National Certificate of Educational Achievement
TAUMATA MĀTAURANGA Ā-MOTU KUA TAEA

Exemplar for Internal Assessment Resource

Mathematics Level 3

Resource title: Tricky questions

This exemplar supports assessment against:

Achievement Standard 91583

Conduct an experiment to investigate a situation using experimental design principles

Student and grade boundary specific exemplar The material has been gathered from student material specific to an A or B assessment resource.

Date version published by Ministry of Education — December 2012
To support internal assessment from 2013

The task asks students to conduct an experiment to investigate the effects of one type of question bias in the design of a questionnaire.

	Grade Boundary: Low Excellence
1.	For Excellence the student is required to conduct an experiment to investigate a situation using experimental design principles, with statistical insight. This involves integrating statistical and contextual knowledge throughout the investigation process, and may include reflecting about the process; discussing how possible sources of variation were dealt with during the design phase; considering other relevant variables. There is evidence of: • Background contextual knowledge by referencing research related to using anchors to influence estimates when people are uncertain, leading to an appropriate question (1) • A logical and clearly described plan for the experiment (2) • An appropriate formal statistical inference by assessing and interpreting the strength of evidence about the causal relationship (3) • Consideration of broadening the experimental situation in the discussion of their findings (4). For a more secure Excellence the student could have discussed in greater depth how the background knowledge relates to this experiment and how the results of the experiment were consistent with other research studies.

I investigated possible bias in questionnaires from using anchors. I wanted to find out if you could get people to give higher answers for a question by using an anchoring question before it. This would be important to take into account when using questionnaires to collect data so you don't unknowingly influence answers, or something that might be used in questionnaires to trick people into giving a certain answer. I spent some time researching each of the different types of questionnaire bias given in resource A, and it seemed to me that anchoring effect would be interesting to investigate because of the applications of this to everyday life. In my research, I found examples of how people use anchoring bias when selling cars to persuade people into paying more. The research about anchoring bias shows that when people are uncertain about something, they use whatever information is available to help them decide, even if the information is not valid or reliable. I wanted to do a similar experiment, so my investigative question was "will higher estimates be given for the number of people who live in Venezuela if a larger number is used for the anchor?" I used Venezuela for my experiment as I thought that it was a country not many people would know exactly how many people lived there, and for the experiment to work people have to be unsure about the answer or amount for the anchor to have an effect. The research suggested that the higher the number I used in the anchor, the higher the estimates of the number of people who live in Venezuela would be, and so this is what I expected to find in my experiment. 1

I used a comparison of two independent groups for my experiment, where one group was given a high anchor and one group was given a value close to the real value. The groups had to be independent because if I gave someone both anchors then they might realise what the experiment was about and I wanted their behaviour to be natural. I used two year 13 Mathematics classes for my experiment (57 students in all), which was convenient because they were on at the same time as my Statistics class. The experiment was conducted during my Statistics lesson period 4 on Wednesday. I used single blinding, where the participants didn't know which treatment they were getting. In fact, I concealed the fact that it was an experiment at all by presenting the questionnaire as a general knowledge survey.

The response variable was the estimate for the number of people in Venezuela (in millions). I decided to use millions for the unit as this generally what populations for different countries as measured in. The treatment variable was the number used for the anchor. I had two treatment groups: for one group the anchor was 60 million, for the other group the anchor was 30 million. I used these two numbers as the population of Venezuela is around 29 million.

I created two different questionnaires for my experiment – these are provided in the appendix of this report. I included an introduction for the questionnaire used in the experiment that said it was a general knowledge survey, and asked a couple of other questions in the questionnaire (What is your gender? How many continents are there in the world?) so that people would not guess the point of my experiment. In my questionnaire, I decided to ask people to estimate of the number of people in Venezuela to the nearest million because I was confident this would still give me enough variability in the estimates and it would be easier for people to answer the question. The two versions of the questionnaire are exactly the same, except for the number used for the anchor in the question before the one that asks people to estimate the number of people in Venezuela (e.g. Is the number of people in Venezuela bigger or smaller than 30 million?). 2

For this experiment, it was important that people didn't realise there were two different versions of the questionnaire. Before we gave them to students to fill out, we turned the questionnaires upside down and thoroughly shuffled them into a pile to hand out. This was done to make sure that each student did not know which of the two questionnaires they were getting. In this way, we would be randomly allocating students to one of the two treatment groups when we gave out the questionnaires to complete.

There are some factors I couldn't control for my experiment, like:

- whether people already know the population of Venezuela (maybe people who had been travelling or international students from South America)
- whether people would take the survey seriously and not give silly answers
- whether people were aware of the anchoring bias (which maybe students doing psychology might know about)
- whether some people had better general knowledge than others

By randomly assigning people to one of the two treatment groups, I attempted to balance the possible effect of these variables on the estimates across the two groups. For example, people who may be better at estimating the population of Venezuela should be balanced across both groups and should not significantly affect the results for one group only.

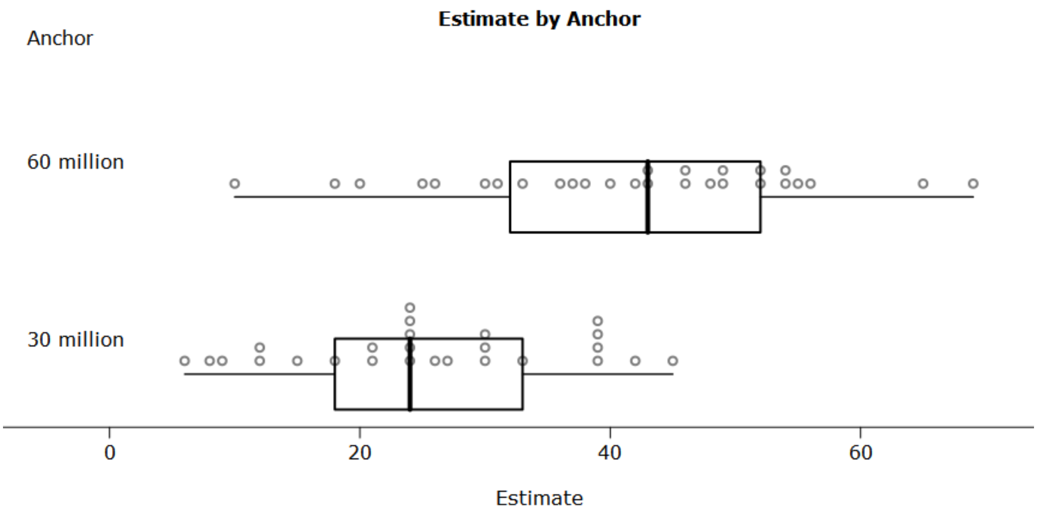
The other variables I controlled for my experiment were:

