Science – Investigating Science – Evaluating first hand investigations transcript  
   
(Duration 21 minutes 41 seconds)

(gentle music)

Instructor: Hello, and welcome to this presentation on planning and conducting investigations. This presentation was put together for Investigating Science, but it may be used for any Stage six science subject. Firstly, I would like to pay my respect and acknowledge the traditional custodians of the land on which this meeting takes place, and also pay my respect to elders both past and present.

This presentation accompanies the document ‘Investigating Science guides and scaffolds for secondary source and first-hand investigations’. As indicated by the title, we will be looking at secondary source research and how to evaluate secondary information. We'll be looking at how to evaluate quantitative data and we will also look at some scaffolds and self-evaluation checklist for firsthand investigations. The page numbers you see listed, there are those from the accompanying document.

So now we're moving onto the second topic; evaluating data. So first up, we're going to look at accuracy, which can also be thought of as exactness. So the true value of a measurement cannot be accurately known. So really all measurements are estimations. So error is a difference between the measured quantity and the true value. So the smaller, the error, the closer the measurement is to the true value and the more accurate your measurement is. Now there are different types of errors. So, the first step we have random error. It's just normal, natural part of everyday life because we can't actually measure the exact true value. Now, the great thing with random errors is that when we take multiple measurements and take an average, that average will be closer to the true value. So that's why repeating measurements and taking an average improves accuracy. Now over here, we have a picture of a domino in the top side and we've put a ruler up. And as you can see, it's a bit hard to tell, is it actually, where is it between 43 or 44 millimetres? So the true value might be 43.24, but we'll never be able to measure that. We just know it's within that range. And so that's an example of random error.

Another type of error is called systematic error. So this is where there's some problem in the setup or method, and so you're out by the same amount every time. So if we have a look at the bottom domino here, instead of measuring the domino from the zero millimetre point, we're actually measuring it from the end of the ruler. So what that means is every single measurement is going to be out by the same amount. And it doesn't matter how many times we repeat the measurement, if we're using that same method, all our measurements will be out and we're not going to be able to improve accuracy by taking an average. And the third type of error is called a gross error. You might know of them as blunders. Often they're the cause of outliers in your data.

So next, we're going to take a quick look at precision. Now people commonly mistake this for accuracy, but it is a different thing. Precision is actually internal reliability. So this is probably what you think of when you're talking about reliability of an experiment because measurement precision is the extent to which repeated measurements made under the same conditions, agree with each other. In terms of instrument precision, so this is related to the error associated with the instrument itself. Now it's important to know that digital instruments do not always have greater precision or lower error than analogue instruments. So for example, you might have an electronic scale at school that fluctuates a lot when you're trying to measure something. So that has very low precision.

So this is another way to look at accuracy and precision. So we can see on the diagram here, this blue dot represents the true value. And then we are trying to measure this value and we can see the red Xs represent all the measurements for that value. Now, if we take an average of those red Xs, we get the average of the measured value. Now with what we've seen here, we can see that you've got a spread of those measurements and that's the spread of your random error. And in the middle of that is your average. Now because we have systematic error in this particular setup, that means our average and indeed all of our measurements, are actually quite a distance away from our true value. So if we want to try and minimise the systematic error, hopefully have none in the system at all. And when we do that, the red dot and all our measurements would shift over towards the left, towards the blue dot. And hopefully if we take enough measurements that our average value should actually be very close to our true value. So with your accuracy. Accuracy is the overall error in the system. So it's the random plus systematic errors. So the smaller the systematic error, the greater the accuracy, because you're shifting all these measurements closer to the true value. So, taking an average from repeat trials only increases accuracy if you have no systematic errors. So, if you have a systematic error and everything has shifted, like it is here to the right, it doesn't matter how many repeat trials you do. And you take an average, your average value is still going to be a long way from the true value. Whereas if we look at this graph in terms of precision, precision is all about random error. So the smaller the random error, the greater the precision. So the smaller the spread of these Xs, the more precise your measurements are. And precision is independent of systematic error because it doesn't matter what systematic error is in your system, precision is all about random error.

