CAT – Development of a German Computer Proficiency Test

Master Thesis, July 5th, 2006

B. Sc. Martin Rohner

Institut für Psychologie

Universität Basel

Supervisor: Prof. Dr. Klaus Opwis

Co-Supervisor: lic. phil. Javier Bargas-Avila

Contact:

Martin Rohner

Socinstrasse 11

CH-4051 Basel

martin.rohner@unibas.ch

Table of Contents

1 Abstract

The aim of the present study is the development and evaluation of an instrument to measure computer proficiency. The resulting instrument, the Computer User Proficiency Test (in German: "Computer-Anwender-Wissens-Test", thus CAT) is designed especially for the German speaking population. In contrast to other computer proficiency tests, CAT measures computer knowledge and the attitude towards computers. CAT is quick to complete and easily transformable to an online version. To validate CAT a sample of 1,119 subjects were recruited. In this investigation, CAT demonstrated moderate to high internal consistency. Computer usage and Internet usage correlated moderately with CAT. Furthermore, computer anxiety showed an impact on computer knowledge. The results of this study indicate that CAT appears to be an adequate tool to assess computer proficiency. Scientists like psychologists, usability experts, assessment centers, and others can use it.

Keywords: computer proficiency test, computer knowledge, computer skills

2 Introduction

Nowadays, computers and the Internet are almost everywhere and accessible for almost everyone in Western Europe. For instance, 70% of all Swiss people older than 14 years used the Internet at least once during the last 6 months, and 57 % use it several times per week (Swiss Federal Statistical Office, 2006). In June 2005, 20% of all Swiss households were equipped with a highspeed Internet access such as ADSL (Swiss Federal Statistical Office, 2006). Hence, the importance of both computers and the Internet for our society are beyond. During the last thirty years a number of computer proficiency tests have been developed (i.e. Bradlow, Hoch & Hutchinson, 2002). A brief look at the closer history concerning computer proficiency tests shows that during the late seventies and during the eighties researchers were mostly interested in finding predictors of programming capabilities (Evans & Simkin, 1989). This was necessary since computers were mostly used in specialized areas like programming and engineering (Evans et al., 1989). People engaged in these areas were expensive and hardly available. In order to increase their availability, there was a great interest in identifying individuals who could be trained efficiently. In the late eighties and nineties, computers became important in several other areas like service economy, non-technical science, and the office in general (Bradlow et al., 2002). As a consequence, new proficiency tests started to focus on other capabilities like office application knowledge. In the late nineties, the Internet became more and more important and highly accessible (see figure 1). Due to this fact, the Internet became a new factor in economics. The focus of interest moved to Internet knowledge as well, and new questionnaires arose like the Internet Skills Proficiency Test (O'Hanlon, 1999) or the computer proficiency test developed by Bradlow et al. (2002).

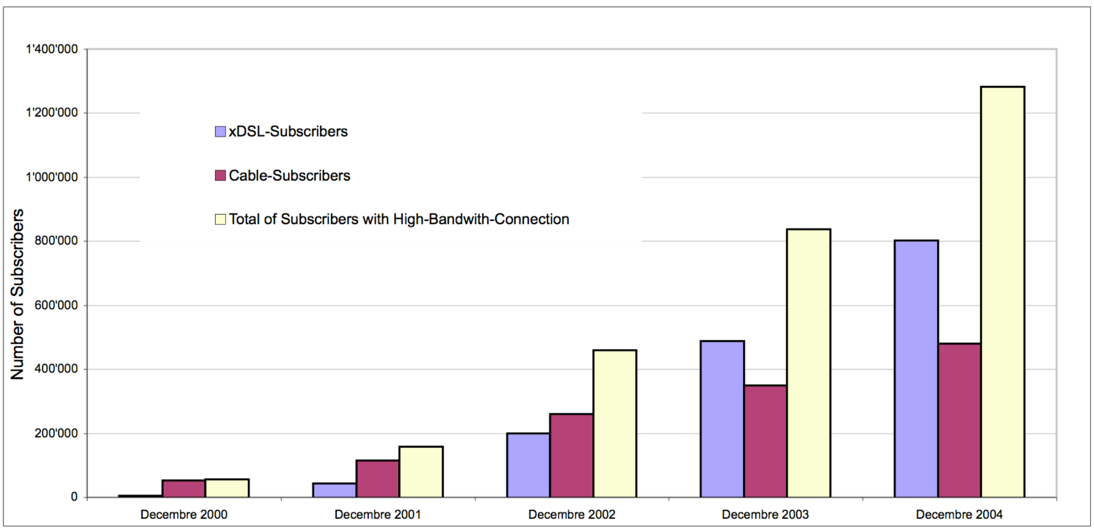

Figure 1. Development of the number of subscribers with high bandwith Internet connections in Switzerland (Swiss Federal Statistical Office, 2006)

In the following section we introduce Bradlow's computer proficiency test and show reasons why we decided to create a new one called CAT (Computer-Anwender-Wissens-Test, translated: computer user proficiency test).

