the british
psychological society
promoting excellence in psychology

Teaching mathematical skills through key studies in psychology

September 2021

Contents

© 2021 The British Psychological Society

Teaching mathematical skills through key studies in psychology

Students beginning A-Level psychology are often disheartened to learn that mathematics is an inescapable part of the subject. Many students are drawn to psychology by the promise of gaining a deeper understanding of human behaviour, or possibly even a deeper understanding of themselves. There are few students who choose to study A-Level psychology because of the opportunities that it provides for data analysis and statistics!

However, without mathematics, it would be impossible for psychologists to pick apart what is true from what is sheer conjecture. Statistical analysis separates the signal from the noise, telling us whether what we have found is meaningful, or simply a product of blind chance.

Despite its importance, many students (and a fair few teachers) have a hesitance in teaching mathematical skills. This may be due to the perceived difficulty of the subject or a lack of confidence in teaching the necessary skills. Perhaps the reluctance comes from a feeling that mathematics is dull and unengaging.

For this reason, many psychology teachers will relegate mathematical skills to a standalone unit of learning; a week-long slog of lessons where all mathematical skills are taught at once. It is tempting to quarantine mathematics to its own walled-off corner of the curriculum lest we risk contaminating the rest of the course with its dullness. But by doing so, we do little to help students understand that mathematics is an integral part of the subject, and at the heart of all the "interesting" stuff they learn.

Therefore, a better way of teaching these essential mathematical skills may be to integrate mathematics into lessons, in particular lessons where students are learning about a key piece of research. By bringing in a little bit of mathematics here and there, students should start to gain a more cohesive understanding of the links between data analysis, statistics, and psychology. Additionally, this "little but often" form of teaching may feel less overwhelming for students than the "all at once" method often used, a gentle flurry of snowflakes rather than an avalanche!

What follows are some examples of how this can be done. It is not an exhaustive list, and teachers should feel free to experiment and be creative. Think of the key studies in your specification as tools at your disposal; how can you put them to work to ease the load of teaching the mathematical skills the students need to know? Some studies (as you will see) lend themselves more readily to teaching mathematical skills than others. Additionally, for some studies it may be impossible to find the raw participant data, meaning that you may be limited in the analysis you can do. In these cases (such as in the Bandura study below), you can present the students with data from a hypothetical replication.

The following pages contain activities and exercises based on the studies below. Answers can be found at the back of the handout.

1. Loftus and Palmer (1974), *Reconstruction of automobile destruction: an example of the interaction between language and memory*

2. Milgram (1963), *Behavioural Study of Obedience*

3. Rosenhan (1973) *On Being Sane in Insane Places*

4. Bandura (1963) *Imitation of film-mediated aggressive models*

(Note: The exercises below assume prior knowledge of the studies, and so only a summary of each study is given)

1. LOFTUS AND PALMER (1974), RECONSTRUCTION OF AUTOMOBILE DESTRUCTION: AN EXAMPLE OF THE INTERACTION BETWEEN LANGUAGE AND MEMORY

Summary: participants watched a video of a car accident, and then had their recall tested. Some participants were questioned with leading questions, specifically with regards to the speed of the car and the presence of broken glass.

Look at the results from experiment 1 and answer the questions that follow.

VERB USED IN CRITICAL QUESTION	MEAN SPEED ESTIMATE (MPH)
Smashed	40.5
Collided	39.3
Bumped	38.1
Hit	34.0
Contacted	31.8

a. Describe how the researchers calculated the mean speed given by the participants for each of the verbs.

b. Give one strength and one weakness of using the mean as a measure of central tendency in this study.

c. How much higher is the average speed estimate for "smashed" than for "contacted"? Express this as a percentage. Round to 1 decimal place.

d. Use the data in this table to plot a bar chart to show the results. Ensure that you correctly label your axes and give it a title.

Look at the results from experiment 2 and answer the questions that follow.

