

How to make your research useful and trustworthy – the three U’s and the CRITIC

Michael Wood
University of Portsmouth Business School
<http://woodm.myweb.port.ac.uk/SL/researchmethods.htm>
August 2015

Introduction.....	2
The three U’s.....	2
The CRITIC	2
C (Causes).....	3
R (Representative)	4
I (Indicators).....	6
T (Triangulation).....	6
I (Imaginative)	7
C (Chance)	7
Exercises	8
Appendix: choosing a random sample	9
References.....	10

Introduction

Checking usefulness and trustworthiness is important for checking the credibility of published research (and to help you do a *critical* literature review), and as a checklist for your own work. My suggestion is that you should check the three *Us* below, and the *CRITIC* – in the sense of the list below. The concepts here are standard ones which should be covered in most research methods texts, although I have tried to avoid jargon where it does not seem helpful. The three *U*'s and the acronym *CRITIC* are my invention with some help from an MBA group in late 2009.

Please treat them as a checklist for your research and other research. Checklists are routinely used by pilots to make sure they remember to fill up with fuel before taking off. A similar checklist might stop you doing silly things in your research.

The three *U*'s

First check the *three U*'s. The research should be:

- As User-friendly as possible. If people can't understand it, what's the point? Is it as simple as possible? Use short sentences and short words whenever possible.
- As Useful or interesting as possible. Are there practical conclusions that will make the world a better place? Is it fascinating? Does it go beyond the obvious? Or does it just leave you thinking "so what?"
- As Uncritizable or TrUstworthy) as possible. Trustworthiness or credibility is particularly important. Can you trust the conclusions? Are they right? Do you believe them? Are there any flaws? It's essential to give readers enough detail to check.

The *CRITIC*

Trustworthiness is more difficult to assess than the other two. The main issues, as I see them, are summarized by the acronym *CRITIC*. Each letter stands for something that needs checking to assess the trustworthiness of research. I think they are, roughly, in order of importance, with the first three (*CRI*) being particularly important, but you may disagree. Some may not be relevant to some projects, but they are all worth checking.

I have avoided using jargon where it doesn't seem helpful. A lot of the points below are described in terms of various forms of *validity* in the textbooks – *C*, *R* and *I* correspond to internal validity, external validity and construct validity. "Threats to validity" are simply reasons why your research may not be trustworthy, which is what the *CRITIC* is all about.

But does this really matter? Yes! Research which is wrong, or can't be trusted, wastes time, and may lead to harmful decisions.

C (Causes)

Useful recommendations for the future depend on understanding what **Causes** success and failure. For example

- If sales go up at the same time as a new marketing strategy is introduced, did the new strategy *cause* the increased sales, or was it something else like a new manager, or better market conditions, or something you haven't even thought of?
- A survey (Heras et al, 2002) revealed that firms with ISO 9000 registration tend to be more profitable than similar firms without ISO 9000. Is this because ISO 9000 *causes* firms to become more profitable, or perhaps being more profitable means the firm can afford ISO 9000 registration?
- A survey in Japan showed that moderate drinkers tended to be slightly more intelligent than people who drank no alcohol (see Hadfield, 2000). Does drinking *cause* increased intelligence, or does increased intelligence cause drinking, or does something else cause both? Will telling non-drinkers to drink make them more intelligent?
- What *caused* the high death rate in Iraq since the 2003 war?

If something is a cause of something that matters to you (good or bad), it may be useful to change it. If it is not a cause, changing it will make no difference, so it is important to understand what causes what. This may be difficult, or even impossible, because there may be a very large number of possible causes, all interacting in complicated and unpredictable ways.

Possible methods for investigating causes are experiments (controlled trials) and quasi-experiments, looking at how things change over time, and common sense and logical reasoning (understanding *how* the marketing strategy might work may help to work out if it is a likely cause of better performance – although you may be wrong).

Sometimes just doing research may have an effect on what you are researching. The well known *Hawthorne effect* (http://en.wikipedia.org/wiki/Hawthorne_effect) refers to the fact that people may change their behaviour in response to the fact that they are being studied. If you decided to find out how much time workers spend on activities not related to their work by sitting in the room and observing them, the workers in question will probably reduce the amount of time spent on non-work-related activities *because* they are being watched. So doing this research by observation would be a waste of time.

Remember that it is very easy to jump to the wrong conclusions about causes. People tend to like simple stories about one event causing the next because it helps to make sense of the world and predict events, but these simple stories may be wrong (see Taleb 2008 for much more on this theme). Take care!

