Master Thesis
Handed in April 29, 2009

Master program “Cognitive- and Neuroscience”
Faculty for Psychology
University of Basel

The Indirect Influence of Counter System Usability on Customer Satisfaction

Prof. Dr. Klaus Opwis (primary reviewer)
M.Sc. Peter Schmutz (secondary reviewer)

B.Sc. Jakob Boss
Buchenstrasse 42
CH-4054 Basel
+41 79 757 34 94
jakob.boss@stud.unibas.ch

Abstract

Helpdesks and counters in public transportation are often only waypoints on a tight scheduled multipurpose journey. The personnel of such service points need a system that support their work efficiently and in a usable way. The aim of this study was to examine the relationship between the usability of the counter system and the overall customer satisfaction. In order to record required data, an adapted usability questionnaire and a customer interview was used as well as observations of performance values. Statistical analysis reveiled two significant interactions. The first relationship exists between the expertise of the counter personnel and the duration of an interaction. The second finding was a negative association between the usability values and the customer satisfaction. These results indicate the personnel not only being a transmitter of information, but rather a powerful filter and translator of inputs and outputs on his system. The practical implication of these findings is to improve the counter system usability in order to effectively facilitate customer service and, therefore, overall customer satisfaction.

Key words: Usability, customer satisfaction, expertise, counter, ticketing, helpdesk, public transportation

Acknowledgements

This study is a cooperative project between the *Swiss Federal Railways* company (SBB) and the *University Basel*. I would specially thank Bruno Spicher from the prior and Peter Schmutz from the latter as well as the management and all involved personnel from the three trainstations for their kind and helpful support. Olgu Günal specially drew figure 1 for this study and Beda Kamm revisited the text for errors which both improved this work in a significant matter.

Contents

1 Introduction

The modern human is closely linked with both increased mobility and limited time. Whereas mankind managed to domesticate electricity for the transport of informantion where physical dislocation is negligible the vibrating competition between time and mobility still exists for situations where the physical substitution with alternating current is not sufficient. The origin of the race between the two unlike twins time and mobility remains unclear but the result is always the catch up of the temporarily second twin as soon as the the other makes even a little step ahead – when schedules become thighter, the urge for better, more reliable mobility arises and vice versa. The knots in the public transportation system form the fast-paced and turbulent peaks in this competition on public grounds.

One way to counter the limited time of travellers consists in improving the functionality and effectiveness of the transportation centers for a higher throughput and passenger workload. But the most common and economical prosperous alternative answer is to allow the execution of multiple tasks along the time-consuming dislocation process. The added complexity and requirements for a transportation center may hinder the actual transit performance, but the gain using otherwise unproductive time totals up in a clear benefit for most use cases. Because of the almost reached physical constraints and limits of pedestrian movement and the uncertainty about and lack of basic insight into the aims and intentions of the travellers (Weidmann, 1993), it can be assumed that significant changes in public transportation in the near future appear in the open field of multitasking during transit. The introduction of wireless internet access on trains (SBB, 2009a) and the furnishing of major transportation centers with shopping, service, gastronomical and leisure facilities with extended opening hours and mall characteristics (SBB, 2009b) give an fascinating outlook on the seamless combination of home, work and recreational activities while travelling in the near future.

The complexity of the modern multitasking traveler demands concentrate at the ticketing and information counters in a trainstation center. Without modern assistive equipment in form of computer hard- and software the work on the counter could not be done nowadays. In the past few years new computer systems have been introduced to the counter. But the application of new and advanced systems does not guarantee the appropriateness and usable support from these systems. Although the equipment of the counter is highly complex and extensively used in professional context there is little insight in the usability of the system and its possible effects on customer satisfaction. The aim of the current study is to assess whether the computer systems used in swiss trainstation counters propose adequate support for the personnel and to identify the impact of the system usability on customer satisfaction.

2 Theoretical background

When a traveler visits the counter of a transportation center a cycle of interactions is set in motion. The initiation could be the demand for a ticket with the subsequent exchange of different types of information as depicted in Appendix B. The informations within the cycle of interactions follows two main paths – the first from the traveller to the counter personnel and back and the second between different inputs of the personnel and the respective feedback of the counter system (c.f. Figure 1). This open triangle is somewhat similar to the one proposed by Bevan, Kirakowski, and Maissel (1991). They suggest three different aspects of usability: the system-, user- and contextual view. For the situation on the couter the personnel forms a human interface for a specialised and complex computer system the traveller could barely interact with at first sight. The traveller is the context (or general task) for the counter system and, at the same time, the user of the computer system by indirectly interacting with the human interface.