	NUMBER OF PARTICIPANTS WHO CLAIMED TO SEE BROKEN GLASS	
CONDITION	YES	NO
Hit	7	43
Smashed	16	34
Control	6	44

e. In all conditions combined, what is the total percentage of participants who claimed to see broken glass?

f. Identify and explain what level of measurement the data in this experiment is.

g. What statistical test would you use to analyse this data? Justify your choice.

The hypothesis for the study was as follows: *Participants who are asked about the car's speed using the verb "smashed" will be more likely to report seeing broken glass when there was none, than those asked with the verb "hit" or those not asked about the speed at all.*

h. The data was analysed using the statistical test identified in question g). The observed value was 7.78. Using a table of critical values, assess whether this finding is significant or not at $P \leq 0.05$. Explain how you found the critical value, and how you determined whether your observed value was significant.

2. MILGRAM (1963), BEHAVIOURAL STUDY OF OBEDIENCE

Summary: participants were deceived into believing that they were administering increasingly severe shocks to another participant in response to the commands of an authority figure. The researchers recorded the maximum shock level that each participant gave before they refused to continue.

Look at the data in the table below. Answer the questions that follow.

MAXIMUM SHOCK LEVEL	NO. PPTS
15V-285V	0
300V	5
315V	4
330V	2
345V	1
360V	1
375V	1
390V	0
405V	0
420V	0
435V	0
450V	26

a. Calculate the range of the shock level administered by the participants. Show your workings.

b. What is the modal score in this set of data? Explain how you identified it.

c. What is the median score in this set of data? Explain how you calculated it.

d. What is the mean score in this set of data? Explain how you calculated it.

e. Explain why the mean may not be an appropriate measure of central tendency for this set of data.

f. Explain why it is difficult to perform an inferential statistical test on the results of Milgram's study.

g. Use the data in this table to plot a line graph to show the results. Ensure that you correctly label your axes and give it a title.

3. ROSENHAN (1973) ON BEING SANE IN INSANE PLACES

Summary: mentally healthy participants got themselves admitted to several psychiatric hospitals (they were known in the study as "pseudopatients"). They recorded their experiences there, including how long it took for the staff to realise that they were not mentally ill.

After the results of the first study were published, staff in another hospital challenged Rosenhan claiming that this could not happen in their hospital. Rosenhan informed the hospital that in

the next three months, one or more pseudopatients would present themselves. Staff and doctors recorded those patients who they judged to be a pseudopatient. (Rosenhan actually sent no one).

Over 3 months, 193 patients were admitted to the hospital.

41 patients were judged to be pseudopatients by at least one staff member.

23 were suspected by at least one psychiatrist.

19 were suspected by one psychiatrist and one staff member.

a. Draw a pie chart to show the proportion of the patients admitted to hospital that were suspected to be pseudopatients (hint: read the information carefully, as it may not be as straightforward as it seems!)

While in the hospital, the pseudopatients approached the staff and doctors, and made a record of the responses they received. This was then compared with the responses that students received from the faculty on a university campus. These are the results expressed as percentages.

	PSYCHIATRIC HOSPITALS		UNIVERSITY CAMPUS
RESPONSES	PSYCHIATRISTS	NURSES AND ATTENDANTS	FACULTY
Moves on, head averted	71%	88%	0%
Makes eye contact	23%	10%	0%
Pauses and chats	2%	2%	0%
Stops and talks	4%	0.5%	100%

b. Rosenhan only gave percentages in his findings, not numbers of participants. However, lets imagine that he tested 154 psychiatrists, 349 nurses and attendants, and 57 faculty. Convert the percentages above into number of participants.

4. BANDURA (1963) IMITATION OF FILM-MEDIATED AGGRESSIVE MODELS

Summary: children were shown video footage of an adult playing with an inflatable "Bobo" doll in either an aggressive or non-aggressive manner. The children's behaviour was recorded, and acts of imitative aggression noted.

A researcher replicated Bandura's study and found the following results.