- giving the same instructions to people about completing the questionnaire
- both groups doing the experiment at the same time of the day
- same test conditions used for completing the questionnaire
- an independent person carrying out the experiment (the teacher of the class) so that I didn't influence the results if I was the handing out the questionnaires

I didn't tell the teacher that the survey was an experiment – I told them it was just a survey to collect data for analysis for our Statistics class. In this way, they were not aware of the two treatment groups. The experiment was done at the beginning of the lesson by the teacher reading out the instructions (which are in the appendix). I entered the responses from each questionnaire into a spread sheet and made notes about any responses that were not clear (see appendix).

The raw data collected from the experiment is included in the appendix of this report.

To analyse the data I used iNZight to generate the following dot and box-and-whisker plots and summary statistics.



Summary of Estimate by Anchor

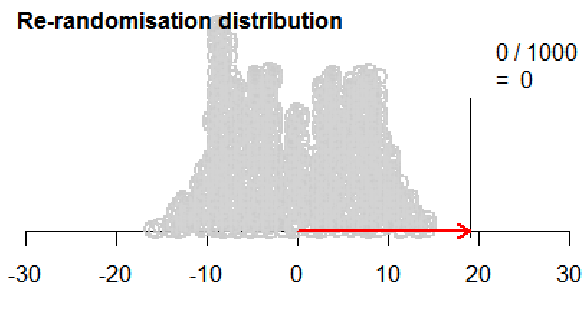
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
30 million	6	21.0	24	25.96	33	45	10.907	25
60 million	10	32.5	43	41.68	52	69	14.053	28

The first thing I noticed about the displays when comparing the two groups is that the estimates for the number of people in Venezuela do not look like what I would expect if was purely by chance. For the amount of data I have, purely by chance, I would expect the data for both groups to look similar – the boxes would be sitting roughly in the same place, they would be roughly the same size, and the shape of the distribution of the data would be similar.

However, there is a difference of 19 million between the two treatment group medians (the treatment group which used an anchor of 60 million had a median estimate of 43 million and the treatment group with the 30 million anchor had a median estimate of 24 million). This difference, and considering the positions of the distributions, suggests that the use of an anchor of 60 million would result in estimates which tend to be higher than when a 30 million anchor is used.

Also, the variability in these estimates for the 30 million anchor group was smaller than the variability for the 60 million anchor group – you can see this in the box-and-whisker plot, as the box for the 60 million anchor group is wider. This shows that people who were in the 60 million anchor group were less consistent in their estimates, and most gave values quite a bit lower than the anchor of 60 million. People who were in the 30 million anchor group gave estimates closer to the anchor of 30 million. The distribution of the estimates for the 60 million anchor group is slightly left skewed. The mean and median for the estimates of both groups are very similar.

I have decided to focus on the difference between the medians of the two groups (I could have also used the means) to look for evidence to answer my investigative question. To explore whether it is highly unlikely to get a difference as big or bigger than my observed difference of 19 million just by chance, I randomly assigned the estimates to the two groups. I did this using a computer package. Each observed value was randomly re-assigned (re-randomised) to one of the two groups and the difference between the re-randomised group medians was calculated. This was repeated 1000 times producing, under chance alone, 1000 differences between the group medians. The graphs and results produced from this method are shown below:



Not one of the 1000 re-randomised differences was equal to or higher than the observed difference between the medians of 19 million. A difference as big as 19 million was highly unlikely to have happened purely by chance.

Using the result from the randomisation test, I have very strong evidence that the use of an anchor of 60 million would cause estimates that tend to be higher than when a 30 million anchor is used. This is because when I compared the observed difference between the group medians (19 million) to the distribution of re-randomised differences, a difference of 19 million or higher never came up in a 1000 re-randomisations. This shows that in this experiment it would be very unlikely that a difference as large as 19 million could happen just by chance. It is this test result that provides me with the very strong evidence that chance was not acting alone in this experiment but something else, namely the anchor effect, was acting along with chance to create the observed difference of 19 million.

I thought that the experiment turned out reasonably as planned, although not everyone completed the survey correctly, so I couldn't use all the results (see my notes in the appendix). However, there weren't many incomplete or invalid responses, so this shouldn't have affected my data too much. I could have also had the teacher check the questionnaires as they were handed back so that responses could be clarified.

The results from the randomisation test, and the fact that my experiment was well- designed and executed, means that I can claim that an anchor of 60 million would cause estimates that tend to be higher than when a 30 million anchor is used. My results are also important in terms of how they apply to questionnaire design, and the importance of making sure that questions are not used that may influence people's answers to other questions in the questionnaire. 3

I wonder how much the numbers I used for each of the treatment groups (the numbers for the anchors) affected the estimates. When I looked at the data for the two treatment groups, it seemed that people were not confident estimating the number of people in Venezuela to be as high as 60 million (the median estimate was 43 million). Maybe if I had used a value like 100 million they would have ignored it because it would have been unrealistic. Would I have got the same result (my conclusion that a larger number for the anchor would result in (cause) estimates that tend to be higher) if I used treatment values of 40 million? So perhaps anchors can influence people's responses to answers, but only if they are a certain value. It would be interesting to repeat the experiment to see how far apart the numbers I use for the two treatment groups have to be to see an effect in the estimates for the number of people in Venezuela. It would also be interesting to see what people's estimates would be for the number of people in Venezuela without any anchor. I also noticed that the 30 million group were not confident that the population was as high as 30 million. The distribution of the estimates for the number of people in Venezuela (both the shape and spread) also suggests that the use of an anchor of 60 would result in (cause) estimates that tend to be more varied than when a 30 million anchor is used. I could investigate by using statistical tests to see whether there is evidence of a difference of spread for the estimations due to the intervention of the anchors. 4

Appendix

Statement read out to the class:

I am going to give you a general knowledge survey to complete. Please answer the questions as best you can and place your survey face down when you have finished for me to collect. Keep your eyes on your own survey and do not talk – just like a test.

