Randomized Exam Generation and Grading

Version 1.1.1

M. West

2013-01-05

1 Concepts

The exam *library* contains N_Q *questions*, each with up to N_V *variants*. Each question variant contains up to N_A *answers*, exactly one of which must be the *correct answer*. The correct answer for variant V of question Q is denoted $C(Q, V)$. The questions within the library are grouped into *zones*.

We randomly generate N_e *exams*. Each exam contains exactly one randomly chosen variant of each library question, with the question order randomly permuted within each zone. The order of the zones is not permuted. Questions are numbered globally with the library and exams, not per-zone. Question q on exam e thus corresponds to variant $V(e, q)$ of question $Q(e, q)$ in the library. The answer order for each exam question is also a random permutation of the original answer order, so $A(e, q, a)$ is the library answer corresponding to exam answer a.

We use capital letters for variables for quantities within the library and small letters for quantities within the randomly generated exams and students. Exams, questions, variants, and students are all numbered with Arabic numerals starting from 1. Answers are numbered with roman letters starting from A.

Each randomly generated exam is identified by a *key* $K(e)$ of length N_K, which is a sequence of answers $K_1, \ldots, K_{N_K}$. The keys are encoded on the Scantron forms as the answers to the final set of questions. For example, on a 96-question Scantron form, the 4-digit key "ECEA" is encoded as the answers (93) E, (94) C, (95) E, and (96) A.

During the exam each student completes a single Scantron form, resulting in N_s Scantron forms for an equal number of students. These forms are scanned and the raw data can be processed to derive the points for each student.

In summary, the size variables used are:

N_Q	number of questions (in the library and on each exam)
N_e	number of exams
N_A	maximum number of answers per question
N_V	maximum number of variants per question
N_s	number of students (equal to the number of Scantron forms)
N_g	number of ranked groups of students for statistics generation
N_D	number of answer-digits required to encode the exam number
N_K	number of answer-digits in the exam key

2 Processing pipeline

The full processing pipeline for making, using, and grading a randomized exam is shown in Figure 1. The purple output files can be inspected and hand-edited before processing runs to control the processing pipeline.