Figure 1. The components and pathways of interaction for a counter visit. The customer demands a service and the clerk operates the computer system accordingly. The computer then requests additional information or prints a result on the screen or on paper. The clerk hands over the final result to the customer or asks for the additional information upon which the circle starts again.

2.1 Counter System Usability

The impact of usability as one possible factor for system acceptability is a wide and extensively explored scientific field. Usability itself segements into different aspects like efficiency, learnability or satisfaction as shown in Figure 2. These supposedly distinct aspects proved to be somehow interdependent in various research. Learnability affects memorability and both make contributions to efficiency and errorless interaction. Younger people for example generally *learn* computer skills easier (Gist, Rosen,

& Schwoerer, 1988; Kang & Yoon, 2008) but the effect of *experience* on performance and therefore *efficiency* is also clearly stated in the context of a database search (Mead, Sit, Rogers, Jamieson, & Rousseau, 2000). Experience benefits were even found to increase with age and therefore experienced older people could compensate their drawbacks in terms of learnability in a study from Czaja (2001). The dependence between different memorability concepts, efficiency and errorless interaction was unveiled in another online library catalog search study where Sit (1998) found *conceptual knowledge* as of more importance on performance and errorless interaction than *semantic knowledge* or *technical skills* especially for older users.

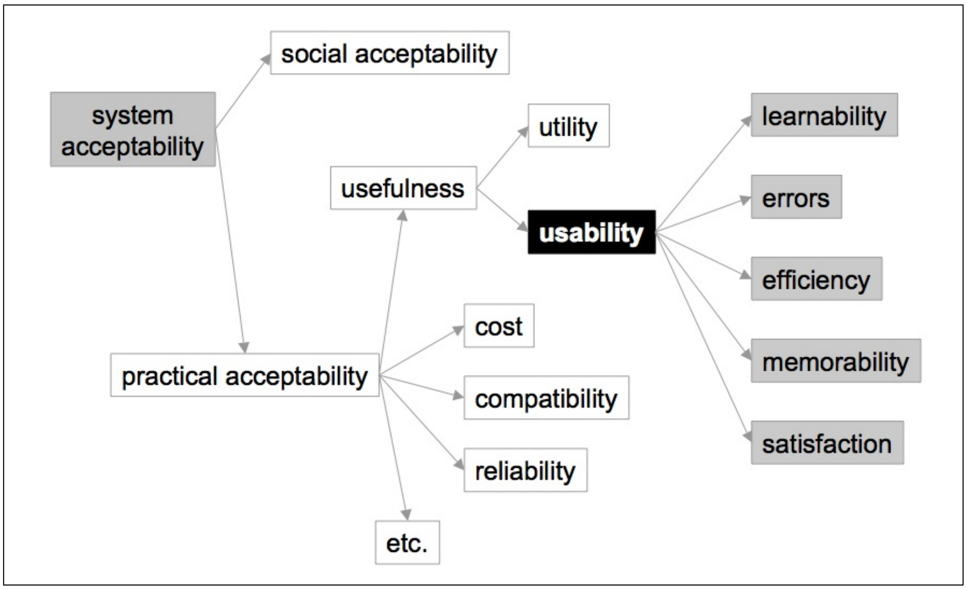

Figure 2. Composites of system acceptability with usability and its aspects on the right-hand side as one possible factor of influence (Nielsen, 1993).

The relationship between satisfaction and usability is a controversial issue in science with different concepts and models applied in research. Whereas Frøkjær, Hertzum, and Hornbæk (2000) suggest treating satisfaction as independent from effectiveness and efficiency, Jeng (2005) found evidence for high correlations between satisfaction and detailed aspects of effectiveness and efficiency. Both autors note the lack of a consistent theoretical model with operationalized parameters to investigate user satisfaction. Apart from the intercorrelations between the aspects of usability the direct effect of expertise on usability is an interesting, but rarely featured issue in usability research. As a by-product of other research, there is some evidence suggesting a negative effect of expertise on usability ratings. In a library search experiment, Wei (2002) found novice users to like the interface in question more than expert groups. A possible explanation is given with the fact that the expert group tends to use more than only one search tool and "that they already know which resources to search for their research topic, and therefore did not need a computer to recommend databases to them" (p. 578).

There is not enough research for the situation in which the user operates a computer system by means of an intermediate specialized human acting as the system interface. In analogy to information processing concepts and general cognitive theories (Anderson, 1996), the task of the personnel could comprise two different aspects – the one of a *signal transmitter* close to the original data and the one of *information translation* where also transformation, alteration and adaptation of the information occurs. The prior view refers to the very first stages of sensory processing whereas the latter resembles the last operations before the well prepared and aggregated information reaches higher cognitive processes. This second alternative provides suitable and useful formats for both the computer and the customer – but with the cost of higher cognitive workload for the personnel in case of corrupt usability.