Questionnaires used:

General knowledge survey A

What is your gender? (please circle answer) male female

How many continents are there in the world? __________

Is the number of people in Venezuela bigger or smaller than 30 million? bigger smaller

Estimate, to the nearest million, the number of people in Venezuela ________________

General knowledge survey B

What is your gender? (please circle answer) male female

How many continents are there in the world? __________

Is the number of people in Venezuela bigger or smaller than 60 million? bigger smaller

Estimate, to the nearest million, the number of people in Venezuela ________________

Results:

Anchor	Estimate
30 million	12
30 million	9
30 million	24
60 million	18
60 million	48
30 million	33
60 million	65
60 million	
60 million	69
60 million	43
60 million	43
30 million	30
60 million	26
60 million	
30 million	6

Anchor	Estimate
30 million	39
30 million	27
60 million	20
30 million	39
30 million	
60 million	10
60 million	25
60 million	38
30 million	15
30 million	8
60 million	31
60 million	46
30 million	42
30 million	21
30 million	18

Anchor	Estimate
30 million	24
60 million	46
60 million	49
30 million	24
60 million	54
30 million	45
60 million	37
30 million	12
60 million	56
60 million	30
60 million	52
60 million	49
30 million	30
60 million	40
30 million	24

Anchor	Estimate
30 million	30
30 million	21
30 million	26
60 million	42
30 million	39
60 million	36
30 million	24
60 million	54
60 million	55
60 million	52
30 million	39
60 million	33

Experiment notes:

The classes were pretty quiet and it didn't look like anyone copied from each other or talked about the questionnaire.

I ended up with a few questionnaires left over, with more of the "30 million" question.

	Grade Boundary: High Merit
2.	For Merit the student is required conduct an experiment to investigate a situation using experimental design principles, with justification. This involves linking components of the process of investigating a situation by experiment to the context, explaining relevant considerations in the investigation process, and supporting findings with statements which refer to evidence gained from the experiment. There is evidence of: • An appropriate causal relationship question with a prediction of likely outcomes (1) • A description of the type of experiment and identification of the variables and how the treatment was allocated to the experimental units (2) • A formal inference by assessing the strength of evidence for the causal relationship (3). To be awarded Excellence the student would need to integrate the information found in the research into the introduction and justification for their prediction, as well as linking the conclusion with contextual knowledge.

Student 2: High Merit

Information/ideas found from research for experiment

- Grouping questions that are similar will make the questions easier to complete and the respondent will feel more comfortable.
- Questions that use the same response formats, or those that cover a specific topic, should be together.
- Each question should follow comfortable from the previous question.
- Transition between each question should be smooth.
- Questionnaire's that jump from one unrelated topic to another feel disjointed and are not likely to produce high response rates.
- Most investigations have found that the order in which questions are presented can affect that way that people respond.
- Presenting general questions before specific questions avoids response contamination.

I will investigate how you frame (word) a question and how this affects the answers. In particular, I am going to focus on the time period I ask someone to remember something over. My investigative question is "Will estimates for the time spent on Facebook be lower if I frame the question to be over a week compared to estimating the time per day?" I expect to find that people will give lower answers when asked about how much time they spend on Facebook over a week because it is too long a time period to remember exactly.

1

I will investigate this by using the students in two year 10 classes as my experimental units. In total there will be 61 students. My experiment will consist of two surveys that are exactly the same (about Facebook) except for one question. These are shown below:

2

How many hours (to the nearest half an hour) do you typically spend on Facebook each day? ____________

Or

How many hours (to the nearest half an hour) do you typically spend on Facebook each week? ____________

My experiment will consist of two treatment groups. 32 students will be getting the "each day" survey and 29 students will be getting the "each week" survey. I put the two surveys together, turned them upside down and shuffled them so that they were completely mixed up. As there was no way of determining which survey each student would get the allocation is completely random. The response variable is the number of hours they spent on Facebook that they write down. I will multiply the estimates for the "each day" group by seven to make it over a week so I can compare this to what people write for estimates for the "each week" group.

Plan for experiment:

- Go to the two year 10 classes and ask the teacher to hand out the questionnaires, but before that mix it up so it not obvious that the students are getting different sheets, and also to randomly allocate students to the two treatment groups
- Every student answers every question in the survey.
- Collect the surveys when they have finished

2

Variables that I can control for the experiment include the following:

- giving the same instructions to all students about completing the questionnaire
- both groups doing the experiment at the same time of the day
- same test conditions used for completing the questionnaire
- I asked the class teacher to allocate the surveys so that I didn't influence the results

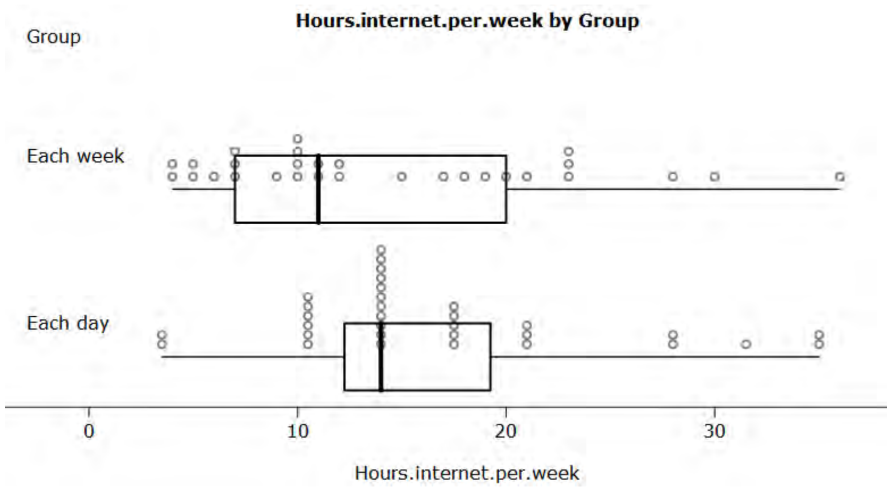
Variables I couldn't control for my experiment, like:

- whether the students had access to the internet at home (because of no computer or maybe punishment)
- whether students would take the survey seriously and not give silly answers
- Students ability to remember (how long they used Facebook)

By randomly allocating one of the two surveys to each student, I attempted to balance the possible effect of these variables on the estimates across the two groups. For example, people who may be better able to estimate the time they used Facebook should be balanced across both groups and should not significantly affect the results for one group only.

