Institute for Medical Informatics, Statistics
and Documentation (IMI)

Medical University of Graz

Scientifically methodological progress in HCI&UE

A guide for seminar works, Master students and Doctoral students

Question – Hypotheses – Experiment – Evaluation – Publication

Andreas Holzinger

andreas.holzinger@meduni-graz.at

Human-Computer Interaction (HCI, in German *Mensch-Maschine Kommunikation*, *MMK*) is a highly interdisciplinary field, which encompasses findings in both Psychology and Computer Science (*Informatics*)[1]. The scientific results from HCI research are solidly incorporated in Usability Engineering (UE). Whilst classic HCI is a bridge between Natural Science and Computer Science, Usability Engineering – as opposed to pure Usability Testing – is a hard Engineering discipline, which applies strictly analytical-experimental methods and is embedded in Software Engineering. As long as something practical is continued to be implemented, the connection between Science and Industry remains essential, since Industry deploys the products. This is a possibility of ensuring real benefits for the end-users. However, it is also important that the experiences from Industry are integrated into Science. In order to document the results and to transfer the findings to the Community, and to enable, initiate and stimulate further research, the documentation of research work is important. This short guide will help future students to organize their experimental work, achieve a methodical approach and advise on successful publications.

[1] The broader terminus is Computer Science, whereas Informatics is used mostly in compounds; for example Medical informatics, which emphasizes the concentration to the management and processing of **data, information** and **knowledge** in the named discipline, i.e. in Medicine and Health Care.

1. Everything begins with curiosity; a Question or Problem statement

The question first …

Every empirical research activity starts with a definite question. However, the question should be relevant and interesting and should somehow be new. Consequently, at the very beginning it is essential to search relevant literature in that direction.

In the best case, we can find contributions (called "papers") which are exactly in the area, which are in the focus of interest. We can look at done research and look for possibilities of altering parameters in order to find something NEW.

Back to the question: It is important that the correspondence between at least two variables is already expressed in the question. For example: "How does variable x influence variable y?" or "What connection exists between x and y?" The basic idea is to introduce one changeable or alterable (dependent) variable (DV) and one unchanging (independent) variable (IV). The most interesting part is: "How does y alter, when I alter x?"

Naturally, this can also be very complicated, e.g.: "in which relationship do the variables X, Y and Z stand to the variables U and W? A typical question would be e.g.: "a higher success in learning is achieved by the employment of the simulation X than the traditional method Y". Whereby, it is necessary to define *a higher success* in detail.

2. Building a working hypothesis or: Formation of a working hypothesis

A working hypothesis develops from the formulation of a question, e.g. in our case:
"*The learning success is greater with the use of the simulation X than with the traditional method Y*". Again, with an exact definition of the terms used. According to

the circumstances, the working hypothesis must be continually modified, until it actually shows what one wants to examine, e.g.: "*by the employment of the simulation X, learning time for the topic Y can be shortened*".

Here is it necessary to very carefully clarify: Which variables are to be examined and which expectation exist (again with reference to the theory and/or to preceding work).

Scientific hypotheses (Popper, 2005) can be brought formally into the following form:

When X then Y

X implies Y

$X \rightarrow Y$

It is important to understand that hypotheses are only conditional statements: The effect occurs, not under all possible circumstances, but only under completely pre-determined circumstances.

The hypothesis must be evaluated experimentally. Whereby there are exactly three possibilities:

1) The hypothesis is significantly proved (verified = is true)
2) The hypothesis is significantly disproved (falsified = is wrong)
3) The hypothesis can (on the basis of the existing data) neither be verified nor falsified (no statement can be made)

Note: Whether a hypothesis is true (thus ALWAYS applies), can not be confirmed due to the universality of the statement. Scientific hypotheses can therefore only be falsified.

Nevertheless hypotheses can be differentiated with respect to the degree of their confirmation, whereby certain criteria is taken into account, such as how frequently and in which critical places the hypothesis has already been confirmed (Popper, 2005).

3. Definition of an experimental Design

Experimental designs must always be formalized, in order to complete investigations according to uniform criteria (Sarris, 1992).

Let us define:
EGr … Experimental Group
CGr … Control Group
TSn … Test Subject n
IV … Independent Variable
DV … Dependent Variable

The most important components of an experimental design are:

a) The experimental groups taking part in the investigation (EGr, experimental group and CGr, control group);
b) The allocation or assignment of individual test subjects (TSn) to the EGr;
c) The experimental conditions (the steps of the IV) as well as the measurement of the DV
d) The investigation sequence, thus the realization times of the IV as well as the measurement times of the DV

Typical example: Pre-test post-test experimental control group design (Manstead and Semin, 1992):

Two given (according to randomized assigned groups) are submitted to a pre-test. The experimental group (EGr) receives the experimental condition (Treatment), the comparison group (CGr) serves as a control. At the end, both groups receive a post test.