AGGRESSIVE ROLE MODEL		NON-AGGRESSIVE ROLE MODEL	
PARTICIPANT NUMBER	AGGRESSION SCORE (OUT OF 100)	PARTICIPANT NUMBER	AGGRESSION SCORE (OUT OF 100)
1	75	11	16
2	54	12	47
3	48	13	57
4	80	14	24
5	24	15	39
6	69	16	48
7	55	17	50

AGGRESSIVE ROLE MODEL		NON-AGGRESSIVE ROLE MODEL	
PARTICIPANT NUMBER	AGGRESSION SCORE (OUT OF 100)	PARTICIPANT NUMBER	AGGRESSION SCORE (OUT OF 100)
8	51	18	66
9	74	19	30
10	36	20	27

The experimental hypothesis was: *"Children who observe a video of an adult playing aggressively with the bobo doll will demonstrate a higher aggression score play than children who observed the video of the non-aggressive adult."*

a. Calculate the mean, mode, median and range for each set of data. What does it tell us about the results?

b. Calculate the standard deviation of each group. What does it tell us about the results?

c. Identify and explain the experimental design used in this study.

d. Identify and explain the level of measurement used in this study.

e. Using your answers from questions c) and d), identify the statistical test that you would use to analyse this data.

f. Identify whether hypothesis is one or two tailed and explain your reasoning.

g. The data was analysed using the statistical test identified in question f). The observed value 24. Using a table of critical values, assess whether this finding is significant or not at $P \leq 0.05$. Explain how you found the critical value, and how you determined whether your observed value was significant.

ANSWERS

1. LOFTUS AND PALMER (1974), RECONSTRUCTION OF AUTOMOBILE DESTRUCTION: AN EXAMPLE OF THE INTERACTION BETWEEN LANGUAGE AND MEMORY

a. The researchers would have added together all the speed estimates from the participants asked the question with that particular verb. This total would then have been divided by the number of participants to give the mean speed.

b. One strength is that the mean uses every piece of data, in this case, every participant's speed estimate would have been used to calculate the mean speed. This is a strength as it means that the mean is a more sensitive measure of the average speed than the mode or median.

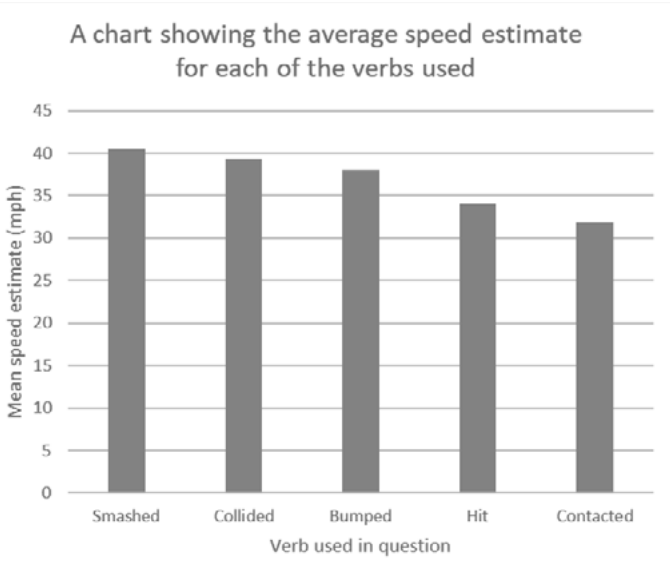
One weakness is that the mean can be skewed by extreme scores. In this case, a single participant in the "smashed" condition might have had guessed 85mph for example. This would have pulled the entire average upwards, making the mean an unrepresentative measure of central tendency.

c. The average speed for the "smashed" condition was 8.7 mph higher than the "contacted" condition.
40.5 – 31.8 = 8.7

The average speed estimated for "smashed" was 27.4% higher than for "contacted.