I had equal numbers of each survey but I did not get all the responses back. I think this may have been because some students didn't want to participate once given the survey. Also I had a suspicion that using year 10 students might produce some silly responses. In each class I asked the teachers to hand out the surveys. I heard one year 10 student ask the teacher "I can't remember how many hours I used Facebook, should I make up a number?" to which the response was "As long as it is a rough estimate".

I used the computer to get the following graphs and statistics:



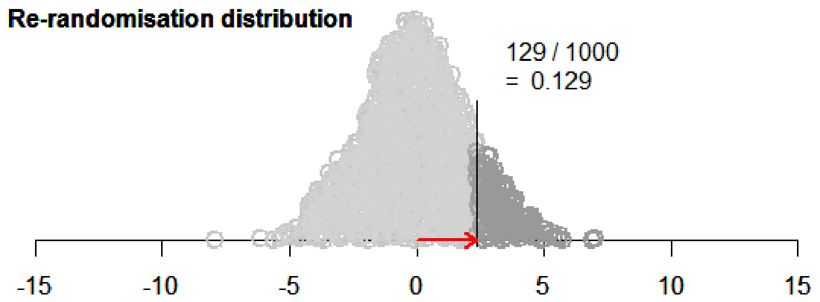
```
Summary of Hours.internet.per.week by Group
          Min. 1st Qu. Median  Mean 3rd Qu. Max. Std.dev Sample.Size
Each day   3.5   13.12     14 16.62   18.38   35   7.750          32
Each week  4.0    7.00     11 14.24   20.00   36   8.454          29
```

The hours spent on the internet per day have bigger gaps between them for the dot plot, because they are multiples of seven.
In the box and whisker graph the overall spread for both groups is very similar, but the estimates for the “each week” group are more evenly spaced while the estimates for the “each day” group are more bunched in the middle. The difference between the means of the two groups was 2.38 hours. But did this difference occur purely by chance? For the size of the groups I used, just by chance, I would expect the box and whisker graphs for the estimates for both groups to look similar – the boxes would be roughly in the same place, they would be roughly the same size, and the shape of each distribution of estimates would be similar.

Comparing the means (I could also have used the medians) the mean of the estimates of the “each week” group is 2.38 hours higher than the mean for the estimates of the “each day” group.

Using the estimates from this group of 61 year 10 students, if I re-randomise the estimates (randomly re-allocate the students’ estimates to either one of the “each week” or “each day” groups) and compare the results, what do I see happening?

Using the iNZight software I did this 1000 times and produced a graph of the differences between the means of the “each week” and “each week” groups and compared this to the difference between the means of my original group.



Conclusion

From the distribution I see that the difference between means that I got (2.38 hours), or more, happened 129 out of 1000 times. That is a probability of 12.9%, so it is possible that the difference (2.38 hours) that I got was due to chance. From this I see that there is not sufficient evidence to support the claim that people who are asked to estimate the time on Facebook per week will estimate less time than those who are asked to estimate the amount of time on Facebook per day.

3

Appendix

Results:

Group	Hours on Facebook /week
Each day	21
Each day	10.5
Each day	10.5
Each day	35
Each day	3.5
Each day	14
Each day	14
Each day	17.5
Each day	28
Each day	14
Each day	21
Each day	28
Each day	35
Each day	17.5
Each day	14
Each day	14

Group	Hours on Facebook /week
Each day	14
Each day	17.5
Each day	14
Each day	14
Each day	10.5
Each day	21
Each day	17.5
Each day	17.5
Each day	31.5
Each day	14
Each day	10.5
Each day	10.5
Each day	3.5
Each day	14
Each day	14
Each day	10.5

Group	Hours on Facebook /week
Each week	9
Each week	12
Each week	7
Each week	7
Each week	10
Each week	10
Each week	6
Each week	36
Each week	5
Each week	4
Each week	28
Each week	30
Each week	21
Each week	23
Each week	18
Each week	11

Group	Hours on Facebook /week
Each week	10
Each week	23
Each week	15
Each week	10
Each week	23
Each week	19
Each week	17
Each week	5
Each week	4
Each week	7
Each week	12
Each week	11
Each week	20

The hours using Facebook per week for the "each day" group was obtained by multiplying by seven

	Grade Boundary: Low Merit
3.	For Merit the student is required conduct an experiment to investigate a situation using experimental design principles, with justification. This involves linking components of the process of investigating a situation by experiment to the context, explaining relevant considerations in the investigation process, and supporting findings with statements which refer to evidence gained from the experiment. There is evidence of conducting an experiment to investigate a situation using experimental design principles, with justification. • The plan justifies the definition of the treatment and response variables (1) • The student links the features of the data to the features of the experiment and the need to do the randomisation test (2) • The student also correctly interprets the results of the randomisation test which is used to assess the strength of evidence (3). For a more secure Merit the student would need to provide more detail linking the research to the situation being investigated, provide a more detailed description of the sources of variation and identify the variables that cannot be controlled.

Student 3: Low Merit

I found this article online about anchoring effects: http://www.overcomingbias.com/2007/09/anchoring-and-a.html

Paragraphs from the article:

- Suppose I spin a Wheel of Fortune device as you watch, and it comes up pointing to 65. Then I ask: Do you think the percentage of African countries in the UN is above or below this number? What do you think is the percentage of African countries in the UN?
- Tversky and Kahneman (1974) recorded the estimates of subjects who saw the Wheel of Fortune showing various numbers. The median estimate of subjects who saw the wheel show 65 was 45%; the median estimate of subjects who saw 10 was 25%.
- The current theory for this and similar experiments is that subjects take the initial, uninformative number as their starting point or anchor; and then they adjust upward or downward from their starting estimate until they reached an answer that "sounded plausible"; and then they stopped adjusting. This typically results in under-adjustment from the anchor - more distant numbers could also be "plausible", but one stops at the first satisfying-sounding answer.

I decided to investigate student's knowledge about the school and whether I can use an anchoring question to influence answers. My question was "Will estimates for the proportion of students who walk to school be higher if I ask people a question with a high anchor number first?" Because I don't think people will know what the actual proportion of students who walk to school is, I think I should be able to trick them into giving higher estimates when I use a high number for the anchor, like what is says in the article.

For my experiment I had to choose a response variable that students would not know the exact answer for (they may have an idea about its value but I would still expect variation in the estimates given). I asked students to estimate the proportion of people who walk to school. I made up a short survey about the school, and asked questions like "What year level are you in?" and "How many students are there at the school?" and other questions that looked like the survey was about finding out what they knew about the school.

