Increasing capability in assessment and self-regulation through multiple choice marking of exemplars

Shane Lindsay – Lecturer in Psychology

Ellie Hemmingway - BSc Student

Stuart McGugan - Academic Practice Advisor

shane.lindsay@gmail.com, S.McGugan@hull.ac.uk, E.F.Hemingway@2015.hull.ac.uk

Why?

Problem 1: Getting students to write better lab/research reports

Problem 2: Improving student understanding of assessment and use of feedback

… and NSS scores

Students must first possess the ability to make a comparison between their own performance and an expected standard *Sadler, 2010, 2013*

Reliant on ability to self-assess and understand quality

Developed from **doing** not **telling**

How?

Pilot of approach

Research Skills 2 – BSc Psychology @

Year 1 - Semester 2

10 weeks

1 x 2 hour lecture ~200 people

1 x 2 hour computer workshop ~20 people

40% Exam, 40% Research Report, 20% fortnightly tests

Previously…

Single two hour lecture on “how to write a lab report”:

- Title
- Abstract
- Introduction
- Method (participants, materials, procedure)
- Results
- Discussion
- References
- Appendix

Assessment criteria for Research Reports

	1st – Exemplary/Outstanding/Excellent 95-85-75%	2.1 – Very good to Good 68-65-62%	2.2 - Satisfactory 58-55-52%	3rd – Adequate but Weak 48-45-42%	Marginal Fail - Unsatisfactory 35%	Fail – Very Unsatisfactory 25-10-0%
Title	**Excellent:** A nice title, though may be somewhat uninspiring. Some word choices could be improved. **Outstanding:** No obvious way to improve. Interesting enough to encourages reader to read report. **Exemplary:** Fulfils function of a title perfectly yet also eye catching, memorable and/or highly intriguing. Near impossible to improve upon.	Not too vague, too short or too long, identifies the IV and DV in a natural way. Encapsulates the aim of the study well. Could be relatively pedestrian with little to make it stand out.	IV and DV is clear, but clear room for improvement, in conciseness or phrasing.	Too long, or too short. Does not identify the IV and/or DV. Contains superfluous phrases, such as "A psychological investigation into.." or "An experiment on...". Uses terminology that a naïve reader would struggle to understand without reading the study.	Very generic (e.g. "Lab report 1", "Free Recall Study").	No title or makes little sense.
Abstract	**Excellent:** Covers all the main areas well. Some lack of clarity at times, with question marks remaining. Could be made shorter without loss of clarity. **Outstanding:** Lacking flair, but very concise, with nothing lacking. Only minor improvements could be made. **Exemplary**: Extremely succinct, with not a word wasted, yet still exceptionally clear.	Clear and succinct (120-200 words) summary of the aims, methods, results and conclusions of the study. Includes all the necessary information, and is well written. There may be an area which lacks detail or has some inappropriate level of detail.	Written correctly, containing most of the key information (e.g. motivation/background, method, results, implications) but has missed out a key detail (e.g. the implications of the results).	Covers most of the important information but missing key information. Hard to follow at times, may rely on additional information for comprehension. Unnecessary details (e.g. ages of participants, p values)	Far too long or too short. Fails completely to adequately and accurately summarise the study, but some details are given that are expected from an abstract.	No abstract included, incomprehensible or a couple of short sentences. Bearing little relation to the study. Little understanding shown of what an abstract should include.
Introduction	**Excellent:** Clearly written, well structured, with evidence of relevant extra reading. Flows well. Identifies the main aims, and ends with a clear outline of the study's hypotheses. At times novel. Convincing rationale for performing the study. Little that is not relevant. **Outstanding:** Literature cited highly relevant, including recent work where appropriate. Demonstrates initiative in considerable extra reading. **Exemplary:** Exceptional understanding of theory and current issues. Compellingly motivated. Not a sentence wasted.	Some evidence of extra reading and a reasonable attempt to formulate an argument, but the structure could be clearer, with a logical argument more apparent. It could be well written, but insufficient discussion of the literature (e.g. generic).	