R (Representative)

Is the sample used **Representative** of the wider context? Is it reasonable to assume that conclusions from your sample apply to the whole population, or to other times and places? Or are your results likely to be *biased* because of the sample of data used?

- If research is based on interviews with three people, to what extent can the conclusions be assumed to apply to other people? Obviously you need to think about (and discuss in your research methods section) which three people you chose, and why.
- Questionnaires are sent out to a sample of 1000 employees in an organization of 10,000 employees. The researcher bases her analysis on the 100 questionnaires that are returned. Can we assume that the results reflect the views in the organization as a whole?
- A study of a construction project comes to conclusions about problems with the project. Can we assume the conclusions apply to all construction projects? To all projects?
- A study of the problems involved in setting up a small business is based on interviews with founders of sample of small businesses (which are still in business). Will this sample lead to any biases in the results?

Almost all research is based on taking a *sample* and using this to come to conclusions about a wider context, or broader *target population*. (The sample and the population might be people, organizations, projects, etc.) You need to think hard about this because many samples are biased, often in ways that are not obvious.

Taleb (2008, p. 308) refers to this as the problem of “silent evidence” – in the last example above, for example, if we only talk to founders of small businesses that have succeeded, we will never find out about the “silent” ones who have gone out of business. The sample is biased, and we will end up with misleading conclusions about setting up a small business. Similarly, if we want to know if a new drug works, we might search the literature for articles reporting clinical trials of the new drug. Suppose we find three articles reporting results that indicate the drug works, and none reporting results that suggest it does not work. At first sight this seems to show the drug works, but be careful! The danger is *publication bias* – the drug company may have simply not reported negative results, or the journals may have rejected papers reporting negative results because they are not as interesting as positive results.

There are two approaches to designing a suitable sample. First, use a *convenience* or *opportunity sample* (i.e a sample which is convenient – perhaps your colleagues, or organizations you know), and then ask yourself how far the results can be generalized. Statistics may help, but the main guide is common sense. But it deserves careful discussion.

The second, better, approach is to define your target population, and then choose a sample which is likely to give as good information as possible about this target population. *Defining the target population is important: don't forget this.*

There are two important approaches to choosing a sample from a target population:

- 1 *Random sampling*. This means each member of the target population has the same probability of being selected, and each member is selected separately so you don't end up with a cluster of similar people etc in the sample. This is an example of what is called *probability sampling*. Most statistical methods are based on the assumption that you have a random sample. If you have not taken a random sample from a target population, you need to consider the interpretation of your results carefully. (See the Appendix for how to take a random sample.)
- 2 *Purposive sampling*. This means you choose the members of your sample for definite reasons relating to your research. These reasons obviously need careful consideration. This is an example of what is called *non-probability sampling* for obvious reasons.

There are many other approaches to sampling – see any book on research methods – but in practice, these two, and *convenience sampling*, are the main contenders.

For a large survey, random sampling is usually best, especially if you want to analyze the results statistically. With a large enough sample, random sampling should ensure that you have a sample which adequately reflects the pattern of the population from which it is drawn. Notice that, if for example, you are sampling people, you don't need to worry about variables like male/female, income level, etc, etc because the random process will ensure you get a mixture which reflects the mixture in the population. However, in practice, there are often problems. Selecting samples for political opinion polls, for example, is difficult, but very important if the results are to be accurate – see http://eprints.ncrm.ac.uk/914/1/MethodsNews_Spring2010.pdf . One important issue is non-response. You may have chosen a random sample, but if most of the people you've chosen don't respond you end up with a biased sample and not a proper random one. It is good practice to give the *response rate* when you report your results. A response rate of 20%, for example, means that only 20% of the people you chose in your sample responded – this is fairly typical, but should be a warning that your results may be seriously biased.

For a small sample, purposive sampling is usually best because you cannot rely on the random numbers giving you examples of each type, so it is best to consider carefully the aims of your research and plan your sample accordingly.

As well as the approach to selecting the sample, you should also consider the size of the sample. In general, the bigger the better: the main constraint is your time. If you are analyzing the data statistically, statistical methods can help decide how large a sample you need (see <http://woodm.myweb.port.ac.uk/stats/statnotes3.pdf>).