2.2 Customer Satisfaction

There are several methods to buy a ticket or get a timetable information apart from the counter: ticketing machines, print-out timetables, using the internet and you can even get both with a modern mobile phone. When visiting the counter, a customer accepts a possibly long queuing time, but on the other hand there is an individual service by another human who handles the interaction with the information system – with lots of experience, ideal efficiency and dealing with all system feedbacks and errors. The usability composites learnability and memorability, error handling and efficiency (c.f. Figure 2) are all handed over to the personnel. A correct and friendly communication can even turn a negative effective result due to a system error or mishandling into a overall satisfactory visit for the customer. Measuring satisfaction, therefore, is the key element when assessing the usability of the counter from the perspective of the customer.

One proposal for simultaneously inspecting effectiveness, efficiency and satisfation is the WAMMI-questionnaire of Kirakowski, Claridge, and Whitehand (1998). This inventory was compared to five categories from an unstructured user interview in a study where Lindgaard (2003) concludes that: "[User satisfaction] is a complex construct comprising several affective components as well as a concern for usability and that a priori expectations seem to play a major role in shaping user satisfaction" (p. 447). The five categories expand the definition of satisfaction by the concepts of aesthetics, expectation and affective components, but the answering pattern is similar to the results from WAMMI-questionnaire.

Whether usability is regarded as one determinant for user satisfaction among others or, on the contrary, satisfaction is assumed as a (vague) factor for overall usability it seems obvious that both share at least parts of each other. This assumption is supported by high correlations between satisfaction, effectiveness and efficiency found by Jeng (2005).Therefore, the concentration on measuring satisfaction of the customers as a comprehensible and acceptable substitute for a more detailed, but otherwise interrelated usability questionnaire, seems adequate.

2.3 Hypothesis

Out of all contemplations on the counter situation and from the theoretical research in usability science and customer satisfaction evaluation so far, the following questions arise. With regard to the personnel-computer interaction (1) a reduction in the duration with higher experience can be expected. This reduction in interaction time is supposed to have (2) a positive effect on the usability evaluation. Some literature idicates (3) a high expertise and thus a pronounced criticism leading to reduced usability ratings. Because of the possible indirect effect from experience over duration on usability the last expectation could become leveled out in total.

For the customer-counter interaction the (4) longer durations are presumed to impair the satisfaction ratings in analogy to the usability evaluation. For the relationship between the usability evaluation and the customer satisfaction ratings there are two possible outcomes corresponding to the actual function of the personnel: (5a) As a mere transmitter of information between customer and computer system, there should be a positive intercorrelation. (5b) As an interpreter and translator of information into suitable formats for the respective receiver, the same correlation should be negative because identified usability deficits result in greater efforts to countervail, and therefore enhance the customer's experience.

3 Method

In order to investigate the supposed effects and interactions, an *in situ* examination was conducted. On different counters in major centers of public transportation the following data were recorded, either by means of questionnaires and interviews, or through observation. The satisfaction of the customer was inquired as the final result of the interaction with the counter. The subsequent usability evaluation of the computer system used by the personnel by means of a questionnaire was then set in relation to the customer's satisfaction. Additionally, objective performance values were registered through observation.

3.1 Study design

The trainstations of Basel, Bern and Zurich were chosen as study centers due to their comparable size, passenger frequency and importance in the swiss public transportation grid. The multicenter design allowed control for slight differences in layout and position of the counter inside the facilities, as well as for influences from probable situational disproportionate demands by the customers. Additionally, the formation of observation blocks consisting of roughly seven customers per respective counter personnel ensured the presence of the frequent demands in all observation blocks. This design also counters the disproportion between few counter personnel and many customers. Great care was taken to include both high expertise senior personnel as well as novice juniors.

3.2 Participants

The data of a total of 24 vendors and 168 customers were included in the current study (c.f. Table 1 for details). Sex and age were tested for possible differences between the three study centers. A Kruskal-Wallis-test could not find a difference in the distribution of sex for the three centers ($\chi^2(2) = 1.32$, $p = .517$). Likewise, the one-way ANOVA showed no differences regarding the age of customers for the three centers ($F(2) = .018$, $p = .982$). Because of the intentional selection of a broad range of expertise levels the resulting distribution was highly skewed. Therefore, no statistics regarding differences on these values could be computed, but the range shows comparable spans for all three centers. After all tests for disproportional distribution for the three study centers, the study sample was treated as uniform regarding sex and age of the customers and experience of the counter personnel.