library.tex

randexam proc-lib

exams.tex

solutions.csv

students, Scantron

points.csv

scantron.dat

specs.csv

randexam proc-scan

answers.csv

randexam proc-ans

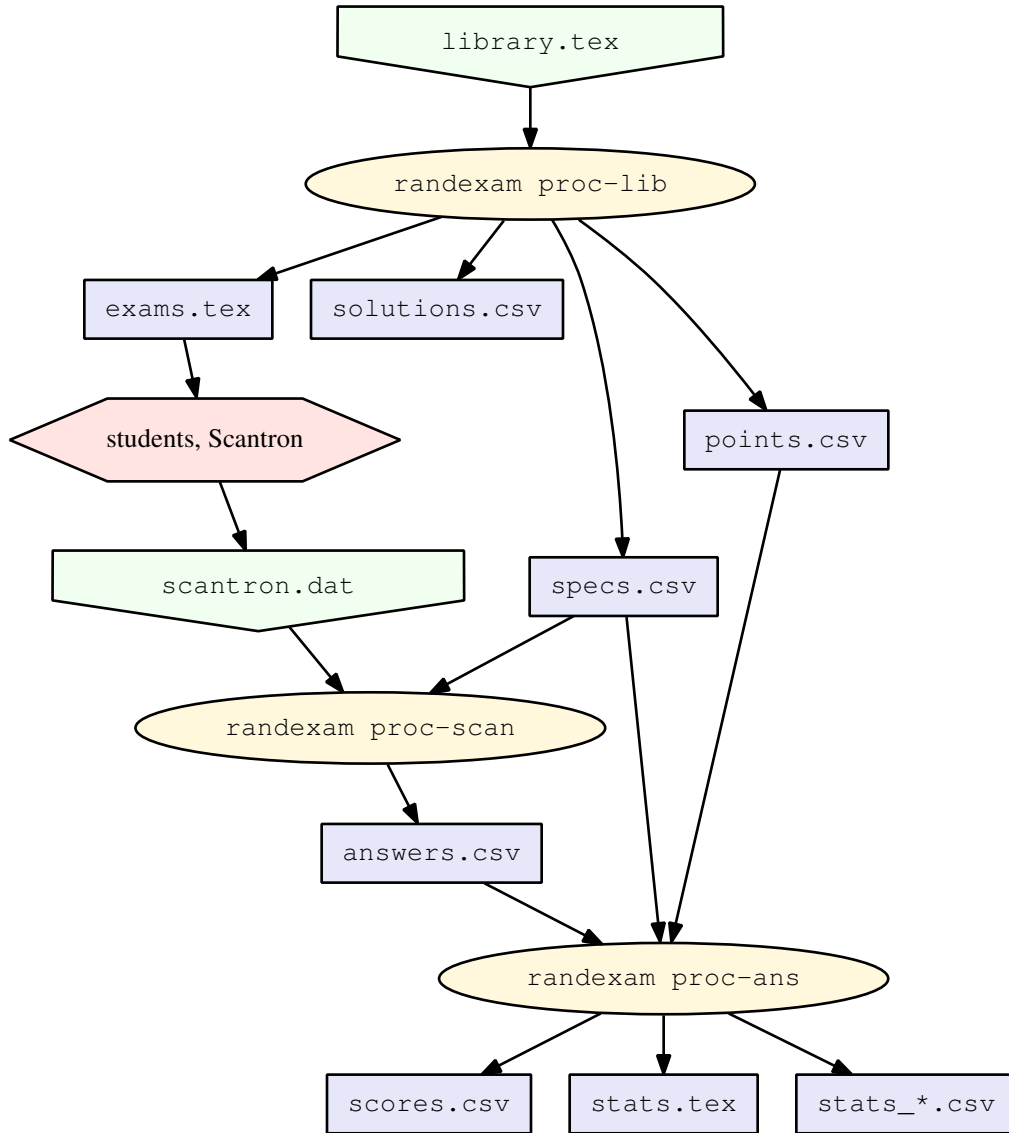scores.csv

stats.tex

stats_*.csv

Figure 1: The pipeline for making and grading the randomized exams. The green pentagons are input files that require human effort to generate, the purple boxes are output files, the yellow ellipses are program runs, and the pink hexagon is the students sitting the exam and the scanning of the Scantron forms.

3 File formats

All library and exam question are formatted in single TeX files. All data is stored in CSV (comma-separated values) files for easy editing by hand in a spreadsheet or text editor.

3.1 `library.tex`

The `library_template.tex` file should be copied to `library.tex` and edited to fill in the questions. Non-comment text may only go into the locations indicated in the file. This file should be a valid LaTeX file that can be typeset and edited to prepare the questions. To generate the randomized exams with the `randexam` script, the format of `library.tex` must conform to the grammar shown in Figure 2. The basic structure of `library.tex` is thus a tree, as shown in Figure 3. An outline of the LaTeX in the `library.tex` is given in Figure 4. The `randexam` script processes `library.tex` using a state machine. The state transition graph is shown in Figure 5. Note the following points:

- Zones serve two purposes: (1) to group questions so that random permutations only occur within a zone, not between zones; and (2) to insert text at a certain point in the exam. Zones do not need to contain any questions, so they can be used just to insert text, for example a formula sheet at the start or end of the exam.

- Points are assigned on a per-question basis, so all variants of a given question are worth the same number of points.

- Only one answer can be marked as correct for each variant. This answer will be awarded the full points for that question, and all other answers will be worth zero points. This can be adjusted by manually editing `points.csv`, which allows points to be assigned to any answer of any variant in an arbitrary way, including negative points and real-valued points.

- Questions do not have any text associated with them directly. If there is text that is common to all variants of a question then it must be repeated in each variant.

- The preamble in `library.tex` is discarded when it is processed and the preamble of `exams.tex` is hard-coded in the `randexam` script. To modify the `exams.tex` preamble, for example to include another package, the `randexam` script should be edited directly.

3.2 `exams.tex`

This is the generated file containing all of the randomized exams. Each coverpage ends with the exam key that needs to be filled in by the student.

3.3 `specs.csv`

This file encodes the arrays:

array	dimensions	description
$K(e)$	N_e	key for exam e
$Q(e, q)$	$N_e \times N_Q$	library question for question q on exam e
$V(e, q)$	$N_e \times N_Q$	library question variant for question q on exam e
$A(e, q, a)$	$N_e \times N_Q \times N_A$	library answer corresponding to answer a for question q on exam e

The first column is the exam number, the second column is the exam key, and then triples of columns identify the library question Q, variant V, and library answer order. As an example:

```
e,K(e),"Q(e,q=1)","V(e,q=1)","A(e,q=1,:)","Q(e,q=2)","V(e,q=2)","A(e,q=2,:)"
1,ADC,2,2,DEBCA,3,2,DCEBA
2,BED,1,2,BDCAE,2,1,DECBA
```

⟨*library*⟩	::=	⟨*preamble text*⟩ '\begin{document}' ⟨*body*⟩ '\end{document}'
⟨*body*⟩	::=	⟨*coverpage text*⟩ ⟨*zone-list*⟩
⟨*zone-list*⟩	::=	⟨*zone*⟩ ⟨*zone-list*⟩ | ⟨*zone*⟩
⟨*zone*⟩	::=	'\zone' ⟨*zone text*⟩
	|	'\zone' ⟨*zone text*⟩ ⟨*question-list*⟩
⟨*question-list*⟩	::=	⟨*question*⟩ ⟨*question-list*⟩ | ⟨*question*⟩
⟨*question*⟩	::=	'\question{' ⟨*points*⟩ '}' ⟨*variant-list*⟩
⟨*variant-list*⟩	::=	⟨*variant*⟩ ⟨*variant-list*⟩ | ⟨*variant*⟩
⟨*variant*⟩	::=	'\variant' ⟨*variant text*⟩ '\begin{answers}' ⟨*answer-list*⟩ '\end{answers}'
⟨*answer-list*⟩	::=	⟨*answer*⟩ ⟨*answer-list*⟩ | ⟨*answer*⟩
⟨*answer*⟩	::=	'\answer' ⟨*answer text*⟩
	|	'\correctanswer' ⟨*answer text*⟩

Figure 2: Grammer for the `library.tex` input file expressed in partial Backus-Naur Form. The *text* entries refer to any valid LaTeX text, and *points* is a numeric value.