3 Summary of Bradlow's Computer Proficiency Test

In 2002, Bradlow et al. published a computer proficiency test. This test is based on nine subdomains, namely (a) terminology, (b) file management, (c) word processing, (d) spreadsheets, (e) data base, (f) printing, (g) e-mail, (h) the Internet, and (i) information search. Each subdomain contains 3 items. Bradlow et al. (2002) chose three different types of items: multiple choice with one correct answer (22), True-False (4), and "check all that apply" (1). Each item is scored as correct-incorrect.

To assess the statistical properties of their test they used a three-parameter item response theory model (Birnbaum, 1968). They found appropriate values, especially for the item difficulties. This computer proficiency test is often used as a covariable in psychological studies (see e.g. Bargas-Avila, Oberholzer, Schmutz, de Vito & Opwis, 2006; or Schmutz, 2004). However, there are some problems concerning this test. To begin, the subdomain “printing” has problematic answer items. For instance, item 17 states, "Which of the following types of printers has the highest print quality (i.e. resolution)?”. The provided answers (dot-matrix; ink jet; laser jet; bubble jet) are all reasonable, depending on the application (graphics vs. letter vs. carbon copies) and the printer model. Furthermore, Bradlow et al. (2002) did not consider the attitude towards computer issues such as computer anxiety. Van Braak (2004), for instance, could show that "computer-confidence" has the most significant influence on self-perceived “computer-competence”. Lastly, this computer proficiency test is constructed for the English speaking population, hence, even with a translated version, a new validation is needed.

4 Method

4.1 Test Construction

Our first goal is to provide psychologists with an instrument which allows them to get a rough idea of how well a subject can handle a computer. This instrument is meant to be a screening tool to use as , e.g., a covariable in scientific experiments. Therefore, the instrument should be quick to complete (and quick to evaluate) as testing time is often expensive. Since hardly any validated instruments in German are available, the second goal is to provide a validated German computer proficiency test. The test should address scientists like psychologists, usability experts, assessment centers, and others. Lastly, the instrument should be easily transformable into an online form to widen its area of application such as online experiments.

To meet these criteria we decided to go for five knowledge and one attitude domain. These are (a) computer, (b) database, (c) Internet, (d) office, (e) search, and (f) attitude towards computers and their environment. The first five categories were also used by Bradlow et al. (2002). With the attitude toward computers category (f) we take van Braak's (2004) findings into account that "computer-confidence" has an effect on self-perceived "computer-competence". Each category contains three items.

4.1.1 The Six Categories

In this section we describe the six categories of CAT.

a) Computer: This category deals with hardware knowledge, operating system issues, and basic system settings.

b) Database: In this category, we address database related knowledge like common known applications and structural concepts.

c) Internet: This category deals with typical Internet related issues, in this case with common terms, e-mail, and coding.

d) Office: This category is connected to Microsoft Office applications such as Word, Excel, and Powerpoint. The items deal with general functions as well as facts provided by the Microsoft Office environment.

e) Search: Here we focus on search strategies as well as common known search engines available on the Internet.

f) Attitude-Toward-Computers: Items in this category rely on the Computer Anxiety Rating Scale by Heinssen, Glass and Knight (1987). Here we use statements with which subjects are asked to indicate their level of agreement on a Likert scale (1 means "Does not apply at all", 5 means "Does fully apply"). Topics are: personal capability, trust towards computer environments (fear of experiencing personal harm like losing money), and fear about assumed computer system capabilities (anticipated loss of control). Subsequently, the attitude towards computer category is called CARS. Table 1 shows the items and their categories.

4.2 Test Development

4.2.1 First Version

We used a three step process to develop this test. In the first version of CAT, each item had five possible answers with an n/a-option except for the CARS items where a five-point Likert scale was used.

Subjects were asked to judge every answer as correct or incorrect. If subjects judge the answer to be correct, they were asked to mark the corresponding checkbox next to the answer, otherwise to leave the checkbox empty. A subject received a point (or a part of it) according to the following rule: Add up all correct answers within an item that were marked as correct by the subject. Further add up all incorrect answers within the same item which were left unmarked. Divide this sum by five (number of answers per item). Thus a subject could receive 0, 1/5, 2/5, 3/5, 4/5, or 1 point respectively for each item (except CARS). Where the n/a-option was used, the subject received zero points for the corresponding item. In the end, all points for every item were summed up to an overall score. Each value of the CARS-item was summed up as well to an overall score (SUMCARS), however, one CARS item needed to be inverted first before the sum could be processed.