1

For the anchors I chose the two numbers 30% and 60%, because around 30% of students walk to school, 60% is double the actual proportion. Below is part of the survey I used:

> You have been randomly assigned a number between 1 and 100.
>
> Your number is ______.
>
> Do you think the proportion of students at our school who walk to school is above or below this number?
>
> Estimate the proportion of students at our school who walk to school. __________

I took this idea from the article, and I hand wrote either 30% or 60% on each survey sheet, to make it look even more like it was a random number (even though I only used 30% or 60%). I made up equal numbers of each version of the survey (with either 30% or 60%) and then shuffled them up into a random order to hand out.

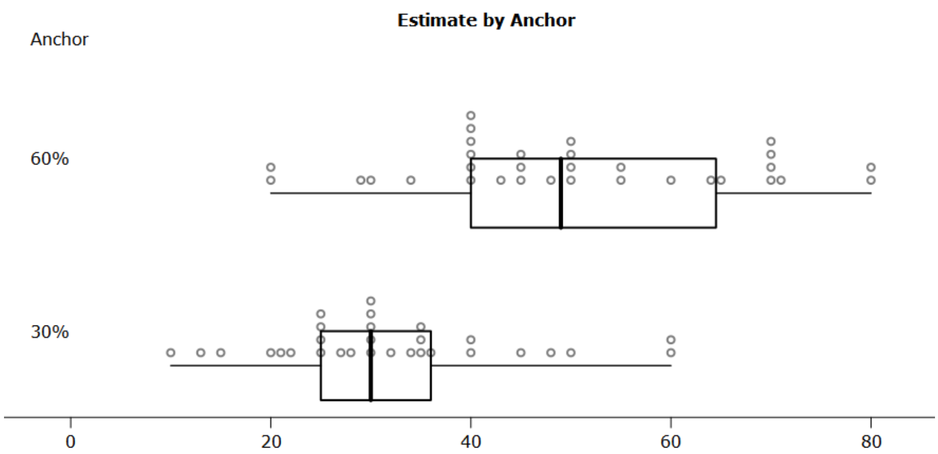
I went to two different classes (both Year 9 classes) and handed out the question sheet, and when the students completed it, I made sure that the students do not look at each other's surveys. I also told them it

1

was a personal survey and that I was interested in their response and how much they knew about the school. My experiment was a comparison of two independent groups design, so students only completed one version of the survey.

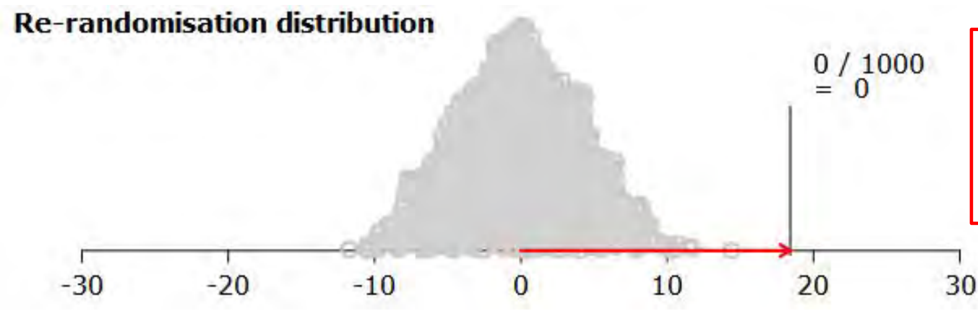
Some students might know the proportion of students who walk to school (if they are involved with travel wise or the school council) but they should be in both of the groups because I randomly mixed up the different surveys before handing out.

Below are my graphs and statistics for the data collected from my experiment:




```
Summary of Estimate by Anchor
    Min. 1st Qu. Median  Mean 3rd Qu. Max. Std.dev Sample.Size
30%   10      25     30 31.87   35.75   60  12.102          30
60%   20      40     49 50.28   64.25   80  16.099          32
```

The data looks good to me because the estimates from the group with the 60% anchor are higher than the estimates from the group with the 30% anchor which suggests that 60% anchor did cause higher estimates overall. It is interesting though that in the 60% group there was one person who did get the answer right (30% for the proportion of our students who walk to school). I don't think that the groups would look like this if only chance was responsible for the estimates of the proportion of students walking to school. The distribution of the estimates for the 30% anchor group is almost symmetrical but the distribution of the estimates for the 60% anchor group is more skewed and slightly more spread out (a slightly wider box and slightly bigger standard deviation). The mean of the estimates for the 60% anchor group is just over 20% more than the real value (30%) and the mean of the estimates for the 30% anchor group is nearly 2% higher than the real value (30%), with a difference between the means of the two groups of 18.41%.



This could happen by chance just by randomly allocating the people to two different groups, so I need to do the randomisation test to see how many times a difference of 18.41% comes up when the estimates people gave are re-randomised to the two groups (30% and 60% anchor) and the differences of the means of the two groups are calculated.

2

The difference of 18.41% or higher never came up.

The design of my experiment was good and I carried it out well, so I am happy that there are no other explanations for what I see in the data (that the 60% anchor group has estimates which tend to be higher than the 30% group, with a difference of 18.41%) apart from chance and the anchor questions I used. The randomisation test gives me very strong evidence as it shows me that in this experiment it would be very unlikely that a difference as large as 18.41% could happen by chance alone. This means I can claim that the use of the anchoring question had an effect on the estimates for the proportion of students at our school who walk to school, in particular that the higher anchor of 60% caused estimates that tended to be higher than the estimates from the anchor of 30%.

3

Appendix

Results:

Anchor	Estimate
60%	65
60%	43
30%	30
60%	40
60%	70
60%	80
60%	70
60%	70
30%	25
60%	40
60%	50
60%	50
30%	25
30%	25
30%	25
60%	40

Anchor	Estimate
30%	13
60%	60
60%	20
60%	40
60%	80
30%	22
30%	40
30%	10
30%	35
30%	45
60%	45
30%	15
30%	30
60%	40
30%	50
30%	30

Anchor	Estimate
60%	40
30%	32
60%	20
30%	60
60%	70
30%	60
60%	48
30%	48
60%	45
30%	28
30%	35
60%	55
30%	20
60%	30
30%	40
60%	50

Anchor	Estimate
30%	30
60%	55
30%	21
60%	64
60%	34
60%	29
60%	71
30%	34
60%	45
60%	50
30%	30
30%	27
30%	35
30%	36

	Grade Boundary: High Achieved
4.	For Achieved the student is required to conduct an experiment to investigate a situation using experimental design principles. This involves showing evidence of using each component of the investigation process. There is evidence of: • Posing an investigative question (1) • Defining and explaining the response variable (2) • Features of the data relevant to the experiment are discussed (3) • A causal inference has been made (4). To be awarded Merit the student would need to provide more reasoning for the hypothesis, describe in more detail the experimental plan (in particular the treatment variable) and correctly interpret the strength of evidence.

