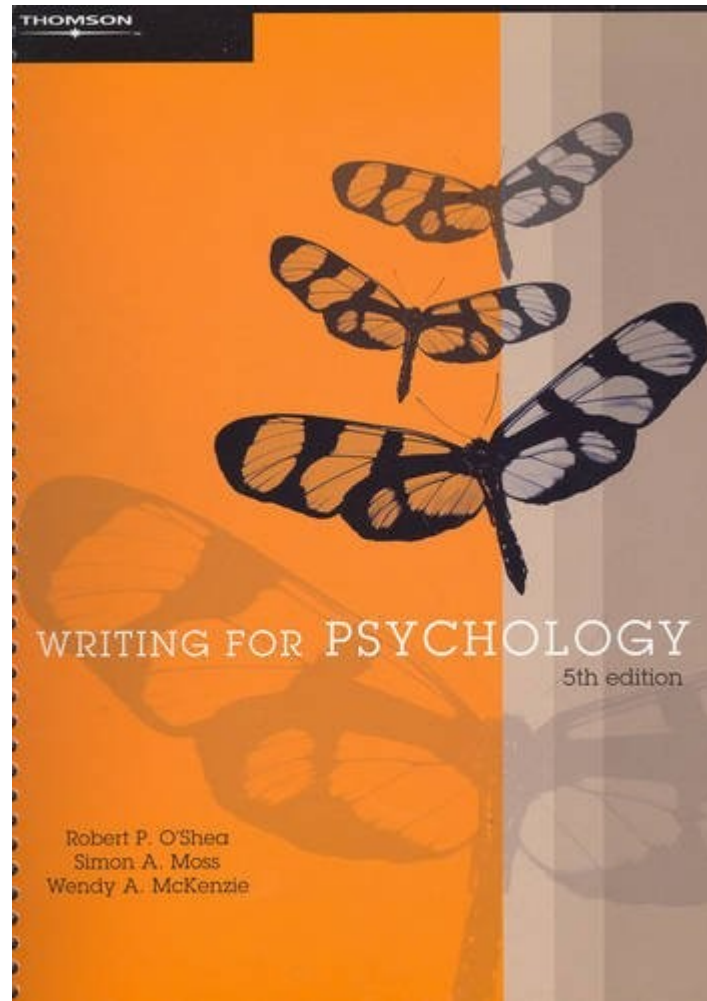# Passing Multichoice Exams Without Studying

Andrew Nisbet

STAT 462 Presentation

O'Shea, R., Moss, S., & McKenzie, W. (2006). *Writing for psychology* (5th edition). Thomson Nelson.

Passing Multichoice Exams Without Studying

1. Finding a dataset

2. Data wrangling

3. Model

4. Results

1 Finding a dataset

Requirements:

- Digital format

- Representative

- Scructured format

1 Finding a dataset

Australian Science Olympiads

1   Finding a dataset

Sample Question

> **26.** What is the primary function of large leaves found on seedlings growing on the forest floor?
>
>   a.   Provision of shade for their root systems.
>   b.   Elimination of excess water that is entering via the roots.
>   c.   To allow for leaf damage by insects.
>   d.   Acquisition of as much light as possible for photosynthesis.

1 Finding a dataset

✓ Digital format

    – PDF format, with solutions

    – 19 exams

    – 402 questions

    – 1800 answers

1 Finding a dataset

✓ Representative

- Six years (2007 - 2012)
- Three subjects (Biology, Chemistry, Physics)
- Eight authors

```
<?xpacket end="w"?>\nendstream\nendobj\n94 0 obj\n<</
CreationDate(D:20070723120350+10'00')/Author(Mary Oli
ver)/Creator(Acrobat PDFMaker 8.1 for Word)/Producer(
Acrobat Distiller 8.1.0 \(Windows\))/ModDate(D:200707
23120431+10'00')/SourceModified(D:20070723020324)/Tit
le()>>\nendobj
```