The first version was distributed in a psychology class at the University of Basel and 242 completed questionnaires were returned within ten minutes. We found similar values as we used the online form of this first version (11.5 minutes). Participants had to create a personal code in each survey we conducted. Since the construction of the code was rather complex, it was the most time consuming issue for the participants as they completed the test.

Due to several problems, the first version of CAT had to be discarded:

(1) Cronbach's alpha: As we conducted the reliability analysis, the first version showed a poor reliability value of $\alpha = .343$ ($n = 242$). We believe that the n/a-option was the main reason for this result. There was a problem with participants who guessed. Their answers were mixed up with those participants who indicated not to know. Since both received zero points for the same item, no clear discrimination between those two types of participants could be measured. To solve such problems, lie-detecting items are needed. However, lie-detecting scales are controversial (Buse, 1976).

(2) Due to the item scoring procedure there were some items with which it was easier to score points (or parts of a point). This was the case, when more than one of the item answer possibilities (up to all of them) were correct statements. Here it was easier to score “lucky points” as a participant was able to guess. When participants indicated not to know, they were “punished” in the sense that they were not rewarded with points where those who guessed were. This might have affected the poor Cronbach’s alpha value, too. Since the item scoring procedure and the overall score building was complicated, the evaluation of the questionnaires was time consuming. Furthermore, the procedure was vulnerable to miscounting. Therefore, the first version of CAT did not achieve the criterion of quick feasibility from the investigator’s point of view.

(3) A five-point Likert scale as used for the CARS items can cause some problems of interpretation. We didn’t offer an n/a-option for these items as there was no defined correct answer. Yet there was the center point interpretation problem, that is “fifty-fifty” vs. not answerable vs. indifference (Bortz & Doering, 2002).

To solve these problems, we decided to reconstruct CAT with regard to the item answer structure. Therefore, a number of the items needed to be modified.

4.2.2 Second Version

This time a multiple choice structure with a forced choice paradigm was chosen. There was one correct answer per item and no n/a-option offered. For the CARS items a six-point Likert scale is used. The advantages are as follows: There is no counting issue with the n/a-option as subjects are forced to choose an answer. There is no difference between subjects who guessed on the one hand, and subjects who did not know on the other hand. Both of them have a chance of .2 to mark the correct answer. Due to this fact, no lie-detecting items are needed as both are treated equally. The one-correct-answer design allows simple counting (quick feasibility). Using a six-point Likert scale for the CARS items avoids problems of interpretation with the center point interpretation.

We are aware that with this procedure we do not completely cover a subject’s computer related knowledge or opinions, as they cannot state missing knowledge anymore (Bortz & Doering, 2002). However, gains in counting, overall score interpretation, the simple transition into an online form, and fast feasibility are strong arguments for applying these changes.

We arranged the items in a way that the further you advanced the more difficult the items would get within a category. Additionally, we mixed the items of the different categories within CAT, so that there is no specific category pattern.

4.2.3 Scoring

The scoring of the new CAT is simple. For each correct answer subjects receive one point. All the points are then summed up to an overall score (CATpoints). This procedure applies to all items except the CARS items. Since these items use a Likert scale, the procedure is different. First, invert the value of item no. 2 (CARS1) according to this rule: inverted item value equals 7 minus raw item value, so that 6 becomes 1, 5 becomes 2, and so forth. Then sum up the values of item no. 6 (CARS2), no. 14 (CARS3), and the inverted value of item no. 2 to an overall-score (SUMCARS). Thereafter, CATpoints can have a value from zero to 15, and SUMCARS from three to 18 respectively.

Table 1

Items with their Category and Substitution. Original german version is in the appendix

Item-position	Item [Answers, "/"separated]	Category/ Substitution
1	Gigaherz is usually used for… [the hard disk / the system memory (RAM) / the display / the processor (CPU) / the power supply]	Computer1
11	Documents which are in the trash... [can be moved back and used again anytime / are irrevocably deleted / can only be restored with specialized software (e.g. "Norton Utilities") / can only be deleted / are usually affected by viruses or worms and have to be deleted]	Computer2
17	Which of the following number-blocks are possible and correct IP-addresses? [212.35.35.35 / 192.168.0.212.94 / 164.123.385.63 / 354.286.212.2/ All above mentioned IP-addresses are correct]	Computer3

Table 1 (*continued*)
Items with their Category and Substitution. Original german version is in the appendix