Will leading questions influence the treatment groups to give a specific answer? So will changing how a question is worded affect the way our data is produced meaning answers will be higher or lower? According to this website http://www.busreslab.com/index.php/articles-and-stories/research-tips/general-research-tips/leading-questions/ a leading question is one which attempts to guide the person's answer. You are supposed to avoid using leading questions in questionnaires so that you get truthful answers.

I investigated whether people will be honest about how many times they used their phone in class. If I remind them that it is against the school rules - I think that they will give lower numbers. 1

Our experimental units will be two Year 12 physics classes of 58 students. Out of the 58 student 29 students will receive one survey and the other 29 will receive another, at random we will be handing out the survey.

One group will receive a survey asking "Even though using a cell phone class is against school rules, how many times did you use your cell phone in class last week?" and another group of students will receive a survey asking "How many times did you use your cell phone in class last week?" 29 students were given the survey which had the leading question about 'school rules' in it and the other 29 were given the question without any mention of 'school rules'. This makes the experiment design one of comparing two independent groups.

The response variable for our experiment will be the number of times the student writes down they used their cell phone in class over the last week. We chose the last week for the question so we would get range 2 of answers – if we had just said yesterday, then maybe the answers would only range between 0 and 5 times (one time per lesson). The variables that we can control for the experiment include the following:

- Same test conditions
- Same time

The students will be given the same time of day to complete the survey. I will be telling both the groups the rules before handing out the survey so students don't copy other student's answers. They will also be told to hand in the survey straight after they have finished answering the questions so they don't change their answers.

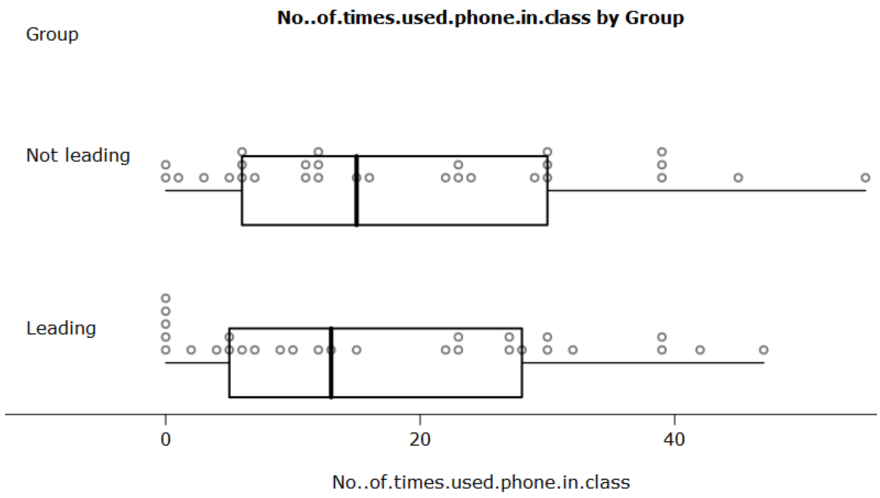
Variables that we can't control include the following:

- A student's memory is a variable we cannot control because some students won't be able to remember how many times they used their cell phone in class.

The uncontrolled variables will be randomly assigned to the treatment groups to balance them.

The experiment will be conducted in the following way:

- The classes will randomly split evenly in to 2 groups. Two students will be handing out the survey and one student will collect them in.

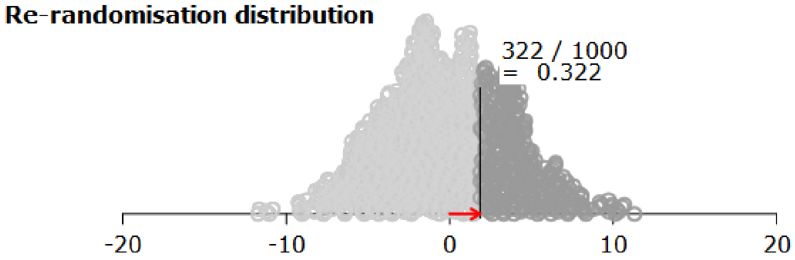


Summary of No..of.times.used.phone.in.class by Group

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
Leading	0	5	13	17.14	28	47	14.613	29
Not leading	0	6	15	19.00	30	55	14.820	29

The graphs and statistics above are from the experiment data. It shows the difference of 2 between the medians of the two groups, and a difference of 1.86 between the means, which are pretty similar. Referring to the graphs above the group that got the leading question has a lower median and mean than the "not leading question group", but not by much. The distribution of the number of times a phone was used in class for both groups is skewed, so the mean for each group is different from the median for each group. The two groups look pretty similar in terms of their spread and shape. The groups look like what I would expect just by chance, and there isn't anything I can see in the data to suggest that the leading question caused answers that tend to be lower. 3

To confirm this, I used the randomisation test with the means. This will re-randomise the answers to the two groups (leading question, no leading question) and record the difference between the means of the two re-randomised groups each of the 1000 times.



A difference of 1.86 or higher came up 332 times out of 1000. 4

In my investigation I found that the leading question I used wasn't effective in changing people's answers. I could have got a difference between the means of the two groups this size by chance without me doing anything (just by shuffling up the groups) because the value was 33.2%.