8.7 / 31.8 = 0.274
0.274 × 100 = 27.4

d.



e. 19.3% of all participants claimed to see broken glass.
7 + 16 + 6 = 29
29 / 150 = 0.193
19.3 × 100 = 19.3

f. The data in this study is nominal. This is because the responses to the question "Did you see any broken glass" are categories; it is either "Yes" or "No".

g. The test used would be Chi Square. This is because the data is nominal (whether the participant answered yes or no to the question) and the experimental design was independent as each participant was only in one of the three conditions; they were either asked the question with the verb "smashed", "hit" or were not asked about the speed at all.

h. To find the critical value, we need to determine whether the test was one tailed or two tailed. Because the hypothesis was directional, we know that we will be looking for a critical value for a one tailed test.

Next, we need to identify the degrees of freedom (df). To do this, we multiply the number of rows minus one, by the number of columns minus one.
(3–1) × (2–1)
2 × 1 = 2
Degrees of freedom is therefore 2.

Now we can find our critical value on the table. First, we find the table for P≤0.05. Next, we look at the column for one tailed test, and the row for df=2, and we find the critical value is **4.60**.

df	ONE-TAILED TEST	TWO-TAILED TEST
1	2.71	3.84
2	4.60	5.99
3	6.25	7.82
4	7.78	9.49
5	9.24	11.07

For a Chi Square result to be significant, the observed value of must be higher than the

critical value. Therefore, as the observed value of 7.78 found in this experiment is higher than the critical value of 4.60, we can conclude that the results found in this study are significant, and therefore unlikely to be due to chance.

2. MILGRAM (1963), BEHAVIOURAL STUDY OF OBEDIENCE

a. The range is 150 volts.
450 – 300 = 150

b. The mode is 450 volts. The mode is the result that appears most often in a set of data, and 450 was the most common result.

c. The median score is 450 volts. When all 40 results are placed in ascending order, there is no middle score as there is an even number of results. Therefore, we look at the two middle scores (position 20 and 21) and then find the halfway point between them. As both numbers are 450, the median is therefore 450.

300	300	300	300	300	315	315	315	315	330
330	345	360	375	450	450	450	450	450	450
450	450	450	450	450	450	450	450	450	450
450	450	450	450	450	450	450	450	450	450

d. The mean is calculated by adding up all the scores, and then dividing by the number of scores. There are two ways to do this for Milgram's study.

- Get a calculator and add up each of the 40 scores (using the numbers you wrote out for the previous question), then divide by 40. While this would work, it is time consuming and easy for mistakes to occur.
- An easier way would be to use the table of data and multiply each voltage by the number of participants, then add up these totals. This total would then be divided by 40 to give a mean of 405. See the table below.

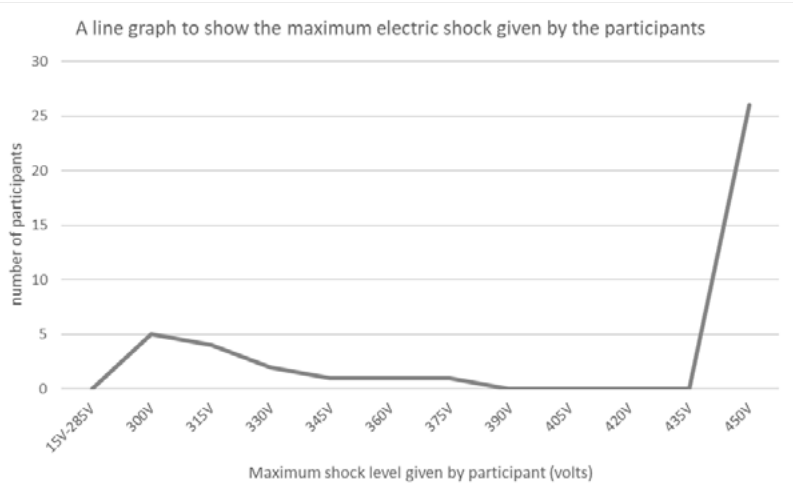
		NO OF PPTS × VOLTAGE
300V	5	5 × 300 = 1,500
315V	4	4 × 315 = 1,260
330V	2	2 × 330 = 660
345V	1	1 × 345 = 345
360V	1	1 × 360 = 360
375V	1	1 × 375 = 375
390V	0	0 × 390 = 0
405V	0	0 × 405 = 0
420V	0	0 × 420 = 0
435V	0	0 × 435 = 0
450V	26	26 × 450 = 11,700
	TOTAL	16,200
	MEAN	16,200 / 40 = 405