1   Finding a dataset

✗ Structured

   – PDF is a presentation focussed format

2   Data wrangling

Text Extraction

$$PDF \longrightarrow text$$

```
a.\tProvision\sof\sshade\sfor\sroots \n
```

```
Provision\sof\sshade\sfor\sroots <left>a.\t</left> \n
```

2 Data wrangling
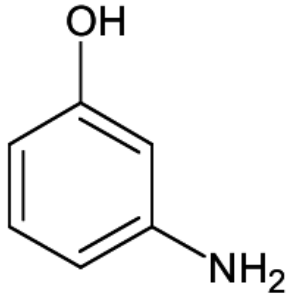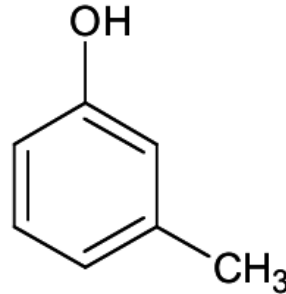
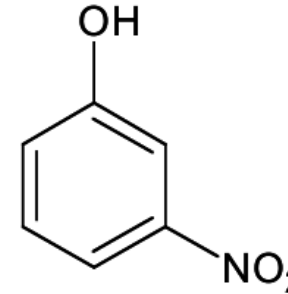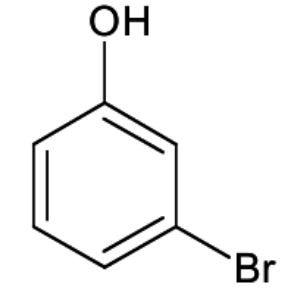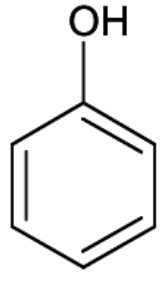Optical Character Recognition (OCR)

$$PDF \longrightarrow image \longrightarrow text$$

2 Data wrangling

Unavoidable Errors

Q10 Ethanoic acid was heated with an unknown compound **X** and a catalytic amount of concentrated sulfuric acid. The reaction mixture was shown by mass spectrometry to contain a compound of molar mass 193 g mol^{-1}.

Which of the following could be compound **X**?

OH ⬡ NH$_2$	OH ⬡ CH$_3$	OH ⬡ NO$_2$	OH ⬡ Br	OH ⬡
A	**B**	**C**	**D**	**E**

2 Data wrangling

Avoidable Errors

```
It lowers the activation energy of the reaction.
It lowers the amount of heat gained by the system.
It lowers the amount of heat released by the system.
It lowers the potential energy of the products.
It lowers the potential energy of the reactants.
a.
b.
c.
d.
e.
```

2   Data wrangling

80% of questions were successfully parsed

Text extraction	49%
OCR	32%
Sucessfully parsed	**80%**

Unavoidable failure	10%
Avoidable failure	10%
Unsucessful	**20%**

2 Data wrangling

The parsing failure rate was biased to certain subjects. This affects the final ratios of the subjects in the dataset:

Biology	66%
Chemistry	22%
Physics	12%

2   Data wrangling

Feature extraction

- Numbers (`number_of_words`)

- Booleans (`is_inverse_logic`)