Appendix

Questionnaires/tools used:

Sighted by teacher but not provided

Results:

Group	No. of times used phone in class
Not leading	30
Not leading	39
Not leading	0
Not leading	11
Not leading	6
Not leading	29
Not leading	22
Not leading	5
Not leading	55
Not leading	6
Not leading	39
Not leading	3
Not leading	1
Not leading	12
Not leading	39

Group	No. of times used phone in class
Not leading	30
Not leading	24
Not leading	11
Not leading	45
Not leading	7
Not leading	6
Not leading	23
Not leading	0
Not leading	23
Not leading	30
Not leading	12
Not leading	12
Not leading	15
Not leading	16
Leading	42

Group	No. of times used phone in class
Leading	0
Leading	13
Leading	23
Leading	27
Leading	4
Leading	2
Leading	0
Leading	47
Leading	39
Leading	7
Leading	6
Leading	0
Leading	9
Leading	27
Leading	0

Group	No. of times used phone in class
Leading	30
Leading	0
Leading	23
Leading	5
Leading	32
Leading	22
Leading	5
Leading	28
Leading	39
Leading	30
Leading	10
Leading	15
Leading	12

	Grade Boundary: Low Achieved
5.	For Achieved the student is required to conduct an experiment to investigate a situation using experimental design principles. This involves showing evidence of using each component of the investigation process. There is evidence of conducting an experiment to investigate a situation using experimental design principles, showing evidence of using each component of the investigation process (1). For a more secure Achieved grade the student would need to indicate how the surveys were distributed, provide detailed steps for the experimental plan and provide a more detailed answer to the question.

Student 5: Low Achieved

Question: Does the order of questions in a questionnaire affect a student's answers to the questions?

Plan:

My questionnaire was about how many Facebook friends a person has. I wanted to test whether mentioning a link between Facebook friends and popularity first will affect the number of Facebook friends that participants put down.

To investigate this I will use two Year 13 classes as my experimental units. There will be 53 students, 30 female students and 23 male students who will be split up randomly into two groups.

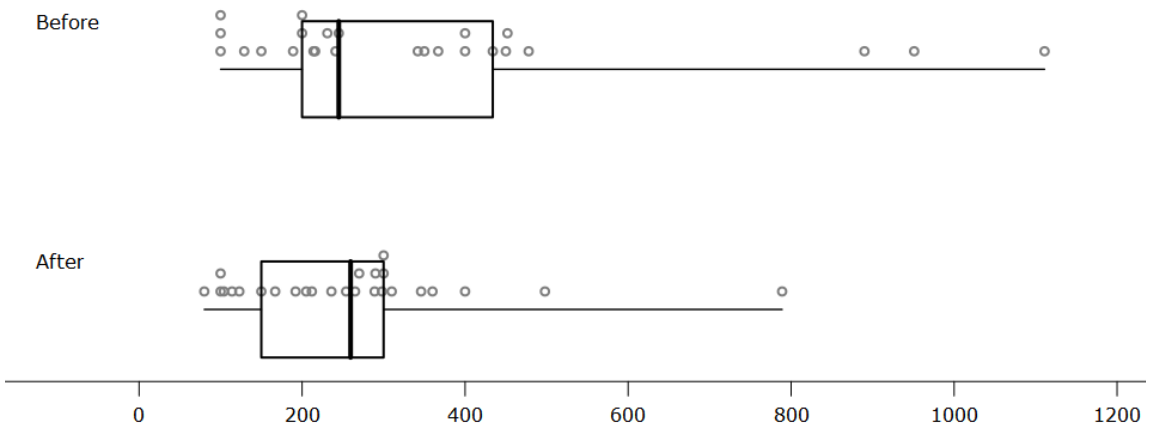
Two surveys have been typed up particularly with question order in mind. One asks if popularity is defined by the amount of Facebook friends a person has then is immediately followed by a question that asks the user to state how many Facebook friends they have. The other reverses this order, asking how many friends a person has first, then asking if the number of Facebook friends defines popularity. 1

The treatment variable for my experiment is whether the popularity question is before or after the question asking how many Facebook friends a person has. Students only complete one survey.

The two surveys will be carefully distributed to students in the class, by members of the experiment team in order to control the number of boys and girls who receive each kind of survey removing any possible variables that might arise from the difference in gender.

The response variable from my experiment will be the amount of Facebook friends listed by participants. To ensure that students do not discuss the question order of their surveys or how many friends they think they might have, participants are made to fill out the survey in silence without any communication to those around them.

The data from the surveys was typed into a spreadsheet.



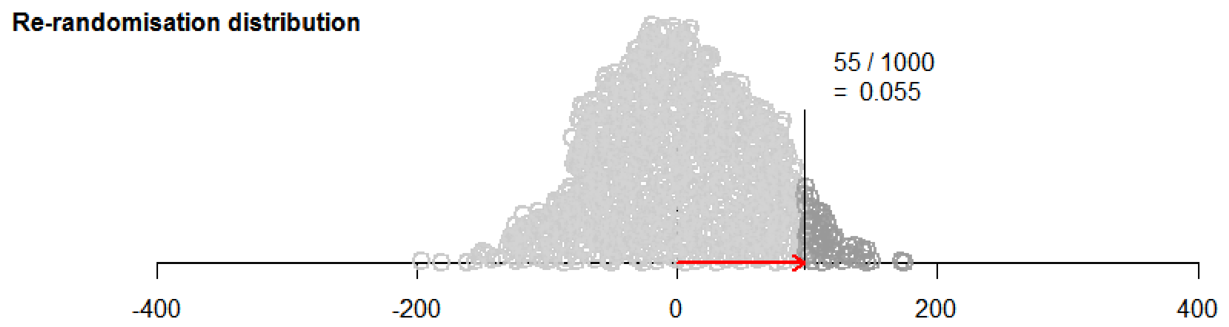
Summary of number by When

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
After	80	154.2	259.5	259.7	300	789	149.910	26
Before	100	200.0	245.0	357.6	434	1111	266.181	25

There is a gap in the data for the group that got the popularity question first, with three very high numbers for Facebook friends (890, 951 and 1111). I don't think the two groups look like what I would expect purely by chance with these sized groups. The medians for the two groups are not that different, with only 14.5 friends different. The means are much more different (difference of 97.9 friends). There is a greater spread in the answers for the number of Facebook friends for the group that got the popularity question first. I noticed when I entered that data from the surveys that some people were giving numbers to the nearest hundred or ten, which makes me think they were estimating the number of friends.

I carried out the randomisation test with the means. The test will see if the difference I got between the means of the two groups is likely by chance.

1



The difference of 97.9 only came up 55 times in the 1000 re-randomisations. Based on the evidence (the randomisation test results) I would say that the order of questions does affect answers. Having the popularity question first caused answers for the number of Facebook friends that tended to be higher then when the popularity question was second.