Table 1

Descriptive Characteristics for the three Trainstations in the Study

	Basel	Bern	Zurich	Total
Personnel	8	9	7	24
Expertise [years]	1.5-22	2.5-16	0.25-17	0.25-22
Customers	54	68	46	168
Female [%]	52	62	54	56.5
Male [%]	48	38	46	43.5
Age (*SD*)	50.8 (*20.1*)	50.3 (*20.9*)	51.0 (*20.4*)	50.7 (*20.4*)
Age range	17-86	14-87	18-95	14-95

3.3 Measures

In order to collect the satisfaction ratings, the customers were asked to verbally rate their satisfaction level with the service of the counter they just visited on a ten-point Likert scale ranging from 1 (*not satisfied at all*) to 10 (*absolutely satisfied*). They were also asked about their year of birth. An interview with more questions tested in an early pilot study led to a considerable increase of drop-outs because customers reached their goal after the visit and available time thereafter was low. Furthermore, those drop-outs are of the particularly interesting category of highly time-limited customers.

The usability of the counter computer system was measured using the adapted Computer System Usability Questionnaire (CSUQ) developed by Lewis (1993). For the purpose of this study, the original questionnaire was translated from English into German. A further adaptation to the sitation of the counter was done by reducing the original 19 items to 12 items. This reduction also removed redundance and concentrated on the adequate items for the counter context (c.f. Appendix A for the adapted questionnaire). The items were provided as full sentences to which the counter personnel indicated their level of agreement by means of Likert scales ranging from 1 (*strongly agree*) to 7 (*strongly disagree*). The added entity *no answer* in contrast to the original questionnaire was little used by the respondents.

The limited sample size discouraged the use of a factor analysis for comparison with the original questionnaire. Backhaus, Erichson, Plinke, and Weiber (2005) suggest at least three times more participants than explored variables. An interrater reliability analysis exhibited a Cronbach's Alpha of 0.69 but a total of 7 subjects were eliminated from this calculation because of listwise exclusion. The subscales were subsequently computed using the similarity to the original factors of the CSUQ. The original instruction suggests equal weighting as acceptable. Items 1 to 5 were averaged on subscale *System Use*, 6 to 9 formed *Information Quality* and 10 and 11 counted on subscale *Interface Quality.* Item 12 represents the overall satisfaction and was not assigned to a specific scale. A general score with the average of all items was computed and used for correlative comparisons. In addition to the questionnaire, the personnel were asked about their expertise as counter personnel at the trainstation in years, the frequency of their private computer use and the number of computer operating systems they ever used in total. As the personnel sample was small, sex was not recorded but treated similarly for all study centers.

The duration of the visit and the type of service the customers demanded were gathered through observation. The demands of the customers were categorised as *ticket only*, *subscription tickets*, and *timetable information* – the rest was marked as *other* (c.f. Table 2). The queueing duration could not be registered for practical reasons. Overlapping observations and the frequent changing of customers from a long queue to a shorter one would have required a multiple of the actually disposable observation personnel.

Table 2
Distribution of Customer Demands

Customer needs (N)	Basel	Bern	Zurich	Total
Ticketing only	30	30	34	94
Conscription tickets	10	13	1	24
Timetable information	2	9	2	13
Other	12	16	9	37

3.4 Procedure

The selected personnel were informed about the ongoing investigation at the beginning of each observation block. They were instructed to serve customers as normally as possible. The customers were observed and interviewed consecutively after their visit in front of the counter. When necessary, the customer was accompanied and asked *on the go*. To facilitate the study for the single observer present every second customer was skipped. An observation block with 5 to 10 customers lasted for about 30 to 45 minutes. At the end of each block the personnel was asked to fill out the adapted CSU-Questionnaire in retrospect of the whole block. The observation block was then repeated with a different counter or another counter clerk.

3.5 Data Preparation

Customer data was separately checked for normal distribution before computing aggregated values from all customers of one observation block. The satisfaction ratings of the customers showed a pronounced ceiling effect with over 3/4 giving a full 10 on the scale. Only ten (or 6%) customers gave a rating below 8. The merging of values of the same observation block encourages the further treatment as a continuous variable because the resulting group means show differences in fractions of the original rather discrete nature of Likert scales.

The type of observed demands differed significantly for the three study centers ($\chi^2(6) = 15.54$, $p < .05$). The only demand that was present in approximatively even proportions in sufficent high numbers in all three centers was *ticketing only* (c.f. Table 2). An additional cluster analysis for the duration data – using the type of observed demands as to be predicted – showed an unsurprisingly good separation between two groups (c.f. Table 3). The duration values of customers asking for a ticket only are located around the peak between 30 and 60 seconds with a relatively narrow distribution, whereas complex or multiple demands spread on a long slope towards the maximum duration of seven minutes. These two premises endorsed the decision to concentrate on the *ticketing only* case for further statistical analysis.