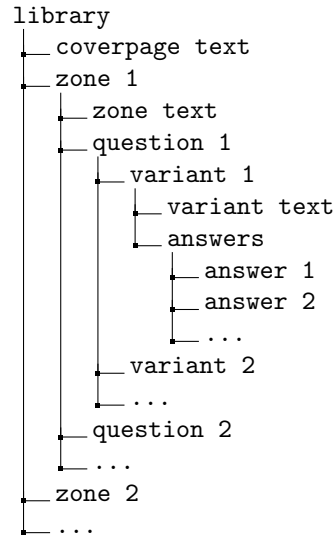Figure 3: Tree structure of the `library.tex` input file.

```
<preamble text>
\begin{document}
    <coverpage text>
    \zone
        <zone text>
        \question{<points>}
            \variant
                <variant text>
                \begin{answers}
                    \answer <answer text>
                    \correctanswer <answer text>
                    <more answers...>
                \end{answers}
            <more variants...>
        <more questions...>
    <more zones...>
\end{document}
```

Figure 4: Outline of the `library.tex` input file.

Here the key of the first exam is $K(1) = \text{ADC}$ and the second question on this exam is variant $V(1,2) = 2$ of library question $Q(1,2) = 3$. The first answer for this question on the exam paper is answer $A(1,2,1) = \text{D}$ from the library question, while the second answer on the exam paper is answer $A(1,2,2) = \text{C}$ from the library question.

3.4 `points.csv`

The points file encodes the array:

array	dimensions	description
$P(Q,V,A)$	$N_\mathrm{Q} \times N_\mathrm{V} \times N_\mathrm{A}$	points awarded for giving answer A to variant V of question Q

This is stored as a four-column file with entries like:

```
Q,V,A,"P(Q,V,A)"
...
2,1,A,0.0
2,1,B,0.0
2,1,C,0.0
2,1,D,1.0
2,1,E,0.0
```

Here some lines have been removed, as indicated by the ellipsis. This shows that for variant 1 of question 2, answer D is worth 1 point and all other answers are not worth any points, so $P(2,1,4) = 1.0$ and $P(2,1,A) = 0$ for $A \neq 4$.

3.5 `solutions.csv`

This file gives the correct solutions for each exam in case grading by hand is necessary. This thus encodes the array:

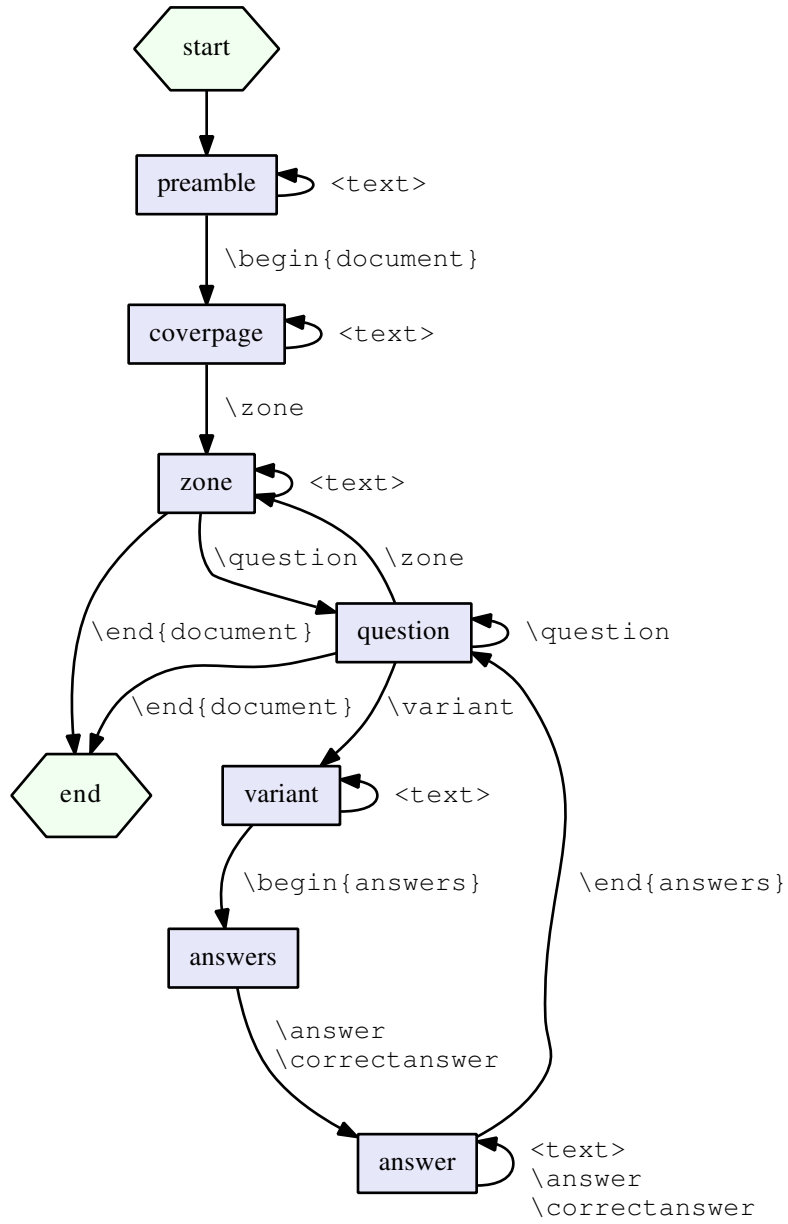| array	dimensions	description
$c(e,q)$	$N_\mathrm{e} \times N_\mathrm{Q}$	the correct answer for question q on exam e