Appendix

<u>Questionnaires/tools used:</u>

Sighted by teacher but not provided

Results:

When	number
Before	100
Before	400
Before	478
Before	100
Before	0
Before	100
Before	200
Before	400
Before	951
Before	214
Before	434
Before	200
Before	1111
Before	0

When	number
After	100
After	289
After	300
After	205
After	310
After	360
After	100
After	270
After	80
After	290
After	400
After	150
After	300
Before	890

When	number
Before	450
Before	367
Before	150
Before	216
Before	350
Before	245
Before	129
Before	342
Before	231
Before	241
Before	452
Before	189
After	167
After	265

When	number
After	123
After	254
After	212
After	192
After	498
After	789
After	346
After	114
After	104
After	236
After	298

Removed the 0 friend responses before analysis as these people did not use Facebook.

	Grade Boundary: High Not Achieved
6.	In order to achieve the standard the student is required to conduct an experiment to investigate a situation using experimental design principles. This involves showing evidence of using each component of the investigation process. The student has posed a question, attempted to carry out an experiment, and has answered the question (1). To be awarded Achieved the student would need to clearly define the variables and the method of allocation of the treatment to the groups, and provide more evidence to support the conclusion.

Student 6: High Not Achieved

Teenage Relationships

What year level would students think would be an appropriate level to be in a relationship? Can I influence people to say a higher level? I decided to investigate question order in questionnaires and how this might affect the answers. 1

In this investigation our group will be using two year 13 statistics classes as the experimental units. In total there were 63 students, 25 male students and 38 female students. This experiment consisted of two random treatment groups. Both groups were determined by the questionnaires that were given to them. One treatment group had the bias question "Do you think Year 10 students are too young to be dating?" before a question asking “At what year level do you think students are old enough to be dating?” and the other treatment group received just the question “At what year level do you think students are old enough to be dating?” without the bias question about Year 10 students being too young to be dating.

In our investigation we conducted the 63 students each received a questionnaire with four questions on teenage relationships. After answering the questions, the questionnaires were then collected and recorded on a table. In our questionnaire the questions asked were referring to Year 10 students but we later found out that it was Year 13 students as our experimental units. This was one of the first variables that we could not control. This also included the amount of male and female students. This produced an uneven amount of results with 25 males and 38 females taking part in the experiment. Also other variables that cannot be controlled are the weather conditions. Both of the two treatment groups were assigned at random with a mixture of year 13 boys and girls.

The weather conditions was sunny and hot, which inside the classroom was stuffy from the amount of students inside the classroom but was also cool because of the fans. This also affected the way students answered our questionnaire, due to the weather it may have caused some of the experimental units to not be bothered answering our questions and write down anything (too hot and tired!) And other students were interested in seeing new faces in their classroom. Not only were half of the experimental units working on their own work on the computers and the other small amount of students were conversing this showed that this could have also impacted the end results of our experiment.

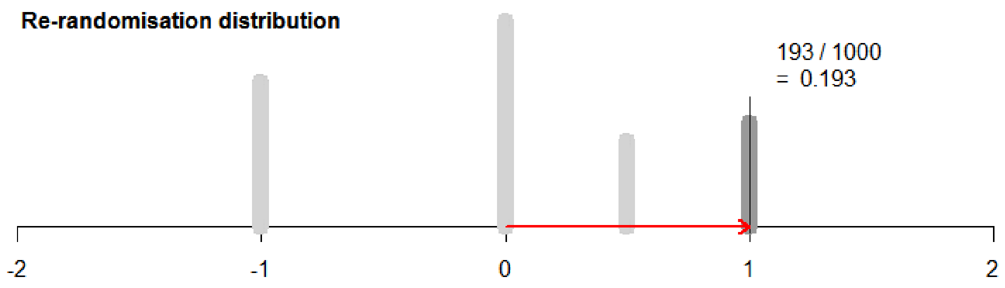
Variables that can be controlled in this experiment are doing it at same time of the day. This will show no unbiased in the amount of time given for the questions to be answered and it will be fair on the students in their participation in the experiment.


```
Summary of Year.started.dating by Bias.Question
  Min. 1st Qu. Median  Mean 3rd Qu. Max. Std.dev Sample.Size
N    9      11     11 11.06      12   13   1.063          31
Y    9      11     12 11.72      13   13   1.114          32
```

I notice the median and mean are higher for the group that got the bias question, with the median being one year level higher for the group that got the bias question. There is a greater spread of year levels given for the group that got the bias question – you can see this with the wider box and the higher standard deviation. Over a third of people gave Year 11 for the year they think students should start dating if they did not get the bias question.

1

I used the randomisation test for the medians of each group, and the results are shown below:



The difference of one year came up 193 out of 1000 times when the groups were re-randomised.

In conclusion, I would have to say that I was not able to influence people to say a higher level for when students are old enough to start dating by having the bias question before this question.

Appendix

<u>Questionnaires/tools used:</u>

Teenage relationships survey A

This survey is being conducted as part of the Statistics programme. Please answer the questions honestly and sensibly ☺. Please circle your answers.

Your gender	Male		Female			
Please circle your year level	9	10	11	12	13	
Are you in a relationship with someone?	Yes	No				
Do you think Year 10 students are too young to be dating?	Yes	No				
At what year level do you think students are old enough to be dating?		9	10	11	12	13

Teenage relationships survey B

This survey is being conducted as part of the Statistics programme. Please answer the questions honestly and sensibly ☺. Please circle your answers.

Please circle your gender		Male		Female		
Please circle your year level		9	10	11	12	13
Are you in a relationship with someone?		Yes	No			
At what year level do you think students are old enough to be dating?	9	10	11	12	13	

<u>Results:</u>

Bias Question	Year started dating
N	11
N	11
N	12
N	11
N	9
N	12
N	11
N	11
N	11
N	11
N	11
N	12
N	13
N	11
N	13
N	11

Bias Question	Year started dating
Y	12
Y	12
Y	10
Y	11
Y	12
Y	12
Y	11
Y	13
Y	10
Y	9
Y	10
Y	13
Y	13
Y	12
Y	13
N	10

Bias Question	Year started dating
Y	10
N	9
Y	12
N	12
Y	13
Y	13
Y	12
N	11
N	11
Y	11
Y	12
Y	12
Y	13
N	10
N	11
N	9

Bias Question	Year started dating
N	10
N	12
Y	12
Y	11
Y	13
N	13
N	12
Y	11
Y	11
N	11
Y	11
N	10
N	11
Y	13
Y	12

<u>Experiment notes:</u>

Incorporated into report.