Item-position	Item [Answers, “/“separated]	Category/ Substitution
4	Which of the following products is a database-software? [Microsoft Word / Microsoft Windows-Media-Player / Microsoft MSN / Microsoft Access / Microsoft Internet-Explorer]	Database1
10	A relational database is... [a set of data which are similar to each other / datasets that are delicate due to protection of privacy / a collection of data-files (e.g., Word or Excel documents) / datasets which are connected over one or several key-variables / an address-index of related persons]	Database2
16	A database can store the following types of data: [Pictures / word documents / film documents / list of addresses / all above mentioned]	Database3
7	The abbreviation "URL" stands for: [United Reference List / Universal Receiving License / Uniform Resource Locator / Universal Reallocate List / Unified Recording Language]	Internet1
15	If you use the BCC function in an e-mail program, then.... [everybody knows, that a copy was mailed to the BCC-recipient(s) / only the main-recipient(s) knows that copies of this e-mail were sent to the BCC-recipient(s) / the sender automatically received a copy of this e-mail / no recipient (not even the main-recipient(s)) knows that a copy is mailed to the BCC-recipient(s) / None of these statements are correct.]	Internet2
18	The following HTML-code is given: Television What will happen? [www.sfdrs.ch will be shown / "Television" will be displayed / when www.sfdrs.ch is clicked, the website of Television will be loaded / Television is a paragraph of the third order / None of the statements are correct.]	Internet3

Table 1 (*continued*)
Items with their Category and Substitution. Original german version is in the appendix

Item-position	Item [Answers, "/"separated]	Category/ Substitution
5	Which file-extension does Microsoft Word usually use for Microsoft Word documents? [.txt / .doc / .xls / .dat / .wrd]	Office1
9	If you copy a paragraph within Microsoft Word into the clipboard (ctrl-c), then.... [It will be available in Microsoft Word / you can use it within Microsoft Word, Excel and PowerPoint / it will be available system-wide, e.g. within an address-bar of an Internet-browser / it will be available until you copy something new into the clipboard or shut down the computer / All of the above statements are correct.]	Office2
13	When "=sum(c3:c7)" is written in an Excel-cell, this means that in this cell... [the sum from c3 and c7 will be calculated / first the division from c3 and c7 will be calculated and afterwards the sum of them / the sum of c3, c4, c5, c6 and c7 will be calculated / the values of c3 and c7 will be formatted as a sum / This expression is wrong. Letters cannot be divided.]	Office3
2	Your are searching for the exact phrase: The rabbit is dead With the help of a search engine you want to find all documents with this exact content. You will write in the search-field: [The + rabbit + is + dead / the rabbit is dead / "the rabbit is dead" / rabbit + dead / findphrase "therabbitisdead"]	Search1
8	Which of the following terms is not a search-engine on the Internet? [Webcrawler / Yahoo / Explorer / Alta Vista / Google]	Search2

Table 1 (*continued*)
Items with their Category and Substitution. Original german version is in the appendix

Item-position	Item [Answers, "/"separated]	Category/ Substitution
12	If you enter the exact following words and string in Google (www.google.de): "the green tom thumb" site:.de then.... [everything which contains the proper words and string: "the green tom thumb" site:.de will be searched / only .de-domains will be searched (e.g., www.netzeitung.de) / everything will be searched for which contains at least one of the words from the string, so "the" or "green" or "tom" or "thumb" or "site:.de" / it is the same as if you would enter findphrase "thegreentomthumb" in the search-field / All of the above statements are correct.]	Search3
3	How much do you agree with this statement? "I am convinced that I am able to learn a programming language." [Six-point Likert-Scale]	CARS1
6	How much do you agree with this statement? "I prefer to handle bank-transactions at the counter (and not via Internet)." [Six-point Likert-Scale]	CARS2
14	How much do you agree with this statement? "The thought that computers work more precisely than humans makes me queasy." [Six-point Likert-Scale]	CARS3

4.3 Participants

To validate CAT, we recruited participants using three different ways. First, we handed out CAT to two different schools in different places of Switzerland. One school is a comprehensive school in Sissach containing all levels from secondary school to grammar school. The school is a grammar school in the canton Schwyz. Further, we distributed CAT to psychology students as well as others (no special group characteristics). Lastly, we recruited participants using an online version of CAT. This version was announced via the website of the Faculty of Psychology (University of Basel), newsletters, and via word-of-mouth recommendation. Altogether, we recruited 1,119 participants (see table 2). None of the participants was rewarded except the online participants. One of them received an iPod Nano.

We used different numbers of participants to conduct the diverse analyses. Mainly this was the case because some participants did not fill out the questionnaire completely or provided multiple answers where only a single answer was appropriate. Furthermore, we excluded subjects who indicated no data (e.g. no gender specification) or unqualified data (e.g. spending 168hours per week at a computer). Beyond that, all participants who were younger than the age of 18 were excluded for analyses with CARS item. People younger than 18 are assumed not to be able to answer the e-banking item since they don't have the possibility (by law) to get familiar with this topic yet.