Figure 5: The state machine used by the `randexam` script to read the `library.tex` input file.

The first two columns give the exam number and key for reference, and then subsequent columns give the correct answers. For example:

```
e,K(e),"C(e,q=1)","C(e,q=2)","C(e,q=3)"
1,ADC,A,A,D
2,BED,D,A,A
3,CAE,A,E,D
4,DBA,A,A,D
5,ECB,B,E,A
```

Here the third exam has key $K(3) = \text{CAE}$ and the correct answer for the first question on this exam is $c(3,1) = A$, while the correct answer for the second question is $c(3,2) = E$, etc.

The correct answers each exam are derived from the correct answers in the library file, given for each library question and each variant in the array:

array	dimensions	description
$C(Q,V)$	$N_Q \times N_V$	the correct answer for variant V of library question Q

The derivation of $c(e,q)$ is then:

$$A(e, q, c(e,q)) = C\big(Q(e,q), V(e,q)\big). \tag{1}$$

3.6   scantron.dat

This is the raw data file from the Scantron machine. It contains lines like:

```
721000001001120412001Y   5380      1    N DEVILLE    R993502067    RDEVILLE 434
721000002001120412001Y   5380      1    N WEST       M795969582    MWEST    252
```

The actual files send from the scanning office should be renamed to `scantron.dat`.

3.7   answers.csv

This file is essentially just a CSV version of the data in `scantron.dat`. It encodes the arrays:

array	dimensions	description
$u(s,i)$	$N_s \times 4$	student last name, first initial, student ID number, and NetID
$a(s,q)$	$N_s \times N_Q$	answer given to question q by student s
$k(s)$	N_s	key filled in by student s

The first column in the data file is the student number, then the next four columns store the student information, the sixth column stores the exam key, and then columns 7 and up store the answers to the questions in exam order. As an example:

```
s,Name,Initial,Number,NetID,k(s),"a(s,q=1)","a(s,q=2)","a(s,q=3)"
1,DEVILLE,R,993502067,RDEVILLE,CAE,D,C,D
2,WEST,M,795969582,MWEST,ECB,B,E,B
```

Here we see the second student has information giving the last name $u(2,1) = \text{WEST}$, the first initial $u(2,2) = \text{M}$, the student number $u(2,3) = 795969582$, and the NetID $u(2,4) = \text{MWEST}$. The exam key for this student is $k(2) = \text{ECB}$ and the student gave answer $a(2,1) = B$ for question 1 on this exam paper, answer $a(2,2) = E$ for question 2, etc.

3.8 `scores.csv`

This file encodes the array $P_{\mathrm{se}}(s)$ described in Section 4. For reference, the student number and information in array $u(s, i)$ is also included in the first columns. As an example:

```
s,Name,Initial,Number,NetID,P_s(s)
1,DEVILLE,R,993502067,RDEVILLE,2.0
2,WEST,M,795969582,MWEST,2.0
```

Here we see that both students have total points $P_{\mathrm{s}}(1) = P_{\mathrm{s}}(2) = 2.0$.

3.9 `stats.tex`

This file presents various summary statistics relating to the student and question performance in the exam. It should be processed with `pdflatex` to produce the output `stats.pdf` file.

3.10 `stats_*.csv`

Multiple files are written to encode the statistics arrays described in Section 5. These are named after the variables they contain, so $n_{\mathrm{QV}}^{\mathrm{a}}$ is be stored in `stats_n_a_QV.csv`, for example. The format is broadly similar to those files described in detail above, with column headings that specify the contents.