So next we're going to look at reliability. And what we mean by reliability is that our measurements are stable. So there's two parts to this. So we've got internal reliability, which is our precision that we've already talked about, but we also have external reliability. So reliable measurements are similar in value over multiple experiments. So to improve and evaluate reliability, you can do two things. So we can repeat an experiment and compare the results. So this is our internal reliability, this is our precision. This is what you often do in a class when you do repeat trials. What we also need to do is look at external reliability. So we need to compare our results with the group that carried out the same method. So if different groups also carrying out the same method, come to the same conclusions, have the same measurements, then we can say that our information is reliable.

So now we come to uncertainty, which is really a measure of accuracy and reliability. So earlier we said, accuracy is basically how close is our measured value to our true value and that difference is the error. Now, if we don't know our true value, we don't actually know what our error is, but we can actually come up with the best estimate of the magnitude of possible errors that we could see. And so that depends on a few things. So it depends on the smallest unit of the instrument, which we also call the limit of resolution or the limit of reading. So for example, if you've got a ruler, often the smallest unit is one millimetre. We also look at instrument precision. So how reliable is the instrument? So will it give you the same measurement time after time? And the other part is your method. So how have we set the equipment up? What's the procedure, have we followed repeat trials and then taking the average? So if we have a look in the video on the right, what we see here are little planes that are flying round and round in a circle, and the students had to measure the height of that plane above the floor. Now the smallest unit of that ruler is one millimetre. But as you can see, from the distance that we're trying to measure it from, you can't get to that limit of resolution. And similarly, it's very difficult to know where the bottom of the aircraft is and to be measuring at the same point in time. So even though we're using a really with one millimetre as a limit of resolution, the realistic uncertainty is probably about plus or minus five, about five centimetres. So you can only be confident that your value is somewhere within a 10 centimetre range. So when we talk about uncertainty, we talk about a measurement. Well, we're actually talking about a measurand, which is the quantity being measured or calculated. So our measurand is actually the best estimate for that quantity, plus or minus what our uncertainty is.

So how do we estimate uncertainty? So for analogue measurements, we start with plus or minus the smallest unit, and then we increase the uncertainty based on the methodology. So if we're using a ruler and we're looking directly on the ruler, our uncertainty is going to be much lower than if we use the same ruler, but we have to look at it from one metre away. So depending on the method, we'll change your level of uncertainty. Now for digital measurements, the preferred way is to actually look up the manual and find out what the error is in the manual. So sometimes it's usually presented as a percentage, for example, 2% at 10 degrees Celsius or 5% at 90 degrees Celsius. And interestingly, the error report in the manual is usually much higher than you think it would be based on the decimal points of that particular instrument. But if you don't know the precision from the manual, then we need to account for unknown rounding areas in electronics. So for example, when you have a scale and it might read 10.2 grams, we don't know if the actual value behind that is 10.21, 10.22. Sometimes electronics only round up and down. So in that case, 10.29 would be reported as 10.2. Similarly, it might be a different kind of electronics where 10.11 would also be recorded as 10.2. So, because we don't know how those particular electronics are going to work for that piece of equipment, we need to account for that. So for a stable reading, we take the measurement and we use plus or minus the smallest unit, becomes our uncertainty. But in some cases you might have a fluctuating reading in your digital measurement, so the precision is quite low and if your reading is fluctuating, you need to look at the range of the reading. So the highest value minus the lowest value and your measurement becomes the average reading and the mean for that. And your uncertainty is plus or minus half the range of readings. So often what happens is we think that analogue instruments have a lower level of accuracy or not as accurate as digital measurements, but quite often, when you look at the level of uncertainty, analogue measurements can actually be better than digital measurements.