Beside the test results we collected other data as well. Each participant was asked for age, gender, native language, hours spent at the computer per week (in average), and hours spent on the Internet per week (in average).

The online population was further asked to rate each item separately with respect to the self-perceived difficulty after it was processed. A 6-point Likert scale was used ("1" means "This was a difficult item", "6" means "This was an easy item"). Lastly, we recorded the time the online participants needed to complete the online questionnaire.

Table 2

Participants with their characteristics

Origin	Age (*SD*)	Computer Usage in hours per week (*SD*)	Internet Usage in hours per week (*SD*)	N
Sissach[1]	13.53 (3.02)	7.55 (7.95)	5.41 (7.51)	626
Gym. Schwyz[2]	15.48 (3.94)	7.99 (15.42)	10.44 (15.76)	146
Miscellaneous[3]	20.35 (12.17)	14.66 (15.28)	5.15 (6.13)	160
Online	28.45 (12.05)	24.90 (20.10)	13.21 (14.07)	187

Note. [1]*:"Gesamtschule Sissach", the comprehensive-school of Sissach;* [2]*: "Gym. Schwyz" is a high-school in the canton Schwyz;* [3]*: Participants in these groups do not have specific properties;*

5 Assumptions

In this section, we provide assumptions with their corresponding hypotheses regarding the applicability of CAT as a computer proficiency test.

First, we assume that male participants perform better compared to female participants in both the online version as well as in the paper-pencil version. It is a common fact that gender effects are found in technical domains (e.g., Beckwith, Burnett, Wiedenbeck, Cook, Sorte & Hastings; 2005).

Hypothesis 1: Male subjects perform better than female subjects.

Secondly, we expect that participants with higher values in computer usage and Internet usage perform better on CAT. We expect a higher correlation between computer usage and CAT compared to the correlation between Internet usage and CAT. Computer usage covers several domains and CAT tests a wide area of computer knowledge.

Hypothesis 2: Subjects with a higher value in computer usage perform better than participants with a lower value.

Hypothesis 3: Subjects with a higher value in Internet usage perform better than participants with a lower value.

Thirdly, we expect the online participants to perform better than the paper-pencil participants. On average, participants who entered the survey online are more likely to have acquired computer knowledge as they probably use computers more than participants who completed the paper-pencil survey (computer usage in hours; online, $M = 29.64$, $SD = 19.86$; offline, $M = 17.55$, $SD = 23.25$; $d = 1.41$)

Hypothesis 4: Participants who complete the online version of CAT perform better than participants who complete the paper-pencil version.

Lastly, we predict that subjects who score high in the CARS category (meaning high computer anxiety) perform worse on CAT compared to others.

Hypothesis 5: The higher participants score on the CARS scale the worse they perform on CAT.

6 Results

In this part we provide classical test theory results followed by the different results of the hypotheses. All analyses were computed with SPSS 11.0.4 for Macintosh.

6.1 Cronbach's alpha

To calculate Cronbach's alpha, we used only the categories that build CATpoint (15 items), the CARS category was not included. We did this, because CAT and CARS measure two different constructs, computer knowledge and computer anxiety. A separate reliability analysis for the CARS item was not conducted as there are only three items.

We found an environment effect as we computed Cronbach's alpha for the different versions of CAT (online vs. paper-pencil). Therefore, we provide three different values, one for the online version, one for the paper-pencil version, and one for the mixed dataset (both online and paper-pencil). Table 3 shows the corrected item-total-correlations and "alpha-if-item-deleted" for the mixed dataset. To conduct this analysis we used all participants who completed the questionnaire and provided useful data (such as gender and age). Cronbach's alpha for the online version of CAT is $\alpha = .88$ (n = 172), for the paper-pencil version $\alpha = .67$ (n = 646), and for the mixed version $\alpha = .75$ (n = 818). We ascribe this effect to the poor computer knowledge we found in the paper-pencil survey (CATpoints, $M = 6.07$, $SD = 2.64$). By chance a mean CATpoint of 3 (0.2 by 15 items) is reachable. On the other hand, the online survey shows different results in regard to CATpoints ($M = 8.60$, $SD = 3.82$). Figure 2 shows the distributions. Furthermore, the mean age of the paper pencil survey is low ($M = 16.05$, $SD = 5.86$) compared to the online survey ($M = 31.80$, $SD = 10.71$). A chunk is recognizable for the paper-pencil survey indicating that participants were in average young and performed weakly on CAT (CATpoints). We assume that this is the main reason for the moderate Cronbach's alpha we found for the paper-pencil version.