4 Grading

The map from student number s to exam number e is given by the array

array	dimensions	description
$e(s)$	N_{s}	exam number corresponding to student number s

This array is determined so that

$$K(e(s)) = k(s). \tag{2}$$

The points awarded to each student can be determined using the grading key for any exam, not just the particular exam that the student actually took. To make this clear, we consider the arrays

array	dimensions	description
$P_{\mathrm{seq}}(s, e, q)$	$N_{\mathrm{s}} \times N_{\mathrm{e}} \times N_{\mathrm{Q}}$	points awarded to student s on question q scored for exam e
$P_{\mathrm{se}}(s, e)$	$N_{\mathrm{s}} \times N_{\mathrm{e}}$	total points awarded to student s scored for exam e
$P_{\mathrm{sq}}(s, q)$	$N_{\mathrm{s}} \times N_{\mathrm{Q}}$	points award to student s on question q using the correct exam
$P_{\mathrm{s}}(s)$	N_{s}	total points awarded to student s scored using the correct exam
$P_{sQ}(s, Q)$	$N_{\mathrm{s}} \times N_{\mathrm{Q}}$	points awarded to student s for library question Q

These quantities are computed by:

$$P_{\mathrm{seq}}(s, e, q) = P\Big(Q(e, q), V(e, q), A\big(e, q, a(s, q)\big)\Big) \tag{3}$$

$$P_{\mathrm{se}}(s, e) = \sum_{q=1}^{N_{\mathrm{Q}}} P_{\mathrm{seq}}(s, e, q) \tag{4}$$

$$P_{\mathrm{sq}}(s, q) = P_{\mathrm{seq}}(s, e(s), q) \tag{5}$$

$$P_{\mathrm{s}}(s) = P_{\mathrm{se}}(s, e(s)) \tag{6}$$

$$P_{\mathrm{sQ}}(s, Q(e(s), q)) = P_{\mathrm{seq}}(s, e(s), q). \tag{7}$$

We expect that $e(s)$ maximizes $P_{\mathrm{se}}(s, e)$ for each s, as grading with the "wrong" exam key should give random choices for each answer. Indeed, for $\hat{e} \neq e(s)$, $P_{\mathrm{se}}(s, \hat{e})$ should be approximately $N_{\mathrm{Q}}/N_{\mathrm{A}}$, assuming that each question is worth one point if correct. Some significant deviations are likely, however, when the number of students N_{s} much larger than the number of questions N_{Q}.

5 Statistics

As well as computing the scores for each student, we can assess how the students as a group performed on the different questions and variants. Basic statistics for this assessment are listed below.

array	dimensions	description
$n_{\mathrm{e}}^{\mathrm{s}}(e)$	N_{e}	number of students taking exam e
$n_{\mathrm{QV}}^{\mathrm{s}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	number of students given variant V of question Q
$n_{\mathrm{QVA}}^{\mathrm{a}}(Q,V,A)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}} \times N_{\mathrm{A}}$	number of students giving answer A for variant V of question Q
$n_{\mathrm{QV}}^{\mathrm{a}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	number of students giving an answer for variant V of question Q
$n_{\mathrm{QV}}^{\mathrm{na}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	number of students not giving an answer for variant V of question Q
$n_{\mathrm{Q}}^{\mathrm{a}}(Q)$	N_{Q}	number of students giving an answer for question Q
$r_{\mathrm{QVA}}^{\mathrm{a}}(Q,V,A)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}} \times N_{\mathrm{A}}$	fraction of students giving answer A for variant V of question Q
$r_{\mathrm{QV}}^{\mathrm{na}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	fraction of students not giving an answer for variant V of question Q
$P_{\mathrm{QV}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	total points awarded for variant V of question Q
$P_{\mathrm{Q}}(Q)$	N_{Q}	total points awarded for question Q
$\bar{P}_{\mathrm{QV}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	average points per student for variant V of question Q
$\bar{P}_{\mathrm{Q}}(Q)$	N_{Q}	average points per student for question Q
$\bar{P}$	1	average points per student for entire exam
$P_{\mathrm{Q}}^{\max}(QV)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	maximum points available for variant V of question Q
$P_{\mathrm{Q}}^{\max}(Q)$	N_{Q}	maximum points available for question Q
$P^{\max}$	1	maximum points available for the entire exam
$\hat{P}_{\mathrm{Q}}(Q)$	N_{Q}	normalized average points per student for question Q
$R_{\mathrm{QV}}(Q,V)$	$N_{\mathrm{Q}} \times N_{\mathrm{V}}$	ratio of average points for variant V to average for question Q
$r_{Q\hat{Q}}^{\mathrm{s}}(Q,\hat{Q})$	$N_{\mathrm{Q}} \times N_{\mathrm{Q}}$	correlation between student scores on questions Q and $\hat{Q}$
$r_{\mathrm{ss}}^{\hat{Q}}(s,\hat{s})$	$N_{\mathrm{s}} \times N_{\mathrm{s}}$	correlation between question points for students s and $\hat{s}$
$n_{\mathrm{ss}}^{\mathrm{ai}}(s,\hat{s})$	$N_{\mathrm{s}} \times N_{\mathrm{s}}$	number of mutually incorrect exam answers for students s and $\hat{s}$
$r_{\mathrm{ss}}^{\mathrm{ai}}(s,\hat{s})$	$N_{\mathrm{s}} \times N_{\mathrm{s}}$	proportion of identical incorrect answers for students s and $\hat{s}$
$D_{\mathrm{Q}}(Q)$	N_{Q}	"difficulty" of question Q (one minus normalized averaged points $\hat{P}_{\mathrm{Q}}(Q)$)
$r_{\mathrm{Q}}^{\mathrm{P}}(Q)$	N_{Q}	"discrimination" of question Q (correlation coefficient of $P_{\mathrm{s}}(s)$ and $P_{\mathrm{sQ}}(s,Q)$)