A basic attempt with elements missing, such as limited or no justification for the study, little evidence from the literature, poor structure, no hypotheses or hypotheses very unclear. May be too short or excessively long. Mostly relevant information, though lacking insight into wider issues.	Very short and skimpy, with little or no attempt to include any references to relevant theoretical and empirical work beyond the most basic reading. Argument may be illogical or consistent argument.	Little attempt to address research question. Very short, such as a single paragraph. Hard to understand. Little relevant information.	Section missing completely or little effort shown or makes little sense. Very jumbled and hard to follow. What would be expected by a lay person with no understanding of psychology.
Method and Design	**Excellent:** Little or no repetition, very concise and high level of detail provided where appropriate. Some degree of originality and invention in the design. Potentially information lacking that impacts replication and interpretation. **Outstanding:** Innovation and much initiative shown. Insight and understanding into key design issues. Design may be interesting, though not ground breaking. **Exemplary:** Design choices expertly motivated. Clear originality and independence of thought. Theoretically interesting enough to be published.	**Method:** Contains all or nearly all of the relevant information needed. Clearly and systematically described in such a way that a naive reader could not struggle to replicate the study from this description. Correctly describes the formal design of the study, including an accurate specification of the IV(s) and DV(s) used. **Design:** Sound, and addresses research question well. Some evidence of originality or novelty. Ethical issues well addressed.	**Method:** Good structure, some important relevant information omitted, occasional ambiguity, some unnecessary repetition. Procedure may be verbose. **Design**: Sensible attempt to address research question, but little originality, and of limited theoretical interest. Methodological flaws that could have been easily avoided. Ethical issues raised but lacking thoroughness.	**Method:** One or more sub-sections are missing, confused, information often included under the wrong sub-section, sometimes hard to follow, key information omitted, excessive irrelevant details, materials or examples absent. Much repetition. **Design:** Followed task but no evidence of originality. Numerous flaws. Clear conclusions may be difficult. More insight into ethical issues needed.	**Method:** Sections jumbled, or incoherent. Very hard to understand what happened in study and replication would be impossible. **Design:** Has serious flaws, such as major confounding variables or inability to establish causality in an experiment. Possible ethical concerns.	**Method:** Section missing completely or a token effort, sub-sections absent, incomplete or in note form. **Design:** No study undertaken or not informed by basic psychological research principles or application of a scientific methodology. Ethical standards not maintained.
Results	**Excellent:** Well organised, and competent throughout, though minor errors may be acceptable. Very clear link between design and results. Complicated analysis made comprehensible in a simple way. **Outstanding:** Analysis tools employed in a sophisticated way. Very high quality use of tables and	Logical and clear presentation of relevant data, such as descriptive and inferential statistical results. Clear, well-labelled figures and tables, with accompanying written description of what they show, in the context of the study. Possible to see how results	No major errors. Limitations in linking design with analysis. Measures of variability (e.g. standard deviations) from tables or graphs, figures/tables labelled incorrectly or with lack of clarity (e.g. titles too short or to long).	Lack of understanding results and/or data analysis methods shown, with basic errors in interpretation (e.g. $p < .05$ described as non-significant). No graphs, or graphs are in the wrong place (e.g. in an appendix), or	Incomplete or perfunctory attempt (e.g. only descriptives). Has graphs or tables, but without any accompanying written explanation. It may have some writing, but no tables or graphs. Very incoherent.	Section missing completely or bears little resemblance to what is expected from a scientific report.