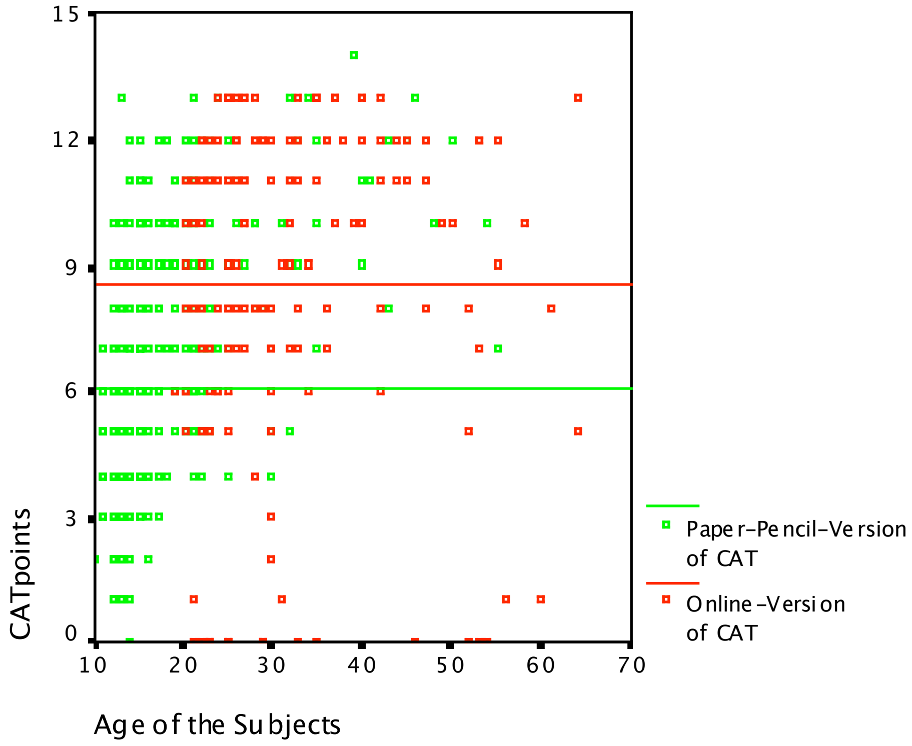

Figure 2. Distribution of the two surveys. Online-participants are older and perform better on CAT.

Table 3 shows a critical corrected-item-total-correlation (-.27) for item 16 (Database3). Further on, it is the only item within the category (Database, see table 1) which is quite easy to solve although it was meant to be the most difficult. Due to these facts we suggest to delete item 16. Finally, the alpha-values show a moderate to high internal consistency although we used a wide area of computer knowledge.

Table 3

Itemcharacteristics for the mixed dataset

Category / Item-Position	Item Difficulty	Corrected-Item-Total-Correlation	Alpha if Item Deleted
Computer1 / 1	.40	.52	.71
Compute2 / 11	.84	.34	.73
Computer3 / 17	.10	.22	.74
Database1 / 4	.45	.49	.72
Database2 / 10	.43	.48	.72
Database3 / 16	.70	-.27	.79
Internet1 / 7	.27	.43	.72
Internet2 / 15	.21	.05	.76
Internet3 / 18	.16	.40	.73
Office1 / 5	.86	.32	.74
Office2 / 9	.48	.40	.73
Office3 / 13	.29	.47	.72

Table 3 *(continued)*

Itemcharacteristics for the mixed dataset

Search1 / 2	.42	.46	.72
Search2 / 8	.64	.45	.72
Search3 / 12	.23	.50	.72

Note: Cronbach's alpha = .75 (n = 818)

6.2 Examining the Hypotheses

To test our first assumption (male subjects perform better than female subjects) all participants who completed the tested categories and provided useful data (gender and age) were taken into account. We compared performance on CAT (CATpoints) of the female and male participants within both the onlinesurvey and the paper-pencil survey. T-tests indicate that in both conditions (online, paper-pencil) male participants performed significantly better than female participants. Results for the paper-pencil survey were $t(1{,}645) = 3.314$, $p = .035$ (one-tailed) (male, $M = 6.28$, $SD = 2.91$; female, $M = 5.89$, $SD = 2.42$; $d = .16$), and for the online survey $t(161) = 3.24$, $p < .001$ (one-tailed) (male, $M = 9.39$, $SD = 3.96$; female, $M = 7.48$, $SD = 3.49$; $d = .51$). Thus gender differences in performance can be confirmed.

The second and third assumptions deal with the correlations of computer usage with the performance in CAT (CATpoints) and Internet usage with the performance in CAT. Pearson's product moment correlations (one-tailed) were used. Pearson's r indicate that computer usage correlates moderately with CATpoints, $r(776) = .51$, $p < .001$. Further Pearson's r between Internet usage and CATpoints also indicates a lower but significant correlation, with $r(776) = .36$, $p < .001$. This shows that the extent of computer usage as well as the extent of Internet usage has an influence on computer knowledge.