These quantities are computed by:

$$n_{\mathrm{e}}^{\mathrm{s}}(e) = \sum_{s=1}^{N_{\mathrm{s}}} \delta\big(K(e), k(s)\big) \tag{8}$$

$$n_{\mathrm{QV}}^{\mathrm{s}}(\hat{Q}, \hat{V}) = \sum_{s=1}^{N_{\mathrm{s}}} \sum_{q=1}^{N_{\mathrm{Q}}} \delta\Big(\hat{Q}, Q\big(e(s),q\big)\Big) \; \delta\Big(\hat{V}, V\big(e(s),q\big)\Big) \tag{9}$$

$$n_{\mathrm{QVA}}^{\mathrm{a}}(\hat{Q}, \hat{V}, \hat{A}) = \sum_{s=1}^{N_{\mathrm{s}}} \sum_{q=1}^{N_{\mathrm{Q}}} \delta\Big(\hat{Q}, Q\big(e(s),q\big)\Big) \; \delta\Big(\hat{V}, V\big(e(s),q\big)\Big) \; \delta\Big(\hat{A}, A\big(e(s),q,a(s,q)\big)\Big) \tag{10}$$

$$n_{\mathrm{QV}}^{\mathrm{a}}(Q,V) = \sum_{A=1}^{N_{\mathrm{A}}} n_{\mathrm{QVA}}^{\mathrm{a}}(Q,V,A) \tag{11}$$

$$n_{\mathrm{QV}}^{\mathrm{na}}(Q,V) = n_{\mathrm{QV}}^{\mathrm{s}}(Q,V) - n_{\mathrm{QV}}^{\mathrm{a}}(Q,V) \tag{12}$$

$$n_{\mathrm{Q}}^{\mathrm{a}}(Q) = \sum_{V=1}^{N_{\mathrm{V}}} n_{\mathrm{QV}}^{\mathrm{a}}(Q,V) \tag{13}$$

$$r_{\mathrm{QVA}}^{\mathrm{a}}(Q,V,A) = \frac{n_{\mathrm{QVA}}^{\mathrm{a}}(Q,V,A)}{n_{\mathrm{QV}}^{\mathrm{s}}(Q,V)} \tag{14}$$

$$r_{\mathrm{QV}}^{\mathrm{na}}(Q,V) = \frac{n_{\mathrm{QV}}^{\mathrm{na}}(Q,V)}{n_{\mathrm{QV}}^{\mathrm{s}}(Q,V)} \tag{15}$$

$$P_{\mathrm{QV}}(\hat{Q},\hat{V}) = \sum_{s=1}^{N_{\mathrm{s}}} \sum_{q=1}^{N_{\mathrm{Q}}} P_{\mathrm{sq}}(s,q)\, \delta\big(\hat{Q}, Q\big(e(s),q\big)\big)\, \delta\big(\hat{V}, V\big(e(s),q\big)\big) \tag{16}$$

$$P_{\mathrm{Q}}(Q) = \sum_{V=1}^{N_{\mathrm{V}}} P_{\mathrm{QV}}(Q,V) \tag{17}$$

$$\bar{P}_{\mathrm{QV}}(Q,V) = \frac{P_{\mathrm{QV}}(Q,V)}{n_{\mathrm{QV}}^{\mathrm{s}}(Q,V)} \tag{18}$$

$$\bar{P}_{\mathrm{Q}}(Q) = \frac{P_{\mathrm{Q}}(Q)}{N_{\mathrm{s}}} \tag{19}$$

$$\bar{P} = \sum_{Q=1}^{N_{\mathrm{Q}}} \bar{P}_{\mathrm{Q}}(Q) = \frac{1}{N_{\mathrm{s}}} \sum_{s=1}^{N_{\mathrm{s}}} P_{\mathrm{s}}(s) \tag{20}$$

$$P_{\mathrm{QV}}^{\mathrm{max}}(Q,V) = \max_{A \in \{1,\ldots,N_{\mathrm{A}}\}} P(Q,V,A) \tag{21}$$

$$P_{\mathrm{Q}}^{\mathrm{max}}(Q) = \max_{V \in \{1,\ldots,N_{\mathrm{V}}\}} P_{\mathrm{QV}}^{\mathrm{max}}(Q,V) \tag{22}$$

$$P^{\mathrm{max}} = \sum_{Q=1}^{N_{\mathrm{Q}}} P_{\mathrm{Q}}^{\mathrm{max}}(Q) \tag{23}$$

$$\hat{P}_{\mathrm{Q}}(Q) = \frac{\bar{P}_{\mathrm{Q}}(Q)}{P_{Q}^{\mathrm{max}}(Q)} \tag{24}$$

$$R_{\mathrm{QV}}(Q,V) = \frac{\bar{P}_{\mathrm{QV}}(Q,V)}{\bar{P}_{\mathrm{Q}}(Q)} \tag{25}$$

