IMT4110 - Assignment 1

Ole André Hauge

September 2021

Poor use of the scientific methodology in academic works can undermine the findings and importance of the study. The methodologies contain a set of tools researchers and academics can use to properly structure and formulate research to ensure causality, reliability, and validity. In doing so they can conduct objective and standardized experiments that minimize the influence of bias in the research. As researchers work in different fields and have different backgrounds, it is important to use scientific methods properly to ensure that the research can be read, understood, and re-validated by other peers. When the researchers apply the scientific methodologies to their work they have to consider what variables they are working with. In this paper, we will focus on the dependent and independent variables. A dependent variable is the subject of the study which receives stimulus from the test. An independent variable, on the other hand, is something the researcher introduces to the study e.g. what he or she wishes to measure the effect of. Let us say we want to measure if sex affects how a person experiences pain when a person is inflicted by cold temperatures of -20 degrees over time. In this case, the girls, boys, and constant temperature in the test are the independent variables, while the pain inflicted by the cold is the dependent variable.

For the study to be reliable and valid, it is further important for the researchers to accurately define the variables in the study. This is known as operationalization. By clearly stating what the different variables and concepts are we implicitly make it possible to measure them. Other peers can also measure them and get the same results. The operationalization further ensures that we measure what we look for in the study since we understand what we are measuring and thus can understand what effect we are measuring upon the variables. In other words, this is what will affect the validity and reliability of the study. The goal of a study is to have good reliability and validity which means that peers should be able to get the same results if they conduct the same test and that the results measure what the study is testing for. This essay begins by briefly describing the use of scientific methodologies in two different papers. It then discusses the variables, operationalization of set variables, and subsequently explores the reliability and validity of the two papers. This is done one paper at a time.

The two papers in question are *"Kansas Is Flatter Than a Pancake"* by Mark Fonstda, William Pugatch, and Brandon Vogl and *"Portable devices as visual noise during lectures. User and non-user differences on distraction from internet access during lectures"* by Volden F., Leiknes H., and Røise T. Both represent different areas of research, the first looks at a geological phenomenon, while the second focus on a psychological problem. As a result, the first paper is based upon quantitative research and the second paper uses qualitative methods like questionnaires to answer the research problem.

Although the papers use different methodologies they are both reliant on the variables in their research questions. The first paper poses the question:

> *"Is Kansas flatter than a pancake?"*

while the second asks:

> *"Investigate attitudes towards computer usage and actual computer access during university type of lectures. How do computer users differ from those who chose not to access*

computers during class? How does it affect the users?"

However, neither clearly state the research question but rather introduces the problem to the reader. Comments on how it affects the users are found later on in the paper, thus the research problem above is a summary of what questions and problems are answered in the paper.

Looking at the first paper, we can identify two main independent variables: "Kansas" and "a pancake". Furthermore, we can identify one dependent variable from the paper: "flat", which is what is being measured. The researchers do not operationalize the terms or concepts, except the flatness measurement where they used a quantification from the field of geodesy, namely the flattening ratio:

$$f = \frac{(a - b)}{a}$$

where a perfectly flat surface will have a flattening ratio f of one. They did not discuss the potential issues and challenges that could be encountered nor do they elaborate on the definition of a pancake. They should have had and tested several pancakes as they differ given different cooking techniques and recipes. If one compares a classical Norwegian pancake to a typical American pancake to a french crêpe one can see huge differences. They should have spent more time on the operationalization of the pancake variable; what defines a pancake? This lack of operationalization affected the reliability and validity of the paper since peers cannot recreate the test using a pancake.

Regarding their Kansas data, they mention that off-the-shelf software could not measure the state's flatness, thus they had to program their own. However, they do not share the code with us leaving us with no way to verify their claims accurately since we do not know how the program calculated the results or what variables are used. Furthermore, we are presented with little data from their findings, only a graph, and their final calculations. This means that we cannot recalculate the final result without performing our test, which will yield different results given the nature of pancakes, the lack of the precise pancake description, and the fact that they do not list their equipment, which will mean that we can end up with different results when we use other equipment.

