenses for continuing education Y (in EUR) were recorded. The statistical summary of the data set is given by:

	Variable X	Variable Y
Mean	58	232
Variance	124	1606

   The correlation between X and Y is equal to 0.65.

   Estimate the expected amount of money spent for continuing education by a firm with 60 employees using least squares regression.

4. Figure 2 shows a scatterplot. Which of the following statements are correct?

Exams: HTML output

`exams2html()`:

- In the *transform* step, LaTeX text is converted to HTML using Ian H. Hutchinson's **TtH** (TeX to HTML) package.
- Mathematical notation is either represented using MathML (`ttm`), requiring a suitable browser (e.g., Firefox), or plain HTML (`tth`).
- No LaTeX installation needed, but also limited to LaTeX commands supported by **TtH**.
- Links to dynamically generated data can be easily included, e.g., `\url{mydata.rda}`.
- The *write* step embeds everything into HTML templates and writes out one HTML file per exam.

A single exam is popped up in a browser, multiple exams are written to an output directory:

```
R> set.seed(1090)
R> exams2html(myexam, n = 3, dir = odir)
```

Exams: HTML output

File Edit View History Bookmarks Tools Help

Exam 1

file:///tmp/RtmpjjIbQz/file2c8d10dbc901/plain1.html

DuckDuckGo

Exam 1

1. **Question**

   In Figure the distributions of a variable given by two samples (A und B) are represented by parallel boxplots. Which of the following statements are correct? *(Comment: The statements are either about correct or clearly wrong.)*

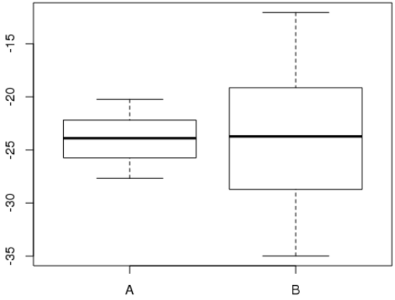


Figure 1: Parallel boxplots.

a. The location of both distributions is about the same.
b. Both distributions contain no outliers.
c. The spread in sample A is clearly bigger than in B.
d. The skewness of both samples is similar.
e. Distribution A is about symmetric.

Exams: Moodle XML

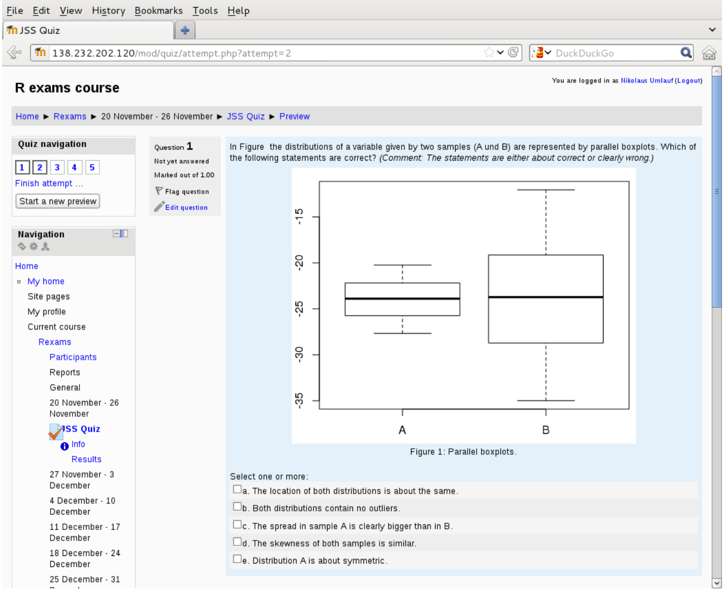
`exams2moodle()`:

- As for HTML output, all LaTeX text is *transformed* to HTML (plus MathML).
- Rather than writing out one file per exam, a single **Moodle** XML file encompassing all exams is produced.
- All supplementary materials (graphics, data, etc.) are embedded into the HTML code directly using Base64 encoding.
- The resulting `.xml` file can be easily imported into a question bank in **Moodle** and then be used within a **Moodle** quiz.