	figures, little or no errors, though several improvements possible. **Exemplary**: Publication quality graphics. Faultless presentation of key data. Suggestions for improvements would only be personal preferences.	Inform research question and link to design, though this could be more direct. Some trivial errors. Minor deviations from APA style. Lacking precision in describing the IV and DV. Units of measurement not always described.	Does show some understanding of main findings, and has presented the information in a logical format. Mostly follows standard presentation of results using APA style, though with some errors.	graphs irrelevant (e.g. raw subject data rather than means) or hard to follow (e.g. no axes). Only skimpy or inaccurate explanations supplied. Inappropriate details, (e.g. raw data in the results section; excessive decimal places).		
Discussion	**Excellent:** Not as focused as for higher grades, but largely avoids irrelevant material. Good understanding of implications and non-trivial critique of their own study, along with non-trivial suggestions for improvements and future research. **Outstanding:** Excellent ability to constructively criticise research study. Very little tangents or that is off topic. Insight into both methodological empirical and theoretical issues that would motivate future research. **Exemplary:** Comprehensive understanding of theoretical issues with originality and flair Insightful grasp of theoretical implications. Clever and worthwhile suggestions for future research. Evidence of extensive reading.	Clear summary of main results, followed by a mostly successful attempt to relate the findings to relevant previous theoretical and empirical research. Generally an intelligent evaluation of the strengths, weaknesses and limitations of the study that was performed, and reasonably sensible suggestions for possible improvements and extensions. However, some points may be trivial and lacking deeper theoretical insight. Use of literature limited.	Poorly structured, but contains the essential elements. OR the structure is good, but elements are missing. An attempt to discuss limitations but may be generic (e.g. results don't generalise to population due to small sample). Possible implications mentioned but not spelled out or generic (e.g. more participants needed as small sample). References to other literature mainly confined to secondary sources.	Poor structure, things in the wrong order, shows little understanding of what the study was about, what the results mean, or how they relate to previous work. Basic conceptual errors. Attempts shown to try and describe limitations but confused or irrelevant. Doesn't advance upon introduction, with no additional discussion of theory or other relevant data.	Failure to demonstrate an understanding of aim of study. Near complete lack of understanding on what conclusions can be drawn from data analysis. Lack of theoretical insight. Very short and/or disorganised.	Absent or very skimpy (e.g. one paragraph) with no or a very poor attempt to relate results to relevant theoretical and empirical research.
Appendices	Very few or no errors. Where appropriate, goes beyond the task requirements. Helpful for replication or for understanding design by allowing examination of properties of stimuli or instructions. Very well supported by the main text.	Used when appropriate or required. Labelled using APA style, though errors possible. May be over zealous leading to some that are not needed.	Appendices are used appropriately. Maybe incomplete, or some that are unnecessary. Labels used though limited use of APA style.	Included where appropriate, but major errors, such as excessively long (e.g. long irrelevant print outs).	Unlabelled. Hand written. Not referred to in the main text.	No appendices when they have been asked for, or impossible to make sense of.
Citing and Referencing	**Excellent:** Correct use of APA style (6th ed.). All references in list cited in text, all citations in text in reference list. Some minor errors. **Outstanding:** Only very trivial errors. **Exemplary**: Perfect use of APA style without any errors in citations and in references.	Correct use of APA style throughout, though with some errors in citing and/or referencing.	Shows an understanding of citations and referencing, though some errors, such as in formatting, citations not in reference list, references not cited.	Referencing used but inconsistent or poorly formatted. References to texts not discussed in main text (e.g. statistics textbooks). Some citations without references.	An attempt at referencing but little or no accordance with APA style. Haphazard.	Citations are almost or entirely absent and awareness of mechanics of scholarship is very weak.
Formatting	**Excellent:** Some minor errors in formatting, but on the whole very close adherence to conventions **Outstanding:** Very few errors (and minor ones) in adherence of report writing conventions of APA style. **Exemplary:** Flawless use of APA style and professional standards throughout.	Correct use of APA style (6th ed.) for the large part, though with errors. Consistency in formatting throughout.	A reasonable attempt to use APA style, though a fair number of errors, with much room for improvement in future. More attention to detail would help.	Poor attention to detail though some effort shown to conform to standard APA style. Misunderstanding of some key aspects (e.g. referring to tables in text; labelling of figures/tables).	Not conforming to APA style. Inconsistency and scrappiness of presentation. Inappropriate use of colour. Inappropriate use of spacing (e.g. huge margins); odd formatting choices (e.g. centred throughout). Figures and Tables not labelled.	No use of APA style. Amateurish in presentation. Clear lack of care taken in presentation.
Writing	**Excellent:** Very clearly written, though lacks flair. Most paragraphs have clear purposes and logical flow can be identified. Transitions between paragraphs generally good but could be improved. **Outstanding:** Signs of an accomplished writer. Highly competent in expression, though sometimes sentences could be paraphrased to be more elegant and concise. **Exemplary:** Exceptional clarity and economy of expression. Ideas conveyed always clear.	Writing style on the whole is clear, concise, and logical. Sentences generally easy to understand, Paragraphs well constructed, but better transitions between paragraphs often needed. Sometimes paragraphs lack focus.	Writing style is understandable, though at times may be clumsy. Understanding of principles of paragraphs, though some not linked with each other.	Basic but poor standard of written communication. Often lacking in clarity. Many corrections necessary indicating much more proof reading was needed. A few sentences might not make sense. Ideas can be poorly expressed. Relatively dull to read. Main ideas do come across.	Significant flaws in spelling, grammar and composition which undermine the clarity of meaning. Sentences often jumbled, incoherent or ambiguous. Paragraphs too short or too long. Little or no linking between paragraphs.	Very hard to read, well below the standard of communication expected at University level. Too many mistakes to annotate. Few sentences make sense. No, or very limited, use of paragraphs.