As a result, the conditions for internal validity are not met, given the aforementioned problems. Although the flatness measured in the paper can be transferred to other cases, thus supporting the external validity, we can question the use of different ratios used to calculate the result. They might have calculated the flatness of the objects at different scales as figure 4 shows Kansas with a scale in meters while the pancake is in millimeters. Further, the sentence from the results "The topographic transects of both Kansas and a pancake at millimeter-scale [...]" is ambiguous in that it can refer to that both are at millimeter-scale or that just the pancake is. This should have been made clearer as the difference in scale can affect the validity of the results. This should at least have been discussed in the paper.

It further looks to be a valid study, but this might be a result of bad face validity. We also struggle with content and construct validity in the choice of pancake and data chosen for Kansas. The criterion validity is also in question. To sum everything up it seems to have bad internal and external validity as we cannot draw accurate conclusions about cause-and-effect and other relationships within the data, or apply the result to situations beyond the study itself. This is partly the result of the bad operationalization of the variables and concepts.

Looking at the second paper about the visual noise created by portable devices during class, we see that the main independent variables are: users and observers. The dependent variables are the attitudes of the students and the value of the lectures. The chosen independent variables seem appropriate, as the study looked for the student's attitudes towards the use of computers during class and their opinions of how the use affected them. The researchers were better at defining the variables and concepts, compared to the aforementioned paper, and defines a user as a student that used computers during a lecture, whereas an observer is a student who took part in the class without using a portable device. They could have been more precise in the definition of a computer as different words like portable device and laptop are used for it in the paper.

To measure the attitude and how the value of the lectures was affected, the researchers used interviews/questionnaires. They asked the students about attitudes towards, and experience with, laptop use during class. They measured the number of computer users and the amount of usage of a computer during class; why they did not use a computer; what activities were performed on the devices, both by the users and the observers; and questions about non-user's reactions to others computer-use. In the method chapter, they describe and discuss how the study was conducted to avoid various biases, as well as methods to analyze the data and make the results comparable. However, the results give an indication of what each participant *felt* about the use of computers in class and how they *felt* it affects them. These qualitative measures should be supported by quantitative measures, e.g. in comparison with the students grades. Here the student's grades could have been evaluated and compared against such claims. During the time, 2007 - 2011, the researchers could have been able to discover a pattern, although they would have to be careful as grades can be affected by many other variables.

This can be a result of the lacking operationalization of the dependent variables. The *"Value of lecture"* measure does not directly measure the value, but rather the feeling the students have of the value of the lectures. Meaning that if the students still learn what they should and get good grades the value of the lectures can be said to be the same as before. Thus a clearer definition of the variable would have interpreted the result more accurately for the reader, who is now left to deduce the value of the variable for him- or herself. Further affecting the construct validity of the study.

To ensure the reliability of the study, we would need to recreate the same questionnaires with equal wording and structure of the different questions. This is important as humans are prone to "prime" their opinions with the previous question when reading the next, thus if we structure the questions in a different order we might get a different results. The study was conducted over a time-span of four years (2007-2011). Due to the further incorporation of computers in lectures and a better understanding of best use, it could be argued that the premise has changed, as the reasoning behind a student from the first year of the study and the last might differ. The sample size seems to be okay with 210 participants where 70 were females, however, the size could be susceptible to the law of small numbers, and statistically fail to take variability into account. A greater size could mitigate this.

The internal validity of the study is fairly good but suffers slightly due to the lack of operationalization of the dependent variables. The external validity seems strong as the research can be applied to other settings, like offices with open floor plans where the use of devices is visible to more than one employee. It could benefit from a larger test group, but other than that the research seems to be capable of being generalized.

The two papers clearly show the importance of using proper scientific concepts like dependent and independent variables and operationalizing to achieve good validity and reliability. We see that the second paper has far better validity, compared to the first, which is mainly given by how well they operationalized their variables and concepts. Furthermore, it is evident that the internal validity is highly affected by the definition of the concepts, while the external validity is more a result of how the study was structured and conducted. The reliability is seen to be affected by the documentation of the methodology the equipment was in the study. As a result, lacking operationalization affects the reliability of the papers, especially the first as the term pancake is left undefined. The quality of the papers and the studies are affected by the quality of these scientific concepts.