So we've just looked at uncertainty for individual measurements, but what about repeat trials? So uncertainty is not just about the instrument and the methodology. So even the measurand itself may vary. Take a bottle rocket, for example, how consistent do you think the measurements would be? So, repeat trials allow you to estimate the size of the variation in the measurand, using the maximum and minimum values. So the measurand is the best estimate plus or minus uncertainty. So our best estimate is our mean or average value. Now there are two ways of figuring out what uncertainty is. If you're only using a few trials, if you've only got, say two to four trials, I probably recommend you use the method where you just find the maximum difference between a measurement from that mean value. But if you're using say five or more trials, an easier method is going be, just find the range of the measurements and take half of that as being our level of uncertainty. Now, just on a side note for science extension students, if you are doing more than 10 trials, uncertainty is calculated using the standard error formula. So hopefully you will know what that means in terms of standard deviation.

So now we come to validity, which is really looking at: do we have results to mean what we actually think they mean? So you need to look at the validity from two parts. You need to look at the method and how we draw conclusions. So to ensure that your method is valid, you can check that your hypothesis is directly relevant to your aim, but you really need to check that your method tests your hypothesis. So if your hypothesis involves the concentration of a certain gas, for example, but your method is looking at the amount of gas in a container, then there is a mismatch between your method and your hypothesis, and your research is not valid. You also need to check that your sampling method results in a sample that actually represents a wider population that you're interested in. And you also need to control your variables so that the test is fair, and the chance of any kind of sneaking or lurking variables affecting your results is minimised.

Now, in terms of your conclusions, you need to first of all check the accuracy and reliability of your measurements. If you don't have accurate and reliable measurements, then any conclusions that you draw are not going to be valid. You also need to identify the limitations of your data. So can you really generalise from that data or extrapolate to other situations? You also need to examine your underlying assumptions to check for possible bias. So you might interpret your data in a certain way based on what you're expecting to see. So you need to be aware of any confirmation bias. And lastly, you need to check the way that you represent your data, that you're not putting potential bias in it. So for example, you can choose the way you present your scales on any graphs to distort the impression of data. So you need to make sure that you haven't done that.

So this NESA Venn diagram looks at the relationship between accuracy, reliability, and validity. So validity encompasses everything. So it's a great big circle around the outside. We've got validity of methods. So we've got our aim, our hypothesis, does our method match? We're looking at sampling method, our control variables and experimental controls. Our validity of conclusions, we're looking at limitations in our data, any assumptions, our possible bias and our accuracy and reliability of our data. So in terms of accuracy and reliability, so the more reliable our sources. So we take repeat measurements and take an average. That average is going to be closer to the true value. So our uncertainty has reduced and our accuracy has increased. But remember, this is only for the situation where we have minimised or have no systematic error. If we have systematic error in our process, then the accuracy of the numbers is going be very low.

So in this last section today, I'm just going to very briefly present you with some scaffolds and self-evaluation checklist that will help you when you're conducting firsthand investigations. So this is a one page template that can help you plan first-hand investigations. So T stands for title. A is for aim, what do you want to investigate. BI stands for background information. So what is the relevant scientific information that will help you plan your investigation. V as you might've guessed is a variable. So we have IV, independent variable. We have our dependent variable, DV, and CV is our controlled variables. Now over to the right here, I've put x and y in the same so that you can try and see the relationships between, and the IV variable goes on the x-axis when you're graphing, the dependent variable goes on the y-axis when you're graphing and our controlled variables is what we keep the same. The next block here, it's for a control or a standard. So what is your control treatment? And if you can't do an experimental control, then what's your standard? What's your base case that you're going to compare everything to? H, as you may have guessed is for hypothesis. So within that, we want to predict the trend and it must include your independent variable and your dependent variable, and it clearly shows how they are being measured. Now we also have M and E, is materials and equipment. So what are you going to use? Then safety is your control measures and an explanation of why you're using those control measures. M is for method as indicated by the footsteps here. R is for results, so how are you going to present your results? What's your table going to look like? What type of graph are you going to use? And D is for discussion, so how, and you want to be thinking about things like what's the variation in your results? Are there any outliers? Do you have any trends or patterns in your data? And what might explain those? In your discussion, you should be looking at any problems that occurred and how you could or did solve these, and you'll be evaluating your method and your results for their validity, reliability, and accuracy. And also explain any changes that you would make if you had to do it again. And then lastly, here at the bottom, we've got our conclusion. Now it is really important that your conclusion links to your aim and your hypothesis. It is a short summary of your findings and how they relate to the aim. You do not include specific results here, but you can include the trends. And you should say whether your observations or results support or disprove your hypothesis.