Table 3
Duration of Customer Demands

	M	*SD*	Range
Ticketing only (Cluster 1)	58.9	31.6	10 – 175
Conscription tickets	146.9	83.7	40 – 411
Timetable information	72.8	45.0	12 – 152
Other	117.4	110.6	10 – 420
Cluster 2 centroid	119.1	96.0	

Note. Correction for skewed values not yet applied.

The last step towards a normal distribution of the still slightly skewed duration data was achieved by taking the logarithm to the basis 60 (a value of 1 would be aequivalent of 1 minute duration). The expertise levels of the counter personnel showed similar skewed distribution caracteristics and after taking the logarithm of the wide spread expertise values a distribution with the two expected peaks on both ends was achieved. The logarithm was chosen to the basis 4 to separate between novices (three year apprenticeship plus one year of ”coming to grips“) and experts with several years or even decades of counter service.

3.6 Outlier Correction

The analysis of the CUSQ subscales revealed two subjects with values exceeding three standard deviations from the respective mean. Another subject had several missing values in both the questionnaire and expertise level data. These subjects were therefore excluded from further analysis. In spite of the small sample size a multivariate outlier analysis using Mahalanobis distances was applied to the four main variables duration, expertise, usability evaluation of the counter personnel and satisfaction ratings from the customers. Two subjects attracted attention, although their Mahalanobis distances (11.9 and 8.9) missed the χ^2 exclusion criterion for $\alpha = .001$ (for $\alpha = .05$ one subject would be clearly identified and the second one nearly). An analysis of particular combinations of variables revealed a crucial impact on specific interactions from these two subjects. Altogether, these findings led to the exclusion of both subjects for further analysis. There were 19 observation blocks left after analysis and exclusion of outlier data.

4 Results

As the present study tries to illuminate the processes and factors involved in the formation of the final customer satisfaction with the counter service of a trainstation, an appropriate structural equation modeling would have been the experimental tool kit of choice. Several restrictions from the experimental situation and practicability issues finally led to an alternative approach. A mere total of 24 interaction profiles would have led to doubtful and unstable results even from a very simple statistical model for the interaction. Apart from descriptive results an intercorrelation analysis was performed as well as methods to *post-hoc* refine possible correlations in more detail. An alpha level of .05 was used for all statistical tests.

4.1 Descriptive Statistics

The *per counter* aggregated customer satisfaction ratings and the derived subscales of the CSU-Questionnaire are shown in Table 4. The expertise of the personnel is listed as original data and after taking the logarithm to the basis 4. The aggregated mean duration (M_c) data, based on the original duration (t_i) observations of the visit, were reverse computed using Formula 1 to facilitate the explanation of the data. The reverse computation resulted in similar results to the original duration data (c.f. Table 3 in Section 3) and can, therefore, be regarded as good estimates of these data. The analysis exhibits a good overall rating for the counter computer system from both customer and personnel.

Table 4
Descriptive Statistics for the Main Variables ($N = 20$)

	M	SD	Range
Customer satisfaction	9.7	.36	9.0 – 10.0
CSUQ overall	5.5	.38	4.8 – 6.1
Subscale *System use*	5.9	.46	5.0 – 6.4
Subscale *Information*	4.9	.71	3.5 – 6.0
Subscale *Interface*	5.4	.61	4.0 – 6.5
Expertise [years]	6.7	6.7	.25 – 22
Expertise (log_4)	.98	.83	-1 – 2.23
Duration (log_{60})	.98	.08	.79 – 1.10
Duration [s] (comp.)	58	16	25 – 90

$$t' = 60^{M_c} \text{ with } M_c = \frac{\sum_{i=1}^{n} log_{60}(t_i)}{n} \quad (1)$$

4.2 Correlations

The data gathering and aggregating finally culminates in the correlative comparison of the main four variables duration of the visit, expertise with the system, usability evaluation of the counter personnel and satisfaction rating of the customers shown in Table 5. The correlation between the expertise and the duration is significant ($r = -.40, p < .05$, one-tailed). The expertise shows another influence on the usability evaluation almost reaching statistical significance ($r = -.37, p = .06$, one-tailed). The negative relationship between the usability evaluation of the personnel and the satisfaction rating of the customer is again significant ($r = -.57, p < .05$, two-tailed).