e. The mean gives an average result of 405 volts. It could be argued that this is not an appropriate measure of central tendency because when we look at the data, none of the

participants actually gained this score. Most of the scores were actually above this value, which makes the mean, in this case, an unrepresentative measure.

f. Inferential tests usually tell us whether there is a difference between two conditions, usually an experimental condition and a control condition. However, in Milgram's study, there was no control condition. Therefore, an inferential statistical test cannot be carried out.

g.



3. ROSENHAN (1973) ON BEING SANE IN INSANE PLACES

a. First, identify the sections of your chart. These will be:

Patients suspected by at least one staff member.
Patients suspected by at least one psychiatrist.
Patients suspected by one psychiatrist and one staff member.
Patients not suspected.

However, we need to ensure that each participant is only included in one section of the pie chart. At the moment, some patients in the first two categories (suspected by a staff member or suspected by a psychiatrist) are also included in the third category (suspected by a staff member AND a psychiatrist). Therefore, we need to remove these participants from the first two categories.

Patients suspected by at least one staff member: 41 – 19 = **22**
Patients suspected by at least one psychiatrist: 23 – 19 = **2**

However, there is also another category of patient that needs to be included in the study: those who were not suspected at all. To calculate this, we take the total number of patients admitted, and subtract all the patients who were suspected.

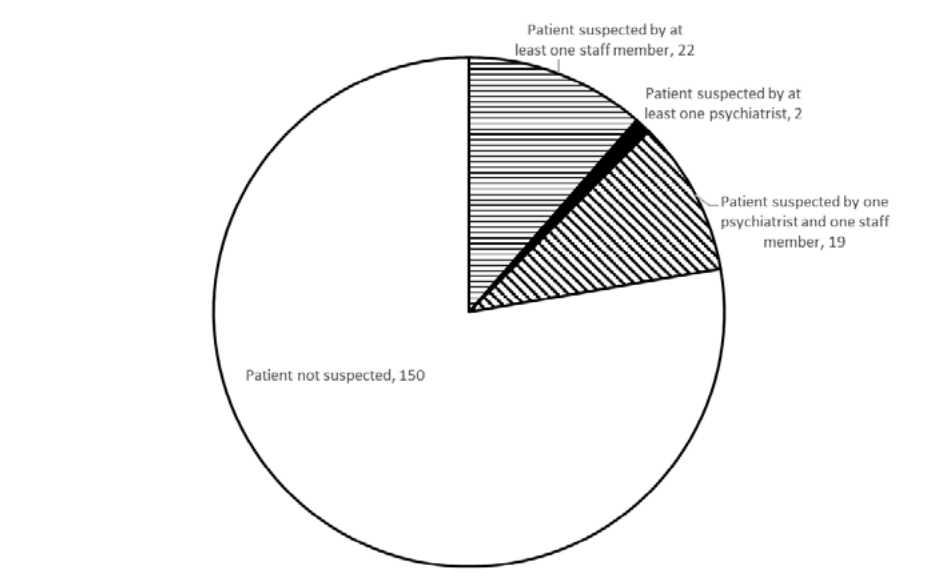
193 – (22 + 2 + 19)
193 – 43 = 150

There are 360 degrees in a circle, so to figure out the angle of each section, take value, divide it by the whole (193), and then multiply that by 360. Round to the nearest degree.