$$r_{\mathrm{QQ}}^{\mathrm{s}}(Q,\hat{Q}) = \frac{\sum_{s=1}^{N_{\mathrm{s}}} (P_{sQ}(s,Q) - \bar{P}_Q(Q))(P_{sQ}(s,\hat{Q}) - \bar{P}_Q(\hat{Q}))}{\sqrt{\sum_{s=1}^{N_{\mathrm{s}}} (P_{sQ}(s,Q) - \bar{P}_Q(Q))^2} \sqrt{\sum_{s=1}^{N_{\mathrm{s}}} (P_{sQ}(s,\hat{Q}) - \bar{P}_Q(\hat{Q}))^2}} \tag{26}$$

$$r_{\mathrm{s}}^{\mathrm{Q}}(s,\hat{s}) = \frac{\sum_{Q=1}^{N_{\mathrm{Q}}} (P_{sQ}(s,Q) - \bar{P}_Q(Q))(P_{sQ}(s,\hat{Q}) - \bar{P}_Q(\hat{Q}))}{\sqrt{\sum_{Q=1}^{N_{\mathrm{Q}}} (P_{sQ}(s,Q) - \bar{P}_Q(Q))^2} \sqrt{\sum_{Q=1}^{N_{\mathrm{Q}}} (P_{sQ}(s,\hat{Q}) - \bar{P}_Q(\hat{Q}))^2}} \tag{27}$$

$$n_{\mathrm{ss}}^{\mathrm{ai}}(s,\hat{s}) = \sum_{q=1}^{N_{\mathrm{Q}}} \delta\big(0, P_{\mathrm{sq}}(s,q)\big)\, \delta\big(0, P_{\mathrm{sq}}(\hat{s},q)\big) \tag{28}$$

$$r_{\mathrm{ss}}^{\mathrm{ai}}(s,\hat{s}) = \frac{1}{n_{\mathrm{ss}}^{\mathrm{ai}}(s,\hat{s})} \sum_{q=1}^{N_{\mathrm{Q}}} \delta\big(a(s,q), a(\hat{s},q)\big)\, \delta\big(0, P_{\mathrm{sq}}(s,q)\big)\, \delta\big(0, P_{\mathrm{sq}}(\hat{s},q)\big) \tag{29}$$

$$D_{\mathrm{Q}}(Q) = 1 - \hat{P}_{\mathrm{Q}}(Q) \tag{30}$$

$$r_{\mathrm{Q}}^{\mathrm{P}}(Q) = \frac{\sum_{s=1}^{N_{\mathrm{s}}}(P_{sQ}(s,Q) - \bar{P}_Q(Q))(P_s(s) - \bar{P})}{\sqrt{\sum_{s=1}^{N_{\mathrm{s}}}(P_{sQ}(s,Q) - \bar{P}_Q(Q))^2}\sqrt{\sum_{s=1}^{N_{\mathrm{s}}}(P_s(s) - \bar{P})^2}} \tag{31}$$

where $\delta(x,y)$ is the Kronecker delta function, equal to one with equal arguments and zero otherwise.

We expect that $R_{\mathrm{QV}}(Q,V) \approx 1$, indicating that all variants were equally difficult for question Q. We also expect that $r_{\mathrm{ss}}^{\mathrm{ai}}(s,\hat{s}) \approx 1/N_{\mathrm{A}}$ for $e(s) = e(\hat{s})$, indicating only random association between incorrect answers given by students with identical exams, although this may be legitimately higher due to correlations between student mistakes.

To understand how students of different ability levels perform it is helpful to group students by their exam scores. We consider N_{g} almost-equal-sized *groups* of students, where all students in group g perform worse than all students group $g+1$, etc. We can then compute the following per-group quantities:

array	dimensions	description
$s(r)$	N_{s}	student number at rank r
$g(s)$	N_{g}	group number of student s
$n_{\mathrm{g}}^{\mathrm{s}}(g)$	N_{g}	number of students in group g
$P_{\mathrm{gQV}}(g,Q,V)$	$N_{\mathrm{g}} \times N_{\mathrm{Q}} \times N_{\mathrm{V}}$	total points awarded for variant V of question Q to students in group g
$P_{\mathrm{gQ}}(g,Q)$	$N_{\mathrm{g}} \times N_{\mathrm{Q}}$	total points awarded for question Q to students in group g
$\bar{P}_{\mathrm{gQ}}(g,Q)$	$N_{\mathrm{g}} \times N_{\mathrm{Q}}$	average points awarded for question Q to students in group g
$\hat{P}_{\mathrm{gQ}}(g,Q)$	$N_{\mathrm{g}} \times N_{\mathrm{Q}}$	normalized average points awarded for question Q to students in group g