Hypothesis 4 predicts that participants who completed the online version of CAT will perform better on CATpoints than participants who completed the paper-pencil version. An independent t-test was conducted. Results indicate that online participants perform better than paper-pencil participants $t(1{,}796) = 9.59$, $p < .001$ (one-tailed) (online, $M = 8.51$, $SD = 3.86$; paper-pencil, $M = 6.05$, $SD = 2.63$; $d = .75$). This shows that online participants have better computer knowledge.

Hypothesis 5 predicts that the higher participants score on the CARS scale the worse they perform on CAT. To test this we excluded participants who were younger than 18 years. Pearson product moment correlation (one-tailed) was used. Pearson's r indicates a negative correlation at a moderate level., $r(298) = -.40$, $p < .001$. This shows that computer anxiety has an impact on computer knowledge.

6.3 Increase in Difficulty

A useful characteristic of a test is an increase-in-difficulty-effect (Bortz & Doering, 2002). As figure 3 shows, there is an increase in difficulty within a category when the first two items are compared to the last one. However, it is worth considering changing some of the items' positions to achieve a smoother increasing in difficulty (Computer1 vs. Computer2, Search1 vs. Search2, see figure 3). But since CAT is a "mixup" of different categories, this enhancement is assumed to have a minor effect. As previously mentioned, we collected further data from online participants, in particular self-perceived difficulty of the items. For each item participants rated how difficult they considered the item to be. We supposed that participants rate the later items as more difficult compared to the first ones. To test this we built a mean value of perceived difficulty for the first 8 items and an analogous value for the last 7 items for each online participant. A paired-sample t-test (one-tailed) indicates that there is a significant difference in self-perceived difficulty, $t(145) = 7.551$, $p < .001$ (first-item-group, $M = 3.85$, $SD = .99$; .last-item-group, $M = 4.28$, $SD = 1.10$, $d = -.41$).

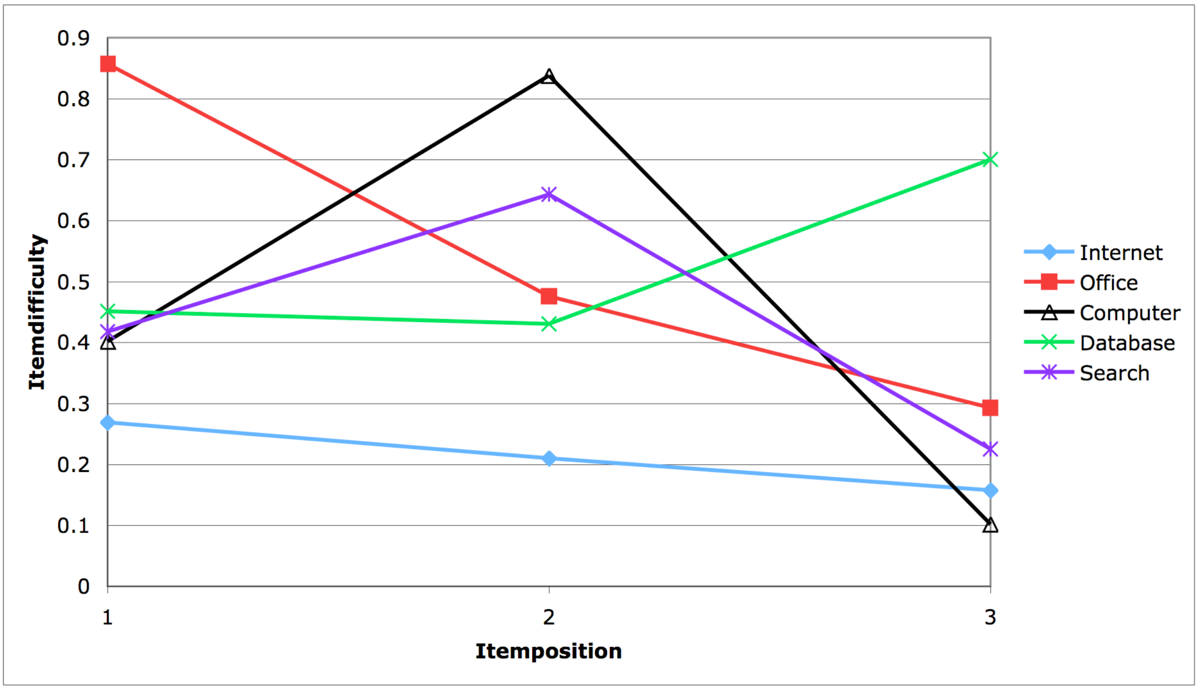

Figure 3. Item difficulties for the "mixed" data set.