  - Normalised to $(0, 1)$ or $(-1, 1)$
  - Replace `NaN`s with mean values
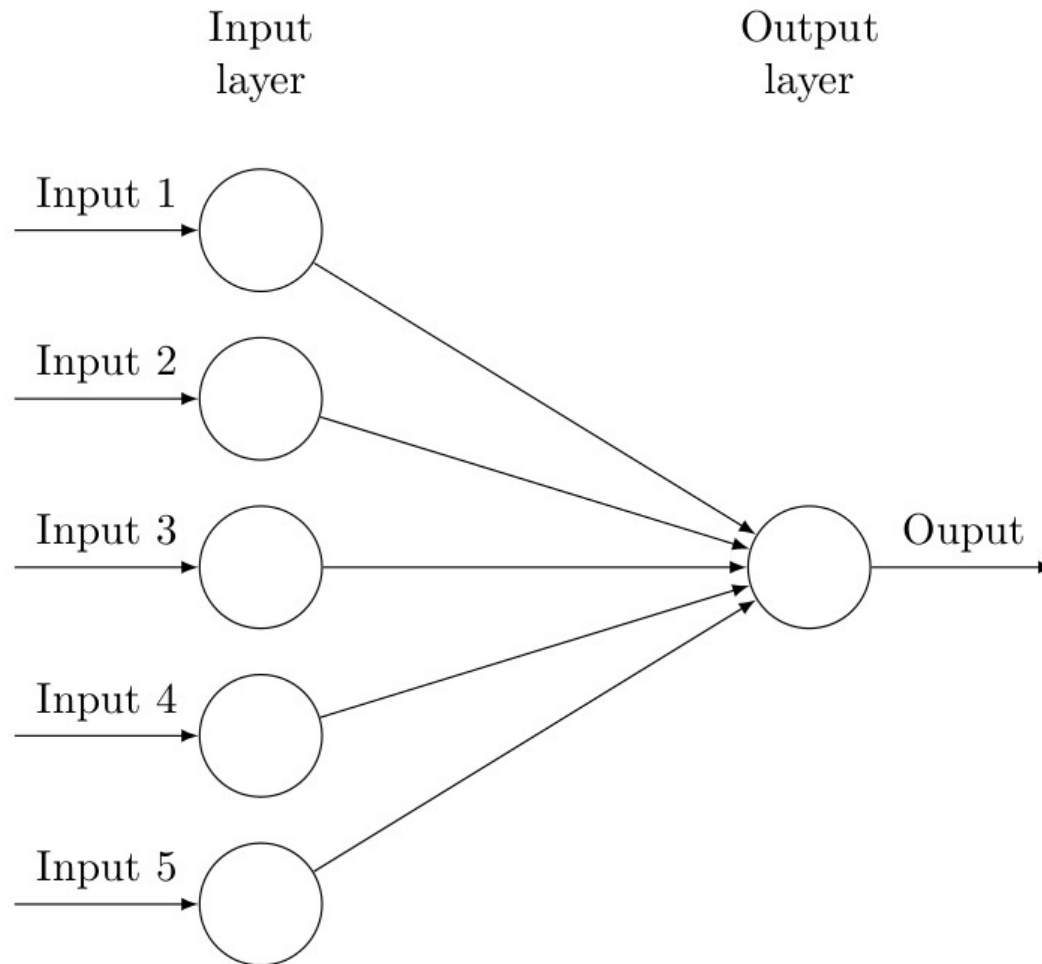
2 Data wrangling