Previously…

Assessment criteria given at end of an assignment brief

Componential weekly focus

Lecture

Section of weekly lecture with instruction on report section with examples of best practice

Workshop

Complete usual lab class worksheet finishing with dedicated section for report section exercise

Multiple Choice Exercise

Use of work from previous years (via TurnItIn)

~5 examples for each report section

Spread of quality standards

Implemented with 

Question 1

A paired samples t-test was performed on the data from the experiment. It revealed that the difference in mean reaction time between 30 degrees and 150 degrees was significant, $t(106) = -8.757(106)$, $p < .001$. So the null hypothesis which is there is no difference can be rejected. As can be seen from Table 1, the mean reaction time for the 150 degrees angular difference was 3.9330 seconds, whereas for the 30 degrees angular difference the mean reaction time was 2.8039 seconds.

- 1st - Well organised, and competent throughout. Very clear link between design and results. Complicated analysis made comprehensible in a simple way. High quality use of tables and figures, little or no errors.
- 2.1 - Logical and clear presentation of relevant data, such as descriptive and inferential statistical results. Clear, well-labelled figures and tables, with accompanying written description of what they show, in the context of the study. Possible to see how results inform research question and link to design, though this could be more direct. Some trivial errors. Minor deviations from APA style. Lacking precision in describing the IV and DV. Units of measurement not always described.
- 2.2 - No major errors. Limitations in linking design with analysis. Measures of variability (e.g. standard deviations) from tables or graphs, figures/tables labelled incorrectly or with lack of clarity (e.g. titles too short or to long). Does show some understanding of main findings, and has presented the information in a logical format. Mostly follows standard presentation of results using APA style, though with some errors.
- 3rd - Lack of understanding results and/or data analysis methods shown, with basic errors in interpretation (e.g. $p < .05$ described as non-significant). No graphs, or graphs are in the wrong place (e.g. in an appendix), or graphs irrelevant (e.g. raw subject data rather than means) or hard to follow (e.g. no axes). Only skimpy or inaccurate explanations supplied. Inappropriate details, (e.g. raw data in the results section; excessive decimal places).
- Marginal Fail - Incomplete or perfunctory attempt (e.g. only descriptives). Has graphs or tables, but without any accompanying written explanation. It may have some writing, but no tables or graphs. Very incoherent.
- Fail - Section missing completely or bears little resemblance to what is expected from a scientific report.

Quantitative evaluation

Initial mean congruence = 43% (*SD* = 16%)

1st	63%
2.1	30%
2.2	56%
3rd	43%

Qualitative evaluation

Student freeform text response on their learning (N = 60)

View of the marker

Quality

What is required

Understanding the view of the marker

“Helpful to compare, and see it from the view of the marker”

“I found it quite challenging to give a mark to a piece of work, I found I was struggling to choose between bands”

Understanding of quality

“Useful for own work to be able to see what would constitute a high or low grade, or details that doesn't affect grades, for example, going into extra detail”

“Was helpful as it made me consider how work is marked and therefore how I should write my report in order to gain a strong mark”

Understanding of what is required

“I learnt that I have to make sure my descriptive statistics are written in the right place. Not to use too much description of the type of tests used in the results section.”

“Useful to understand how much information to put into my own answer. For example, don't describe the types of t tests.”

Next...

Future plans for implementation and student progression

1st Year

- Section-based peer assessment spread across 2 semesters
- Collaborative context

2nd Year

- Peer assessment on “live” report sections
- Self-assessment
- Solo context

Practical advantages

- Develops *general* and *specific* knowledge of assessment processes **and** task specific requirements
- Allows low stakes peer assessment with inbuilt calibration
- Distribution of learning across time
- Flexible format: workshops/seminars, lectures and online
- Route to expertise through exposure and practice

Applications

Componential exemplar peer review for other assessment types (e.g. essay paragraphs)

Calibration of assessment ability

Thresholding to allow participation in peer review (e.g. in MOOCS)

Unlocking peer assessment in Canvas by threshold

Staff training and development to improve consistency

How far can you go?

"Evaluative activity can be configured to be the primary pedagogical vehicle for teaching a considerable proportion of the substantive content of a course"

Sadler, 2009

Less telling, more doing