7 Discussion

This study introduced a new computer proficiency test (CAT), which is based on a test developed by Bradlow et al. (2002). A new proficiency test was primarily developed as a validated instrument for the German speaking population. Furthermore, the test by Bradlow has some problematic answer items, e.g., the subdomain “printing”.

Van Braak (2004) could show that “computer-confidence” has an influence on “computer-competence”. Therefore, we included a category “attitude towards computers” (CARS), which assesses computer anxiety. This category is based on Heinssen et al.’s (1987) Computer Anxiety Rating Scale.

Analyses of the internal consistency regarding each version of CAT (online, paper-pencil) indicated moderate to high internal consistency, with a Cronbach’s alpha of .88 for the online version, .67 for the paper-pencil version, and .75 for the mixed version. The range of the α-values indicates a single latent construct: computer knowledge. Item analyses indicated that elimination of item 16 leads to a higher internal consistency level.

Between the online and paper-pencil version of CAT, significant differences were found. This is in line with the different Cronbach’s alphas for the online, paper-pencil and mixed version. These results may be due to the age structure of the participants of the paper-pencil survey. The mean age in the paper-pencil survey is significantly lower than in the online version. Thus, participants of the paper-pencil survey had fewer possibilities to acquire computer knowledge than participants of the online survey.

Consistent with findings from Beckwith et al. (2005), our results demonstrated that male participants performed significantly better than female participants.

As expected, the present study found significant correlations between computer- and Internet usage and the performance on CAT. In line with the findings of van Braak (2004), we also found a significant correlation between computer anxiety and computer knowledge. This allows for use of the CARS scale as a control scale. However, inclusion of CARS items is not mandatoty for scientific research when computer knowledge is the main point of interest.

In summary, CAT has been shown to be an appropriate tool to assess computer knowledge. However, some limitations should be noted. As previously mentioned, the significant age difference between the participants of the paper-pencil survey (70.5 % were 15 years old or younger) and the participants of the online survey limits comparability. Therefore, further data collection with an adult sample for the paper-pencil version is essential. CAT is a instrument, which assesses a wide area of computer knowledge. However, when a detailed analysis is needed, specialized instruments or larger interviews are inevitable.

In conclusion, this investigation shows that CAT demonstrates good internal consistency and meets the criteria we set ourselves. It is in German, it is quick to complete, and simple to transform in an online version. Therefore, CAT is an adequate screening tool for scientific research on computer proficiency.

8 References

Bargas-Avila, J. A., Oberholzer, G., Schmutz, P. de Vito, M., & Opwis, K. (2006). *Usable error message presentation in the World Wide Web: Don't show errors right away*. Manuscript submitted for publication.

Beckwith, L., Burnett, M., Wiedenbeck, S., Cook, C., Sorte, S., & Hastings, M. (2005). Effectiveness of end-user debugging software features: Are there gender issues? *ACM Conference on Human Factors in Computing Systems*, Portland, OR, USA, April, 2005, 869-878.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesely.

Bortz, J., & Doering, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschafter* [Research Methods and Evaluation for Social Scientists] (3. überarb. Aufl.). Berlin: Springer.

Bradlow, E. T., Hoch, J., & Hutchinson, J. W. (2002). An assessment of basic computer proficiency among active Internet users: Test construction, calibration, antecedents and consequences. *Journal of Educational and Behavioral Statistics, 27* (3), 237-253.

Buse, L. (1976). Zur Interpretation einer Lügenskala. *Diagnostica, 2*, 34-48.

Evans, G. E., & Simkin, M. G. (1989). What best predicts computer proficiency? *Communications of the ACM, 32* (11), 1322- 1327.

Heinssen, Jr., R., Glass, C, and Knight, L. (1987). Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behavior, 3*, 49-59.

O'Hanlon, N. (2002). Net knowledge: Performance of new college students on an Internet skills proficiency test. *Internet and Higher Education, 5*, 55-66.

Olsson, C. (2000). *The usability concept re-considered: A need for new ways of measuring real web use*. In Proceedings of IRIS 23, Laboratorium of Interaction Technology, Uddevalla, Sweden.

Schmutz, P. (2004). Der Foot-in-the-Door Effekt. Ein Mittel zur Reduktion von Drop-Out in Online-Untersuchungen? Unpublished master's thesis, University of Basel, Switzerland.

Swiss Federal Statistical Office. Internet-Verfügbarkeit 2005. Retrieved April 30, 2006, from:http://www.bfs.admin.ch/bfs/portal/de/index/themen/systemes_d_indicateurs/indicateurs_de_la/approche_globale.indicator.30106.html?open=302

Van Braak, J. P. (2004). Domains and determinants of university students' self-perceived computer competence. *Computers & Education, 43*, 299-312.