Final dataset

- 1200 observations

- 32 features

- $y \equiv$ whether answer is correct

3   Modelling

Linear regression showed no significant correlation for any one parameter. So no rule of thumb model!
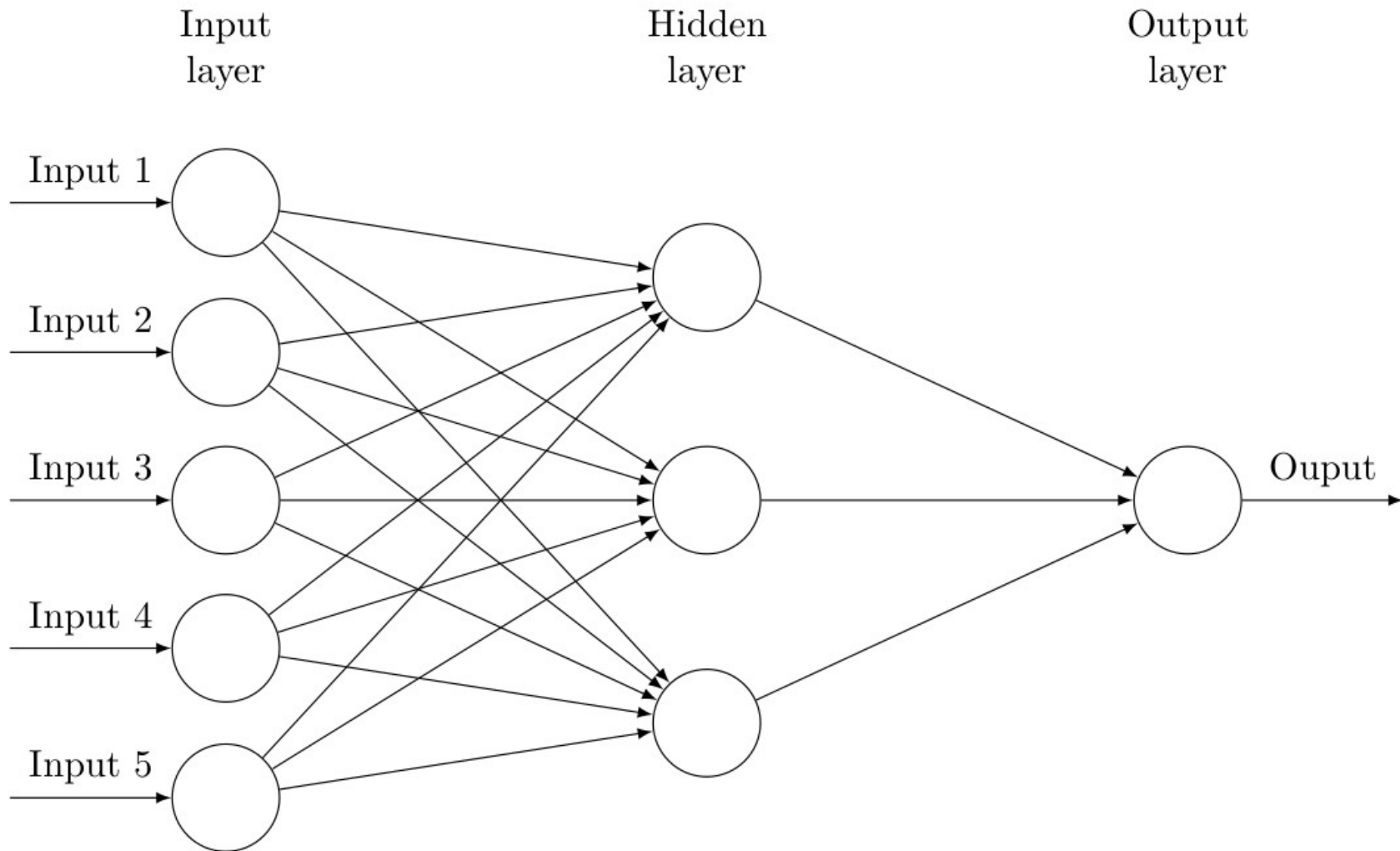
3   Modelling