So this is a working scientifically self-evaluation checklist. As you may notice here, each of these sections are broken up according to the working scientifically outcomes. And so students use these in class in a couple of ways. They use it to help them plan, to make sure they're ticking off all the things they need to do, but they also use it after an experiment so they can see what did they do well, and what do they need to focus on for improving next time. So my students find these really useful to use themselves in class, but also as their teacher, sometimes we'll have a specific experiment and they will just run with it, while I sit back and monitor and I'll monitor their progress using this checklist as well.

So now we come to the very last part of this presentation today, the report checklist. So this is a checklist to show students what they need to have in a band six type report for the first-hand investigations. So for example, as we go through here, do they have an aim? If they've only got an aim that's more looking at a Stage five kind of report and Stage six, I'd be expecting them to have an abstract. So a one page, a one paragraph summary of what they've done. Similarly, if it was a Stage five student, you'd be looking at background information, but as you move through Stage six, I would be expecting by the time you're hitting year 12 to be doing a literature review so that includes in-text citations.

And then have you got all your variables? Do you have an experimental control? It's not always applicable. What is your hypothesis? Have you clearly identified your independent and dependent variables in that hypothesis? Your list of materials and equipment. Safety. So for my Stage six, I would be expecting them to do a proper risk-assessment. They can refer to the risk-assessment template, that's in your guide, which is fantastic, but the students really need to be able to explain the control measures that they've put in place. Now next of all we've got a method or a methodology. So a method is a step by step method that you used in Stage five science. But once you've done module five and you've looked at methodologies, with my year 12 students I expect them to write a methodology rather than a method. So this is like an essay style summary of what they did, but it still needs to be reproducible by an expert. So you still need to have enough detail in there that someone else could reproduce it exactly.

And then we've got our results section. So we're looking at what are observations, tables, graphs and also identify trends. So then in our discussion for a Stage five student, a good Stage five student you'll be looking at describing trends and patterns and using background information to explain them. You'll be looking to describe the variation in the results and identify any outliers and how you may have handled that in the analysis. Look at problems and sources of error and explain what changes you would make to improve the data collected. Now for Stage six students, we're looking at having in-text citations in our discussion, you will be evaluating validity by looking at: did the test measure what they intended to measure and is the data analysis and actions all accurate?

Then we're looking at evaluate reliability by looking at the consistency of the results of change. So ideally you would also be comparing your results to published information as well, but also looking at the measures taken to reduce error there. Now, if you're in Science Extension, Investigating Science or Physics, you're expected in the syllabus to be able to assess accuracy. So that means identifying uncertainty and you're looking at random and systematic error. Now, if you're a Science Extension and Investigating Science, you need to get beyond that. You need to be looking at the potential real world benefit or consequences of this particular research. And then last way we come to conclusions. As I said before, they need to show the overall findings and how they relate to the aim. You need a judgment of whether the results support or disprove the hypothesis and include trends but don't mention specific results. And lastly, we have a few, if you are Stage four or five level, you might be just be doing a bibliography or a list of secondary sources. But if you're in Stage six and this is an assessment task, I'd be expecting a list of all cited sources in an appropriate format.

So that brings us to the end of the presentation today. I know it was a long one, but I hope you found it useful. I just wanted to acknowledge a couple of documents that are on our science website from the Department of Education. So Evaluating scientific data is a great source for looking at how to evaluate data. It also adds a bit more information in terms of how to handle calculations involving uncertainty. And the other document I want to talk about here is Developing a literature review. So this was designed more for science extension students, but it is equally applicable to other Stage six science subjects, right? And that's it for me for today. Thank you for listening. (upbeat music)

End of transcript