Table 5
Correlation Coefficients for the Main Variables Expertise, Duration, Usability and Satisfaction

	Expertise	Duration	Usability
Expertise (log_4)	1		
Duration (log_{60})	$-.40^{*}$	1	
Usability (personnel)	-.37	.19	1
Satisfaction (customer)	.07	-.12	$-.57^{**}$

Note. $^{*}p < .05$ (one-tailed), $^{**}p < .05$ (two-tailed), $N = 19$

To exclude possible influences of other variables on the significant correlations the respective partial correlations were computed. For the correlation between expertise and duration using other variables as control results in missing significance levels ($r_{ed.u} = -.36, p = .071; r_{ed.s} = -.40, p = .053$, all one-tailed). The same applies to the use of both other variables as controls ($r_{ed.su} = -.37, p = .072$, one tailed). When including both the expertise of the personnel and the duration of the interaction as control variables for the bivariate correlation between the usability evaluation and the customer satisfaction ratings, the significance remains ($r_{su.ed} = -.58, p < .05$, two-tailed). Also the tests with single variable as a control still reach significance levels ($r_{su.e} = -.58, p < .05; r_{su.d} = -.56, p < .05$, both two-tailed).

A *post-hoc* median split for the usablity evaluation values of the counter personnel was done with $x < 5.5$ for the lower group. The results of the mean comparisons of all variables for the two resulting groups (low vs. high CSUQ values) are shown in Table 6. The only variable with a significant difference between the means of the median split groups is the customer satisfaction rating.

Table 6
Comparison after post-hoc *Median Split for the Usability Data*

	Usability	M	SD	F	p
Expertise (log_4)	low	1.27	.70		
	high	1.00	.73	.70	.42
Duration (log_{60})	low	.98	.08		
	high	1.00	.06	.39	.54
Usability (personnel)	low	5.1	.34		
	high	5.8	.21	29.6	$< .001$
Satisfaction (customer)	low	9.9	.17		
	high	9.6	.37	5.4	$< .05$

Note. Separation value at $x_U = 5.5; N_{low} = 9, N_{high} = 10$

5 Discussion

The results from this study show a clear negative relationship between customer satisfaction and counter personnel usability evaluation. This finding supports the assumption of the personnel being not only a transmitter but a rather effective translator between computer and customer. The significant intercorrelation between expertise and duration which adds to a long tradition in research showing a clear learning effect on performance. The impact of interaction of duration on both usability evaluation and customer satisfaction as well as the interaction between experience and usability evaluation could not be confirmed.

5.1 The Counter: Expertise, Performance and Usability evaluation

The performance of the entire counter system – measured as the duration of customer visits – is significantly linked with expertise. There is broad psychological evidence for this type of learning effect and its implication for aquisition of skills and the impact on performance. Anderson (1996) states in his comprehensive book on cognition that the effect occurs in memorization as well as with procedural knowledge. Various studies found the analog effect for the field of database search and information retrieval (Czaja, 2001; Mead et al., 2000). In the present study, expertise impacts performance positively although the tasks were different for each customer. Expertise, therefore, seems to affect single actions, but also to account for the general performance of operations of a similar type. In spite of the actual significant relationship, the impact of expertise is somewhat marginal with a small mean reduction of about 9% in duration only after a capacious increase from four to sixteen years of experience. Because of this unfavorable cost-benefit ratio, the impact of actual personnel training on the same subject after the first few years is suspected to rapidly decrease.

No intercorrelation was found for the expected relationship between duration and the usability evaluation. In their literature meta-analysis Nielsen and Levy (1994) list a variety of research results with the expected relation but they also note few examples of contrary findings akin the ones in the present study. They report that some users or user groups seem to prefer techniques with which they perform inferior, compared to other methods at hand, unless a certain threshold in performance deficit difference between the actual und the alternative system is reached. The negative correlation between expertise and the usability evaluation, closely missing significance level in this study, is similar to the indications in the research of Wei (2002). Because no structural equation modeling could be applied, the question whether the usability evaluation in this study is rather based upon preference of the personnel than upon the actual performance can not be answered.

5.2 The Customer-Counter Interaction

The predicted influence of the interaction duration on the customer satisfaction ratings could not be found in the actual data. In analogy to the unconfirmed relation between duration and usability evaluation, this could again be caused by a rather preference than performance based review of the customers (Nielsen & Levy, 1994). Another explanation lies in the fact that only the duration of the actual visit was gathered, but not the overall waiting and queueing time. This detailed examination should be subject of further research in combination with an adequate structural equation modeling.

The last and probably most thrilling question, this study tries to solve, is the one about the role of the personnel between customer and computer. The negative correlation between usability evaluation and customer satisfaction induces the conclusion that the personnel not only serve as a specialized and trained interface, but also as a powerful filter, interpreter and transformer of information into otherwise incompatible formats. An example for such a transformation is portrayed in Appendix B – the demand of the customer had to be collected entirely and several interpretations had to be done before any input in the computer could be made.