This can be graphically represented as follows:

N O X O

N O O

whereby

R = Random assignment (randomized)

N = Non random assignment; arbitrarily given

O = Measurement of the IV, in the case of several Os in a line, always the same IV

Thus, for example, O1, O2... Measurement of IV1, IV2...

X = Treatment (no symbol is control group)

X1, X2, X3... several single factorial Treatments

X11, X12, X21, X22... several Treatments two factorial

4. Operationalization

Operationalization is, in principle, the transfer from the theoretical construction to the observable measurement (Manstead and Semin, 1992), e.g. from theoretical terms (actions), empirical (mathematical) variables must be produced.

This is one of the most difficult aspects of scientific work; on which the result of the experiments is dependent (contruct validity, see for example (Bortz and Döring, 1995)).

The dependent variable (DV) is the variable, whose change due to the IV is measured (DV must measure the effect of the IV).

The DV is described as dependent, because its or their developments are a function of the IV and therefore dependent on the IV. The DV should alter as a function of the IV. The DV is also referred to as reaction variable (Sarris, 1992), since these practically represent the reactions of the VPn (e.g. the answers in the questionnaires).

In addition, there are confounding variables (CV), which exercise unwanted influences on the DV (Bortz and Doering, 1995), (Christensen, 2001). The CVs disturb, because they also affect the DV and which results in ambiguity, causing variation in the DV on IV or the CV. (in German: Konfundierung der Ursachen). Disturbing factors must be eliminated or at least controlled!

5. Overview of Experimental Plans

It is differentiated between pre-experimental plans and "genuine" trial plans. Pre-experimental trial plans are very controversial, because very inaccurate and should also be rarely used:

1. Single investigation of a group X O
2. Pre-test /Post-test single group plan O X O
3. Statistical group comparison
   without previous measurement/testing N X O
   N O

"Real" experimental trial plans (true experiments) are better.
The most important characteristic of a genuine experimental trial plan is the random allocation of the test subjects to the IV levels (Christensen, 2001).

1. Randomized control group allocation without pre-test

R X O
R O

2. Randomized control group with pre-test

R O X O
R O O

By the use of random allocation of the test subjects to the groups (X = experimental group, blank space = control group), almost all interference factors are controlled, with the exception of socially based sources of error. This is the **power of the randomization principle**.

Of course, the random allocation of the TSn on the experimental groups cannot eliminate disruptive factors, but can keep them constant, with which the potential confounding variables work similarly in both groups, however the difference between the groups can only be backtracked to the IV.

There is still another third possibility, i.e. quasi-experimental trial plans:

1. Pre-test/Post-test Control Group Design with non-homogeneous Control Group.

   N O X O

   N O   O

2. Simple time sequence trial plan

   O O O O X O O O O

3. Multiple time sequence trial plan

   N O O O O X O O O O

   N O O O O O O O O O

4. Arrangement with homogeneous time samples

   X_1O X_0O X_1O X_0O

   (X_1 =Experiment Condition X_0 = Control Condition)

1. Pre-test/Post-test Control Group Design with non-homogeneous Control Group.

   N O X O

   N O   O

This trial plan ranks among the most well-known arrangements for the field of education because it is often simply not possible to take the VPs out of their environment and allocate them randomly to new groups.

Time sequence trial plans (Time-series designs)

With time sequence trial plans, the IV is measured several times before and after the implementation of the Treatments.

Simple time sequence trial plan

O O O O X O O O O

In contrast to the pre-experimental pre-test/post-test single group plan (O X O), the simple time sequence trial plan can control most disruptive factors to a large extent.

Multiple time sequence trial plan

N O O O O X O O O O
N O O O O O O O O O

The multiple time sequence trial plan controls the interim occurrence by the inclusion of a control group and can be regarded, on the whole, as an extension and/or an improvement of the pre-test/post-test control group design with a non-homogeneous control group.

In HCI research in particular, there are also other extremely interesting methods, which can be used successfully, especially in Usability Engineering, see in addition (Holzinger, 2005), (Holzinger, 2004), or have a look at:
http://user.meduni-graz.at/andreas.holzinger/holzinger/usability.html

It is enthralling to combine classical research methods from psychology with the new methods from Usability Engineering and definitely provides much room for further work.

6. Data Evaluation

Much effort must be spend on the careful evaluation, the assessing of differences etc. This is done with the help inferential statistics procedures, see in addition e.g. (Christensen, 2001), (Bortz and Doering, 1995). Very helpful are computer programs including SPSS, SAS etc. and essential is the use of MS EXCEL. However, please keep always in mind, that the thinking and understanding of the problem is the main issue in your work … it is not the tools or the handling of impressive data. Therefore, for our work basic knowledge in Statistics is essential!