Multiple replications are written to a single XML file in the output directory:

```
R> set.seed(1090)
R> exams2moodle(myexam, n = 3, dir = odir)
```

Exams: Moodle XML



Exams: QTI 1.2 for OLAT

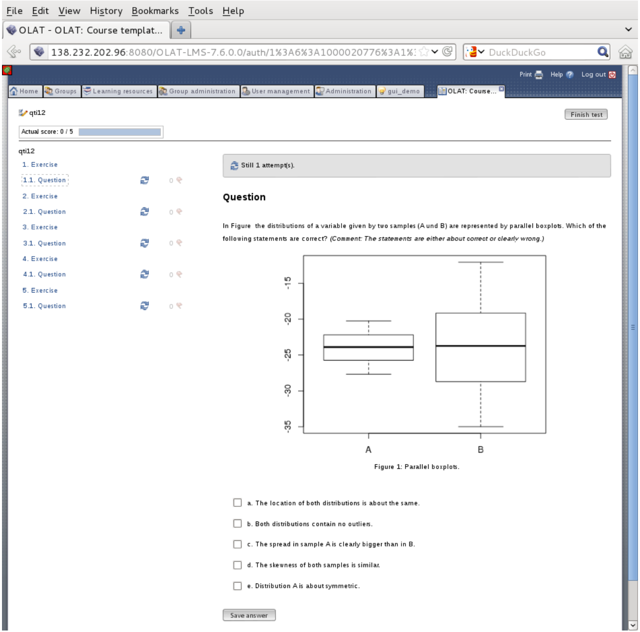
`exams2qti12()`:

- As for HTML output, all LaTeX text is *transformed* to HTML (plus MathML).
- Rather than writing out one file per exam, a single `.zip` archive is produced, containing the QTI 1.2 XML file plus supplementary materials (graphics, data, etc.) if any.
- Base64 encoding is used for graphics by default, but not for other supplements.
- QTI 1.2 is an international standard for e-learning exams.
- The `.zip` files can be easily imported into **OLAT** (or **OpenOLAT**) when configuring an exam.

Multiple replications are written to a single zipped XML file in the output directory:

```
R> set.seed(1090)
R> exams2qti12(myexam, n = 3, dir = odir)
```

Exams: QTI 1.2 for OLAT



Discussion

Package exams:

- Framework for automatic generation of simple (mathematical or statistical) exams and associated self-study materials.
- Based on independent exercises in `Sweave` format which can be compiled into exams (or other collections of exercises).
- Version 1 (Grün and Zeileis 2009) only supported PDF output, version 2 (Zeileis, Umlauf, Leisch 2012) adds an extensible toolbox for various output formats including HTML, **Moodle** XML, and QTI 1.2 (for **OLAT**).
- Contributing to the pool of exercises only requires knowledge of `Sweave` and minimal markup for metainformation.
- Hosted on R-Forge, providing a support forum:
  `http://R-Forge.R-project.org/projects/exams/`

Discussion

At Universität Innsbruck:

- Mathematics course with **OLAT** support (summer/winter term 2012/13 combined: more than 3,000 participants).
- Team of about 10 persons (professors, lecturers, student assistants) contribute to the pool of exercises.
- During the semester, several online tests (and self tests) are carried out in **OLAT** (via `exams2qti12`) using numerical and multiple-choice exercises.
- Two written exams (via `exams2pdf`) are carried out using single-choice exercises. Results are scanned by university services and processed by some optical character recognition.

References

Zeileis A, Grün B, Leisch F, Umlauf N (2013). ***exams**: Automatic Generation of Exams in R.* R package version 1.9-5.
URL `http://CRAN.R-project.org/package=exams`

Zeileis A, Umlauf N, Leisch F (2012). "Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond." *Working Paper 2012-27*, Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universität Innsbruck.
URL `http://EconPapers.RePEc.org/RePEc:inn:wpaper:2012-27`.

Grün B, Zeileis A (2009). "Automatic Generation of Exams in R." *Journal of Statistical Software*, **29**(10), 1–14.
URL `http://www.jstatsoft.org/v29/i10/`