Patients suspected by at least one staff member	(22 / 193) × 360 = **41**˚

Patients suspected by at least one psychiatrist	(2 / 193) × 360 = **4°**
Patients suspected by one psychiatrist and one staff member	(19 / 193) × 360 = **35°**
Patients not suspected	(150 / 193) × 360 = **280°**

Use these angles to draw your pie chart. Make sure to clearly label the different sections.



b. To calculate the actual participant numbers for each group, you need to take the total number of participants, divide by 100, and then multiply by the percentage. Then, round this number to the nearest whole number.

For example, to calculate the number of psychiatrists who moved on, head averted:

(153 / 100) × 71 = 108.63

Rounded: 109 participants

Once done, you should have a table that looks like this:

	PSYCHIATRIC HOSPITALS		UNIVERSITY CAMPUS
RESPONSES	**PSYCHIATRISTS**	**NURSES AND ATTENDANTS**	**FACULTY**
Moves on, head averted	109	307	0
Makes eye contact	35	35	0
Pauses and chats	3	7	0
Stops and talks	6	2	57

4. BANDURA (1963) IMITATION OF FILM-MEDIATED AGGRESSIVE MODELS

a. Mean for aggressive role model = **56.6**

75 + 54 + 48 + 80 + 24 + 69 + 55 + 51 + 74 + 36 = 566
566 / 10 = 56.6

Mean for non-aggressive role model = **40.4**
16 + 47+ 57 + 24 + 39 + 48 + 50 + 66 + 30 + 27 = 404
404 / 10 = 40.4

Mode for aggressive role model = **N/A**
Mode for non-aggressive role model = **N/A**
(There is no mode for either group as there are no repeated scores)

Median for aggressive role model = **54.5**
24 36 48 51 54 55 69 74 75 80
Middle scores are 54 and 55
(54 + 55) / 2 = 54.5

Median for non-aggressive role model = **43**
16 24 27 30 39 47 48 50 57 66
Middle scores are 39 and 47
(39 + 47) / 2 = 43

Range for aggressive role model = **57**
(80 – 24) + 1 = 57

Range for non-aggressive role model = **51**
(66 – 16) + 1 = 51

These descriptive statistics suggest that the children who observed the aggressive role model had a higher aggression score than those who observed the non-aggressive role model. Both the mean and the median were higher for children in the aggressive role model condition. The range for both groups was quite similar, suggesting that there was an even spread of data in both groups.

b. Standard deviation formula

$$\sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Standard deviation for aggressive role model = **18.09**

$\bar{x}$ Calculate the **mean**.

Done in question a) = **56.6**

$x - \bar{x}$ Calculate the **raw score mean**.

75 – 56.6 = **18.4**
54 – 56.6 = **–2.6**
48 – 56.6 = **–8.6**
80 – 56.6 = **23.4**
24 – 56.6 = **–32.6**
69 – 56.6 = **12.4**

55 – 56.6 = **–1.6**
51 – 56.6 = **–5.6**
74 – 56.6 = **17.4**
36 – 56.6 = –**20.6**

$(x - \bar{x})^2$ **Square** each of the numbers

18.4^2 = **338.56**
-2.6^2 = **6.76**
-8.6^2 = **73.96**
23.4^2 = **547.56**
-32.6^2 = **1062.76**
12.4^2 = **153.76**
-1.6^2 = **2.56**
-5.6^2 = **31.36**
17.4^2 = **302.76**
-20.6^2 = **424.36**

$\sum(x - \bar{x})^2$ **Sum** all these numbers together.

338.5 + 6.76 + 73.96 + 547.56 + 1062.76 + 153.76 + 2.56 + 31.36 + 302.76 + 424.36 = **2944.4**

$\frac{\sum(x - \bar{x})^2}{n - 1}$ **Divide** this by the number of scores minus 1.

2994.4 / (10–1) = **327.16**

$\sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$ Find the **square root** of this number.