The definitions for the above quantities are:

$$P_{\mathrm{s}}\big(s(r)\big) \le P_{\mathrm{s}}\big(s(r+1)\big) \tag{32}$$

$$g\big(s(r)\big) = \left\lfloor \frac{(r-1)\,N_{\mathrm{g}}}{N_{\mathrm{s}}} \right\rfloor + 1 \tag{33}$$

$$n_{\mathrm{g}}^{\mathrm{s}}(\hat{g}) = \sum_{s=1}^{N_{\mathrm{s}}} \delta\big(\hat{g}, g(s)\big) \tag{34}$$

$$P_{\mathrm{gQV}}(\hat{g},\hat{Q},\hat{V}) = \sum_{s=1}^{N_{\mathrm{s}}}\sum_{q=1}^{N_{\mathrm{Q}}} P_{\mathrm{sq}}(s,q)\,\delta\big(\hat{g},g(s)\big)\,\delta\Big(\hat{Q},Q\big(e(s),q\big)\Big)\,\delta\Big(\hat{V},V\big(e(s),q\big)\Big) \tag{35}$$

$$P_{\mathrm{gQ}}(g,Q) = \sum_{V=1}^{N_{\mathrm{V}}} P_{\mathrm{gQV}}(g,Q,V) \tag{36}$$

$$\bar{P}_{\mathrm{gQ}}(g,Q) = \frac{P_{\mathrm{gQ}}(g,Q)}{n_{\mathrm{g}}^{\mathrm{s}}(g)} \tag{37}$$

$$\hat{P}_{\mathrm{gQ}}(g,Q) = \frac{\bar{P}_{\mathrm{gQ}}(g,Q)}{P_{\mathrm{Q}}^{\max}(Q)} \tag{38}$$

6 Implementation notes

The randomized exam procedure described in this document is implemented in the `randexam` Python script. The configuration section at the top of the file should be edited as appropriate.

The internal variables in the script follow the notation in this document and use `numpy` for array handling. The only significant departure is that all variables are numbered from zero rather than from

one, as Python uses zero-indexed arrays. The conversion between one-based and zero-based quantities occurs during reading and writing. All internal variables and computations are zero-based.

In this document overloaded notation is used for index variables and arrays, so that we write both $Q(e, q)$ and $C(Q, V)$, where Q is understood to be a dummy variable in the latter expression. In the code, plain variable names like `Q` refer to the array and names like `Qi` and `Qj` are used for dummy index variables.

All variables that represent answers are string-valued and store values A, B, etc, rather than an integer. These correspond to indexes A $\leftrightarrow$ 0, B $\leftrightarrow$ 1, etc. To convert between zero-based integer indexes and string representations, the functions `chr2ind()` and `ind2chr()` can be used. For example, to access a value in the $A(e, q, a)$ array we can write:

```
e = 7
q = 0
a = "D"
print(A(e, q, chr2ind(a)))
```

7 Key generation

The exam keys are generated by encoding the exam number e in base-N_A using N_D digits and then appending two or three checksum digits. The exam-number digits are in little-endian order, with least significant digit first, so for zero-based digits K_i and zero-based indexes i they are

$$K_i = \left\lfloor \frac{e}{(N_A)^i} \right\rfloor \mod N_A, \text{ for } i = 0, \dots, (N_D - 1). \tag{39}$$

The checksum digits are the *parity digit* K_P, a *Fletcher-type* checksum K_F, and possibly a modified Fletcher-type checksum K_M, defined for zero-based digits K_i and zero-based indexes i by

$$K_P = \sum_{i=0}^{N_D-1} K_i \mod N_A \tag{40}$$

$$K_F = \sum_{i=0}^{N_D-1} w_{F,i} K_i \mod N_A \qquad w_{F,i} = \left(i \mod (N_A - 1)\right) + 1 \tag{41}$$

$$K_M = \sum_{i=0}^{N_D-1} w_{M,i} K_i \mod N_A \qquad w_{M,i} = \left(\left\lfloor \frac{i}{N_A - 1} \right\rfloor \mod (N_A - 1)\right) + 1. \tag{42}$$

Using the first two checksum digits ensures a minimum Hamming distance of 3 for at most $(N_A - 1)$ non-checksum digits (up to $N_A^{(N_A-1)}$ different random exams). If more digits are required to encode the exam numbers then the third checksum digit is appended to the key, ensuring a Hamming distance of at least 3 for up to $(N_A - 1)^2$ non-checksum digits, allowing up to $N_A^{(N_A-1)^2}$ different random exams. In the case of $N_A = 5$ the key lengths are:

number of exams	exam-number digits N_D	checksum digits	key length N_K
5	1	2	3
25	2	2	4
125	3	2	5
625	4	2	6
3125	5	3	8
15625	6	3	9

8 Recommended procedures

To prepare a "Final" exam, we assume that we are working within a `exam_final` directory.

8.1 Preparing the exams

- Make a directory called `1_exams` and copy in `randexam` and `library.tex`.

- Rename `library.tex` to `final_library.tex` and set the `PREFIX` in `randexam` to `final_`.

- Edit `library.tex` to add questions and make the coversheet.

- Run `./randexam proc-lib`.

8.2 Administering the exams

- Collect both the Scantron sheets and the exam papers (with names on them) from the students, so that exam papers can be later matched to students if needed.

8.3 Grading the exams

- Make a subdirectory called `2_scantrons` containing a raw unedited copy of the Scantron data files.

- Make a subdirectory called `3_grading` and copy in the contents of both `1_exams` and `2_-scantrons`. All editing should be done in this third directory.

- Run `./randexam proc-scan` and clean up errors in the file as needed.

- Run `./randexam proc-ans`.