7. Conclusions

Finally, based on your results you should be able to answer your questions, to evaluate your hypotheses and to draw conclusions, however, always based on the underlying theory

8. Publication of the Investigation

"*Science is every intersubjective verifiable investigation of facts and the systematic description and explanation of the examined facts which it is based on.*" (Speck, 1980)

A field of knowledge is also often called a Speciality. However, in the case of Human-Computer Interaction we are dealing with a very special combination: On the one hand, computer science is technically constructive and not, per se, a strictly delimited, self-contained subject (Mittermeir, 2003); on the other hand, Psychology, as the science of human experience and behaviour is equally a trans-sectorial science, which can be methodically assigned to the natural sciences but which contains very much in common with the humanities and the social sciences.

The Community (target group) in HCI research is therefore interdisciplinary constructed, whereby, it must be stressed that the subject HCI is assigned to computer science because the practical implementation (must) actually take place in computer science (IT/Informatik)!

Publications (lat. publicus = publicly) are now the medium with which the knowledge and insights attained by the Community are made accessible. Since HCI & UE developed from (Anglo American) computer science, English has a central significance – as today in nearly every discipline of science; English has prevailed as the language of science – much as Latin in the Middle Ages.

However, local languages are significance for what, in English, is referred to as "awareness", i.e. the sensitization, the creation of consciousness for a certain subject.

We differentiate in principle between two kinds of publication:

1) Original work in Proceedings (Conference contributions)
2) Original work in Journals

These are the classical Papers, which document original, peer reviewed, research work (original work). An important criterion is *where* the publication appears. The basic rule is: everything which is listed in the Science Citation index (SCI) has a good reputation; however, that does NOT mean that publications which are not listed are not any good, however, it also does not mean, that all publications which are listed are all good.

http://www.sciencegateway.org/impact/if02c.html

The SCI is a data base, provided by the Institute for Scientific Information (ISI). In it, not only are the titles and authors of contributions listed, but also, which other articles and work are recorded in their bibliography. With this assistance the "renowned" Impact Factor (IF) and other bibliometrical data is determined. The IF of a magazine is a yardstick for how often (statistically), an article from this magazine was referenced in other magazines. In principle, the higher the IF, the more outstanding the magazine, however in the "short-lived" and rapidly developing computer sciences, the Impact Factor does not carry the large value which is does in, for example, medicine.

Also, the computer sciences publications do not, in principle, obtain such high Impact Factors. A substantial criterion is, however, that the publication is listed in the SCI. This is the case with the most important publications of the ACM and/or IEEE (the "Association for Computing Machinery" was created 1947 as the first scientific society for computer science; The Institute of Electrical and Electronics Engineers is the largest technical professional association in the world, with more than 360.000 members in 175 countries (2005)), as well as, for example, Springer Lecture Notes in computer Science (LNCS).

The Name of the Aim is Journal Publication

Although many Proceedings (e.g. LNCS) nearly have the character of a journal, a solid journal still remains the Mercedes among publication goals. In order to achieve this goal, as efficiently as possible, a three step method can be used, whereby the following applies:

First Patent then Publish

A patent is a *right of property* and gives the owners the right to forbid others to use their invention without permission, whereby this applies only to *commercial application*.

Inventors do not have the right to prevent the private use of their invention. The most important criterion for a patent is its innovation:

It is new, if it does not already belong to the state of the art (§ 3 PatG; Kind. 54 EPUe).

The state of the art includes everything that was publicly accessible before the day of registration, (Publications ...). The most difficult aspect is that there are substantial differences between the individual countries, in particular between Europe and the USA. However: If one believes one has found a really good, *economically usable idea*, then one should first look, whether there is already a patent on it,
see:

http://ep.espacenet.com/search97cgi/s97_cgi.exe?Action=FormGen&Template=ep/en/advanced.hts

9. Literature Research

For scientific work, primarily original work is important: again, articles in journals and Proceedings. Compared to these, text books – as a source of reference – have little meaning. The quality of a work rapidly becomes evident when the references are studied. The following data bases represent the most important research sources for our work:

1) *ACM Digital Library* (Association of Computing Machinery)
http://portal.acm.org/dl.cfm

2) *IEEE Digital Library* (Institute of Electrical and Electronics Engineers, usually pronounced I-triple-E)
http://search2.computer.org/advanced/simplesearch.jsp

3) *Science Direct* contains the electronic versions of articles from publishers which include: Academic Press, Baillière Tindall, Cell Press, Churchill Livingstone, Elsevier, Elsevier Advanced Technology, Elsevier Current Trends, Excerpta Medica, Gauthier-Villars, JAI, Mosby, North-Holland, Pergamon, W.B. Saunders u.a.
http://www.sciencedirect.com

4) *Scientific Literature Digital Library (CiteSeer)* Database of publications in Computer Science Informatik and Information Sciences, complete texts are usually linked and cost-free available.
http://citeseer.ist.psu.edu

5) Satisfactory as a scientific *search machine* is:
http://www.scirus.com

The most frequently asked question from students and graduate students (doctoral candidates) is: " *what is enough research resp. when do I have researched enough*".