How you choose samples of data is very important. Badly chosen samples may give biased and misleading results. *There should always be a discussion of the sample, and the how it was selected, in the section on research methods. This is especially important if*

the sample is small! If you are doing one or two case studies as representatives of a wider population, then you must choose the cases very carefully. Or, if you have no choice about your cases, you need to think hard about how far your conclusions can be generalized.

I (Indicators)

Are the **Indicators** used to measure or assess characteristics of interest OK? If you want to find out about performance, or customer satisfaction or complaints, or quality or profits, you must have a sensible way of measuring them, or assessing them in some other way.

- Moreno-Luzon (1993) used managers' "perceived achievement of objectives" as a measure of organizational performance. Can you see any problems with this? Do you think the answers are likely to be biased?
- How would you measure quality of service in a casino?
- IQ (intelligence quotient) is supposed to measure intelligence. Or does it just measure the ability to do stupid tests?
- Do economic indicators like GNP give a good measure of standard of living?

Sometimes it may be possible to measure something as a number. Then you need to consider

- the *validity* of the measure – is it valid in the sense that it really does measure what it is supposed to measure – and
- the *reliability* – it reliable in the sense of giving *consistent* results at different times or with different judges or different people.

Checking validity is largely common sense – the obvious thing is to compare the measurements with other approaches to assessing the same thing. There are various statistical approaches to reliability (see a textbook) – if you need to use these you will need to check carefully that these make sense for your particular measurement.

On other occasions a numerical measurement may not be realistic. But you still need to consider whether the method you are using is assessing the right thing and doing it accurately.

If you can find a way of measuring something which has been used by another researcher, it is usually a good idea to use it – with acknowledgement and, if appropriate, permission. This means the other researcher should have checked validity and reliability, and it should be easy to compare your results with the other researcher's.

T (Triangulation)

Triangulation – compare data and results from different sources. Applies to data, methods, observers, theories (e.g. Robson, 2002: 174). If your interviewees say they are happy and relaxed at work, and this is backed up by physiological measurements, having two independent sources of data obviously makes your evidence much stronger.

I (Imaginative)

Is your research sufficiently **Imaginative**? This is important for at least three reasons:

1. Have you thought of all the possible hypotheses for explaining your results. Perhaps an important cause of poor performance is diet or the weather – these are the sort of possibilities which you may not think of at first.
2. You may need imagination to think of recommendations which may move the business forward in new and exciting ways. You may have researched what is being done at the moment, but possibilities for the future that nobody is trying at the moment may be more interesting (see <http://woodm.myweb.port.ac.uk/NotHappened.ppt>). This is difficult to research, of course!
3. And you will need imagination to think of the best way of doing your research – many projects use questionnaires when more imaginative approaches may be more effective.

C (Chance)

Have you taken account of the possibility that your results may be due to **Chance**?

- If you talk to three people who all have the same problem, it is easy to assume that this is only problem that matters. Another three people might raise completely different issues.
- If the average job satisfaction rating for one department is greater than for another department, this might be just due to chance, to the people who happened to be in your sample. Another sample might give a different result.

Statistical null hypothesis tests and confidence intervals are designed to deal with this problem (<http://woodm.myweb.port.ac.uk/stats/statnotes3.pdf>). Use them when necessary. Don't forget, however, that the problem is likely to be worse with small scale, qualitative, research.

One further, general, tactic is the use of a devil's advocate or critical friend. Remember the problem of *confirmation bias* – you are likely to be more enthusiastic about evidence that confirms your pet ideas than about evidence that undermines it (for some examples, see Taleb, 2008)! Get someone to try and be critical and find difficulties with your research – then fix or (if unfixable) acknowledge the problems.

Can you think of anything important that I have left out? Please be *critical* of this list; it is just based on my fallible ideas. Let me know of any objections to anything here, and anything important I have left out. Be my devil's advocate!

Acknowledgment: I am grateful to Colston Sanger for some helpful comments.