√327.16 = **18.09**

Standard deviation for non-aggressive role model = **15.9**

x Calculate the **mean**.

Done in question a) = **40.4**

$x - \bar{x}$ Calculate the **raw score mean**.

16 – 40.4 = **–24.4**
47 – 40.4 = **6.6**
57 – 40.4 = **16.6**
24 – 40.4 = **–16.4**
39 – 40.4 = **–1.4**
48 – 40.4 = **7.6**
50 – 40.4 = **9.6**
66 – 40.4 = **25.6**

30 – 40.4 = **–10.4**
27 – 40.4 = **–13.4**

$(x - \bar{x})^2$ **Square** each of the numbers

-24.4^2 = **595.36**
6.6^2 = **43.56**
16.6^2 = **275.56**
-16.4^2 = **268.96**
-1.4^2 = **1.96**
7.6^2 = **57.76**
9.6^2 = **92.16**
25.6^2 = **655.36**
-10.4^2 = **108.16**
-13.4^2 = **179.56**

$\sum (x - \bar{x})^2$ **Sum** all these numbers together.

595.36 + 43.56 + 275.56 + 268.96 + 1.96 + 57.76 + 92.16 + 655.36 + 108.16 + 179.56 = **2278.4**

$\frac{\sum (x - \bar{x})^2}{n - 1}$ **Divide** this by the number of scores minus 1.

2278.4 / (10–1) = **253.16**

$\sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$ Find the **square root** of this number.

√253.16 = **15.9**

When we compare the standard deviation of both groups, we can see that the standard deviation of the aggressive group is slightly larger than the standard deviation of the non-aggressive group. Therefore, we can say that the distribution of the scores in the aggressive group are more spread out than in the non-aggressive group. However, this difference is only minor.

c. The experimental design in this study was independent measures. Each of the children was only in one of the two conditions (either they viewed an aggressive or a non-aggressive role model).

d. The level of measurement was ratio data. The score of aggression is a scale with meaningful numbers, and an absolute zero (the score cannot go below 0).

e. The experiment was looking for a difference between the two conditions, and because the data is ratio, and the experiment used independent measures, the Man Whitney U test would be an appropriate choice to analyse the data.

f. The hypothesis is directional, as it states that the group who viewed the aggressive role model will score higher on the score of aggression than those who viewed the

non-aggressive role model. Therefore, as the hypothesis is directional, it is one tailed.

g. To find the critical value, we first need to identify whether test was one tailed or two tailed. Because the hypothesis was directional, we know that we will be looking for a critical value for a one tailed test. Therefore, we will look at the table where P≤0.05 and the test was one tailed.

h. To look up the value on this table, we have to identify N1 and N2. This refers to the number of participants in each group. In this experiment, both and N1 and N2 is 10. Using this information, we look up the critical value, which is **27**.

CRITICAL VALUES FOR A ONE-TAILED TEST (P≤ 0.05)

		N1								
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
N2	**2**				0	0	0	1	1	1
	3		0	0	1	2	2	3	3	4
	4		0	1	2	3	4	5	6	7
	5	0	1	2	4	5	6	8	9	11
	6	0	2	3	5	7	8	10	12	14
	7	0	2	4	6	8	11	13	15	17
	8	1	3	5	8	10	13	15	18	20
	9	1	3	6	9	12	15	18	21	24
	10	1	4	7	11	14	17	20	24	27

For a Man Whitney U test result to be significant, the observed value of must be lower than the critical value. Therefore, as the observed value of 24 found in this experiment is lower than the critical value of 27, we can conclude that the results found in this study are significant, and therefore unlikely to be due to chance.

the british
psychological society
promoting excellence in psychology

St Andrews House
48 Princess Road East
Leicester LE1 7DR, UK

☎ 0116 254 9568 www.bps.org.uk info@bps.org.uk

Incorporated by Royal Charter | Registered Charity No 229642

TT/09.2021