Because there is no similar research at hand, the comparison with rather general psychological findings is undertaken. A first approach in the interpretation of the negative correlation follows the thoughts already portrayed in section 2. Our sensory system provides higher cognitive processes with confortably prepared information we would barely master in full consciousness (Anderson, 1996; Palmer, 1999). Another promising approach provides the concept of cognitive compensation found in clinical neuropsychology and development psychology. The human as the flexible part of the entire counter could try to countervail possible detected deficits of the computer system with enhanced friendlyness, attentiveness or (positive) feedback towards the customer. Those who expressed difficulties with the system seemed to effectively compensate, resulting in higher customer satisfaction ratings.

The practical implication of the negative correlation between the usability evaluation and the customer satisfaction requires interpretative precaution. One could argue to intentionally deficit the usability of the system for decreased usability evaluation and therefore better customer satisfaction. Another faulty interpretation would be to specially train the personnel in identification and expression of usability issues with the system. An approach with higher personnel and customer satisfaction at the same time would be to improve the system usability. On the one hand, the system acceptance and, therefore, work satisfaction of those who express difficulties in using the system would be improved. On the other hand, the customer satisfaction ratings of those who are unaware of system deficits because of unconsciously minimizing cognitive workload.

5.3 Study Limitations and Further Research

The small sample size of counter personnel limits the study to correlative statistics. A structural equation modeling could preferably be applied in order to conduct detailed analysis of the customer-counter-computer interaction but requires considerably more investigated subjects. More subjects would also allow the application of a factor analysis to determine the validity of the adapted CSU-questionnaire in detail. Also the found effects are probably limited to railway station counters. As the situation with an intermediate human operating a computer system exists in public transportation counters as well as in airports, banks, official and commercial services and many more, the effect in such environments should be investigated conducting further research. Apart from this broad field for future experiments the detailed interaction effects could be checked including the observation of waiting time and possible other variables which were not included in the present study.

References

Anderson, J. R. (1996). *Kognitive Psychologie.* Heidelberg; Berlin: Spektrum, Akademischer Verlag.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2005). *Multivariate Analysemethoden.* Heidelberg: Springer Verlag.

Bevan, N., Kirakowski, J., & Maissel, J. (1991). What is usability. In H. J. Bullinger (Ed.), *Advances in human factors / ergonomics* (Vol. 18, pp. 651–655). Stuttgart, Germany.

Czaja, S. (2001). Examining age differences in performance of a complex information search and retrieval task. *Psychology and aging*, *16*(4).

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems* (p. 345-352). New York, NY, USA: ACM.

Gist, M., Rosen, B., & Schwoerer, C. (1988). The influence of training method and trainee age on the acquisition of computer skills. *Personnel Psychology*, *41*(2), 255-265.

Jeng, J. (2005). What is usability in the context of the digital library and how can it be measured? *Information technology and libraries*, *24*(2).

Kang, N. E., & Yoon, W. C. (2008). Age- and experience-related user behavior differences in the use of complicated electronic devices. *International Journal of Human-Computer Interaction*, *66*(6), 425–437.

Kirakowski, J., Claridge, N., & Whitehand, R. (1998). Human centered measures of success in web site design. In *Proceedings of the Fourth Conference on Human Factors & the Web.* Basking Ridge, NJ.

Lewis, J. R. (1993). *IBM computer usability satisfaction questionnaires: Psychometric evaluation and instruction for use* (Tech. Rep. No. 54.786). Human Factors Group, IBM.

Lindgaard, G. (2003). What is this evasive beast we call user satisfaction? *Interacting with computers*, *15*(3).

Mead, S. E., Sit, R. A., Rogers, W. A., Jamieson, B. A., & Rousseau, G. K. (2000). Influences of general computer experience and age on library database search performance. *Behaviour and information technology*, *19*(2), 107–123.

Nielsen, J. (1993). *Usability engineering.* Boston: AP Professional.

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Commun. ACM*, *37*(4), 66–75.

Palmer, S. (1999). *Vision science.* Cambridge, Massachusetts: The MIT Press.

SBB. (2009a). *SBB Online im Zug.* [Online] Retrieved April 7. 2009, from http://mct.sbb.ch/mct/reisezeit/dienstleistungen/online-im-zug.htm.

SBB. (2009b). *SBB RailCity.* [Online] Retrieved April 7. 2009, from http://mct.sbb.ch/mct/immobilien/immobilien_railcity.htm.