Neural networks
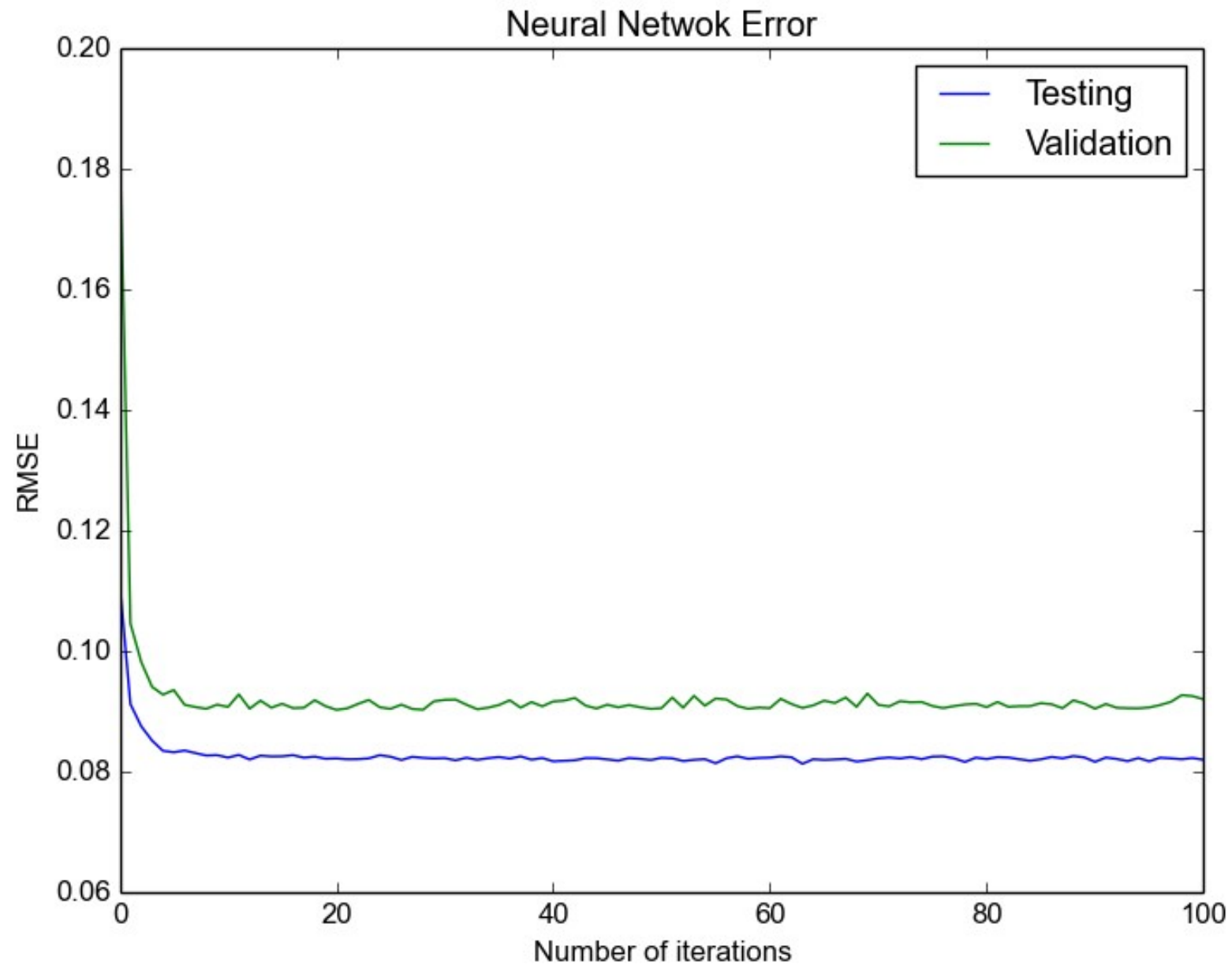
3   Modelling

Neural networks

3 Modelling

Validation

- The multichoice dataset was split into training and testing subsets.

- The neural network was run for a number of iterations.

3 Modelling

4 Results

Objective function

- How many questions can be correctly predicted?

- Compare to chance (21%)

- Network was modelled 10 times, and averaged

4 Results

Random chance: 21%

Model performance: 26%

Is this good?

4   Results

Model performance: 26%

- Model is 5 percentage-points better.

- So if you guessed 20 questions, you would get one more correct with the neural network model, compared to randomly guessing.

- Probably not an improvement on studying.

- Also not an improvement on actual cheating.

4 Results

Model performance: 26%

- Results seem significant, despite small effect size.

- None of the validation errors produced by the different test/training splits were lower than chance.

- There's probably an underlying pattern, but too complicated to extract from neural network.

4   Results

Suggestion: loose the representativeness

- Pull the last few years of exams for a single course from the library website.

- All written by the same lecturer.

- No university-wide multichoice test writing guide.

- Look for individual biases, rather than overall ones.

https://bitbucket.org/ajnisbet/multichoice-modelling