Exercises

1. Suppose you are asked to investigate whether introducing a proposed new IT system in an organization would be a good idea. Clearly you want your recommendations to be based on reliable information about what effects are likely to be *caused* by introducing the new system. What research methods could you use to assess the impact of the new system? What difficulties do you foresee?
2. What caused the latest credit crunch? Does your answer lead to recommendations? Will your recommendations work if your answer about the causes is wrong?
3. Research (e.g. Glebbeek and Bax, 2004) has found a negative relationship between staff turnover and organizational performance (higher staff turnover tends to be linked to lower performance). Do you think higher staff turnover *causes* lower performance, or vice versa? Does it matter? How would you find out?
4. Choose a random sample of 10 companies from the FTSE – 100. Under what circumstances would it be useful to use a random sample? When might you use a purposive sample?
5. How would you select a random sample of 50 staff from Portsmouth University?
6. Suppose you wanted to find out how satisfied people in Portsmouth are with the local NHS medical services. How could you choose a sample to ensure that it is as representative as possible? (Explain how the practical process of choosing people to ask would work.)
 - How big would your sample be?
 - How representative is the sample is likely to be?
 - How easy would it be to carry out your method of choosing a sample? If it's too difficult, think of another method!
7. The National Rail Enquiries service is a phone line for people to ask questions about train journeys in the UK. A few years ago there were concerns about its accuracy, so some surveys were conducted.
 - A Consumer's Association survey used a sample of 60 calls, mainly about fares. The worst mistake was when one caller asking for the cheapest fare from London to Manchester was told £162 instead of the cheaper £52 fare which was available via Sheffield and Chesterfield. The percentage correct was 32%.
 - A reporter rang four times and each time asked for the cheapest route from London to Manchester. The proportion of the four answers which were correct was 25%.
 - An NRE sponsored survey found that the answers were 97% correct.

What do you think of these results? How would you have organized the survey? What would your target population be?

8. It is often assumed that a good way to get a random sample of people is to interview people “at random” in a shopping centre.
- What might the target population be for this approach to sampling?
 - Is the resulting sample likely to be representative of the target population?
 - Do you think this is a satisfactory method of sampling?
9. How would you measure or assess (for research purposes):
- The quality of service in a casino?
 - Job satisfaction?
 - The strength of a brand?
10. From an email circular ... “here's the final word on nutrition and health:
1. The Japanese eat very little fat and suffer fewer heart attacks than the English.
 2. The Mexicans eat a lot of fat and suffer fewer heart attacks than the English.
 3. The Chinese drink very little red wine and suffer fewer heart attacks than the English.
 4. The Italians drink a lot of red wine and suffer fewer heart attacks than the English.
 5. The Germans drink a lot of beers and eat lots of sausages and fats and suffer fewer heart attacks than the English.”
- What does this suggest causes heart attacks?*

Appendix: choosing a random sample

The standard method is

- Make a numbered list of the target population. This is called the sampling frame. Sometimes it's easy – e.g. if your target population is the employees in an organization, you may be able to use the telephone directory. Sometimes it's not easy (e.g. if the target population is everyone living in Portsmouth).
- Use random numbers to choose the sample. (Excel can generate random numbers, or you can find them in some books – e.g. Saunders et al, 2007.)
 - If the same random number comes up twice, ignore it the second time
 - Each member of population has the same chance of being selected
 - Each member of the sample is selected independently (so you won't end up with a cluster of similar individuals)
 - In practice, it is likely that some members of the sample can't be found or won't help, so the sample may be biased towards those that are easy to find. This is often difficult to deal with. Do the best you can.

The principle is to ignore all variables and choose at random. This allows for all variables, including those you haven't thought of. It also means that you cannot be accused of choosing a sample that you think will give you the answers you want!

Stratified sampling is a similar idea. Here you divide the population into groups called strata – e.g. different departments – and then take a random sample for each group. The size of the sample from each group should reflect the size of the group. In theory this is a slightly better way of sampling; in practice the difference is often slight. Do it if it's easy; otherwise don't bother.

References

Glebbeek, A. C., & Bax, E. H. (2004). Is high employee turnover really harmful? An empirical test using company records. *Academy of Management Journal*, 47(2), 277-286.

Hadfield, P. (2000). Drink to think. *New Scientist*, 9 December, p. 10
(<http://www.newscientist.com/article/mg16822681.400-drink-to-think.html>).

Heras, I., Dick, G.P.M., & Casadesus, M. (2002). ISO 9000 registration's impact on sales and profitability: a longitudinal analysis of performance before and after accreditation. *International Journal of Quality and Reliability Management*, 19 (6). 174-191
(<http://tinyurl.com/ydxn8jh>).

Moreno-Luzon, M. D. (1993). Can total quality management make small firms competitive? *Total Quality Management*, 4(2), 165-181.

Robson, C. (1993 and 2002). *Real World Research*. Oxford: Blackwell.

Taleb, N. N. (2008). *The black swan: the impact of the highly improbable*. London: Penguin.