It cannot be said often enough that actually the "reading-in phase" should be completed at the time of starting any work (technical reading takes place during the entire study). If the actual study lies somewhat in the past, a short – reading-in phase can be granted. But the basic rule applies: Begin immediately with the work! The first lines of the problem definition probably results in a substantiated literature research anyway.

Thus: "*what is enough research resp. when do I have researched enough*".

A basic rule is: For each (serious) statement, one should obtain three references: the oldest obtainable, a well-known and, if possible, the latest. Naturally, this will not always be so easily possible, however the student should go to the trouble of research when and who published the first contribution to the current problem, according to the slogan: "back to the roots. It is absolutely essential that one knows "one's" speciality (this is the main focus of the work), i.e. one must also know the people,

1. who worked on this subject in the past (the historically "Outanding Personalities") and
2. those currently working on the subject (the "Competitors").

Basic rule: Reading is important, even very important... one must read much, very much, in order to comprehend the current trends, in order to remain up-to-date; but, writing is at least as important. I have known cases, where students have "finished" with their work ... however not yet written one line. Here it is the task of those responsible to suggest, to instruct, to guide and to move forward and encourage!

Three Steps to the Goal

If one excludes a patent registration, nothing stands in the way of a publication. This can follow in three logical steps:
1) *Technical report*: an informal publication of the findings in the form of a Technical Report; whereby this TR does not necessarily have to be made publicly accessible, but should present an informal basis of the work;
2) *Proceedings*: Publication of the findings at a conference; and
3) Journal paper: based on 1 and 2, this should be a solid documentation of the findings in a journal.

Although, in principle, composing papers takes place in teams, someone must nonetheless take the "Lead" and push the publication forward.

10. Structure of Publications

The structure almost always (there are exceptions) follows according to this pattern, no matter how the "Template" may look:

Title

Authors

Affiliation (affiliation to an institution)

Address

E-mail

Abstract (goals, methods, experimental Setting, result/findings, conclusion)

Keywords

Introduction

Theory, Background

Methods

Experimental Setting

Results/findings

Discussion of the results/findings

Conclusion

Thanks

About the authors

Literature

Or concise:

Introduction and Motivation for Research

Methods and Experimental Setting

Results and Discussion

Conclusion

Acknowledgement: I thank Gig Searle for helpful discussions on this topic.

11. References

Bortz, J. & Döring, N. (1995) *Forschungmethoden und Evaluation für Sozialwissenschaftler.* Berlin, Heidelberg Springer.

Christensen, L. B. (2001) *Experimental methodology.* Boston, London, Toronto et.al., Allyn and Bacon.

Holzinger, A. (2004) Application of Rapid Prototyping to the User Interface Development for a Virtual Medical Campus. *IEEE Software,* 21, 1, 92-99.

Holzinger, A. (2005) Usability Engineering for Software Developers. *Communications of the ACM,* 48, 1, 71-74.

Manstead, A. S. R. & Semin, G. R. (1992) Methoden der Sozialpsychologie: Von der Vorstellung zur Handlung. In: Stroebe, W., Hewstone, M., Codol, J.-P. & Stephenson, G. M. (Eds.) *Sozialpsychologie. Eine Einführung*. Heidelberg, Berlin, New York, Springer.

Mittermeir, R. T. (2003) Schulinformatik - ein Fach oder ein Gebiet? In: Anton Reiter, G. S., Heinz Strohmer, Lydia Tittler, Martin Weissenböck (Ed.) *Schulinformatik in Österreich*. Wien, Ueberreuter.

Popper, K. (2005) *Logik der Forschung.* Tübingen, Mohr.

Sarris, V. (1992) *Methodologische Grundlagen der Experimentalpsychologie. Band 2: Versuchsplanung und Stadien des psychologischen Experiments. : .* München, Reinhardt.

Speck, J. (Ed.) (1980) *Handbuch wissenschaftstheoretischer Begriffe, Bd. 3. 1980,* Göttingen, Vandenhoeck & Ruprecht.

Note:
Dear Students, report any suggestions directly to
andreas.holzinger@meduni-graz.at