Sit, R. A. (1998). Online library catalog search performance by older adult users. *Library & Information Science Research*, *20*(2), 115 - 131.

Wei, M. (2002). A database selection expert system based on reference librarian's database selection strategy: A usability and empirical evaluation. *Journal of the American Society for Information Science and Technology*, *53*(7), 567-580.

Weidmann, U. (1993). *IVT Schriftenreihe: Transporttechnik der Fussgänger* (Vol. 90). Zürich, Switzerland: ETH.

Appendix A

A.1 Computer Use and Counter Expertise Questions

1. Wie viele Jahre (bzw Monate; in ganzen Zahlen) arbeiten Sie bereits an dem Computer im Schalterbereich?

2. Wie oft benützen Sie einen Computer im privaten Rahmen (eine Antwort ankreuzen)?

   O täglich eine Stunde oder mehr O täglich bis eine Stunde O mehrmals pro Woche O etwa ein Mal pro Woche O etwa ein Mal pro Monat O wenige Male pro Jahr O nie

3. Wie viele verschiedene Computersysteme (Bsp: Windows 98, Windows XP, Mac OS X, Linux,...) haben Sie bereits privat und beruflich verwendet (inklusive aktuelles System)?

   Anzahl verschiedener Computersysteme:

A.2 Likert Scale for CUS Questionnaire

The following Likert scale including a "no answer" choice and four lines to add a commentary were used for each of the 12 CSU-Questionnaire items:

Stimmt gar nicht	1	2	3	4	5	6	7	Stimmt genau	keine Antwort
Kommentar:									

A.3 Adapted CSU-Questionnaire

1. Es ist einfach, das Computersystem zu benützen.

2. Ich kann meine Arbeit effektiv erledigen (Ich erreiche mein Ziel), indem ich das Computersystem verwende.

3. Ich kann meine Arbeit effizient erledigen (Ich erreiche mein Ziel schnell und einfach), indem ich das Computersystem verwende.

4. Ich finde es angenehm, das Computersystem zu verwenden.

5. Es war einfach, das Computersystem zu lernen.

6. Das Computersystem zeigt mir Meldungen an, die klar sagen, wie Ich ein Problem schnell und einfach beheben oder einen Fehler rückgängig machen kann.

7. Die Informationen über das Computersystem (Hilfe-Menu, Dokumentationen) sind einfach zu verstehen und effektiv (hilfreich).

8. Die Informationen über das Computersystem (Hilfe-Menu, Dokumentationen) sind einfach und schnell zu finden.

9. Die Organisation der Informationen und Elemente auf dem Bildschirm ist klar.

   Achtung: Mit dem Begriff „Interface“ sind auch sämtliche Geräte und Elemente gemeint, mit der Sie mit dem Computersystem interagieren können. Beispiele für Interface-Elemente sind die Tastatur, die Maus, der Bildschirm, das Programm (inklusive graphische und sprachliche Darstellung).

10. Die Benützung des Interfaces des Computersystems ist angenehm.

11. Das Computersystem hat alle Funktionen und Möglichkeiten, die Ich von ihm erwarte.

12. Ich bin insgesamt zufrieden mit dem Computersystem.

Appendix B

The following tables illustrates an interaction on one counter depicting actions and provided information of the three parts customer, personnel and computer. The computer could be operated with a mouse or with the keyboard but some counters were not equipped with the mouse and none acutally used a mouse regularly. Lists could be accessed using the arrow keys, but experts frequently used direct shortcuts with system numbers. The numbers are shown in the lists but only if learned by heart they can account for shorter interaction with the system. All unrelated communication is ommited, action and comments in italic font.

CUSTOMER	PERSONNEL	COMPUTER
		Initial state input fields: - Ticket type - Destination (*cursor*) - Departure - Number (normal) - Number (half-price) - Date valid *additional list with approximately 200 vouchers and special tickets on the right*
Freiburg im Breisgau	*Interpret: one item only, storage as destination -> infer: departure ≙ in place ≙ Basel; interpret: international ticket; Storage all*	
Second class, half-price, one-way	*Storage* change to and select ticket type international, one-way	*Remove number fields (only single international tickets available); reduces voucher and specials list to 20 applicable items*

CUSTOMER	PERSONNEL	COMPUTER
	Recall and input destination, departure	*Shows changes*
	Recall vouchers change to and select appropriate voucher / class combination from list on the right using keyboard shortcuts	*Shows changes, calculates and shows ticket price*
Timer: 12 seconds		
	tells price to customer, confirms printout	*Prints ticket, informs about ongoing print-out*
Pays for ticket	*Transfers ticket for money*	*Resets to default*
Final timer: 25 seconds		