Predicting Post-Task User Satisfaction With Weibull Analysis of Task Completion Times

Bernard Rummel
User Research Expert
SAP SE
Dietmar Hopp-Allee 16
D-69190 Walldorf
Germany
bernard.rummel@sap.com

Abstract

Task completion times have been shown to follow Weibull distributions, with parameters reflecting different aspects of the task solution process (Rummel, 2017). The offset time matches UI operation time, including system response time, on the shortest path taken by users ("click time"). The characteristic time describes the solution rate in the stochastic process of users solving a task ("think time"). The shape parameter captures non-stochastic positive or negative influences on user performance ("acceleration").

This study investigates how these parameters contribute to user satisfaction. Three-parameter Weibull distribution models were fitted to task completion times from 68 tasks in summative usability tests of business applications. Weibull parameters explained 66.5% of variance in post-task user satisfaction ratings. Estimations of relative importance indicate characteristic (think) time as the dominant predictor, contributing roughly twice as much as the two next important predictors, task completion rate and offset (click) time, which explained roughly equal amounts of variance. The Weibull shape parameter (acceleration) contributed the least.

Keywords

usability metrics, satisfaction, efficiency, effectiveness, task completion time, task completion rate, Weibull analysis

Introduction

User interface efficiency is commonly measured using task completion times (Coursaris & Kim, 2011; Hornbæk, 2006; Molich et al., 2010; Sauro & Lewis, 2009). The extent to which efficiency is related to user satisfaction has been debated by numerous authors, coming to different conclusions. While, for instance, Frøkjær, Hertzum, and Hornbæk (2000) and Hornbæk and Law (2007) claimed independence of the concepts, Sauro and Lewis (2009) found strong correlations between usability metrics that, however, were attenuated when user satisfaction was measured with post-test questionnaires (as opposed to post-task ratings). More recently, Strohmeier, Mikkola, and Raake (2013) found task completion time to even be "the key influencing factor on QoE [Quality of Experience] for task-driven Web-QoE evaluation" (p. 38).

These seemingly contradictory findings highlight the necessity to consider a variety of conceptual and methodological aspects in their interpretation. Conceptually, Hassenzahl (2001) pointed out that user satisfaction is driven by both pragmatic and hedonic aspects of the user experience. Their relative importance depends on the genre of the software under investigation—obviously, pragmatic aspects are more important in business software than in entertainment systems. More pointedly, in a later paper, Hassenzahl, Kekez, and Burmester (2002) postulated that the importance of a software's pragmatic quality depends on whether the user is more in a "goal mode" vs. "activity mode." In goal mode, pragmatic quality would play a greater role than in activity mode. In a usability test, the systematic instructions given to test participants to perform certain tasks potentially induce either one of these modes. If the researcher's interest is to test mainly within task performance parameters, they are likely to choose a procedure that involves clear task goals and success criteria, which will most likely induce goal-oriented behaviors and reactions. In such a context, higher correlations between user performance and satisfaction can be expected. In a more experience-oriented test with open-ended and exploratory tasks, activity mode may be more prevalent, with lesser correlations between performance and satisfaction.

Within the domain of performance-oriented tests, methodological considerations can further explain diverging results. Sauro and Lewis (2009) stressed the importance of different data aggregation schemes (e.g., averaging over tasks vs. averaging over test participants), as well as the point in time when satisfaction is being measured. Satisfaction assessments conducted immediately after task performance typically show higher correlations with performance metrics than post-test satisfaction questionnaires where respondents integrate over a variety of factors that influenced their experience.

Another methodological aspect is the method used for calculating aggregated metrics. Because task completion times are not normal-distributed, it makes a great difference whether arithmetic means, medians, or geometric means are calculated. Using logarithmized task times (corresponding to geometric means when averaged), Xu and Mease (2009) found correlations with satisfaction ratings in the -.80 order of magnitude for web search tasks.

Xu and Mease's findings point at the importance of correctly dealing with the peculiarities of time distributions. The log transformation they used is a much recommended fix for the non-normal distributions typically found in task completion times (Sauro, 2011; Sauro & Lewis, 2012, p. 66f). Assuming that task completion times follow a lognormal distribution, logarithmizing data mathematically leads to a normal distribution that can be used in common statistical procedures. This effectively eliminates spurious error variance, created by the skewedness of the original distribution, that would account for lesser correlations and reduced power of statistical tests.

The lognormal distribution model, however, has certain conceptual shortcomings with regard to usability test data. First, the lognormal distribution starts at zero. But in usability test data, typically, there is a minimum time necessary to solve the task (for instance, because the system takes a finite time to render screens) and to respond to user input. Second, the parameters of the standard normal distribution (and by association, the lognormal) are conceptually misleading (Rummel, 2017). Naively, one would conceive of the mean as a midpoint and the standard deviation as a more or less symmetrical dispersion around it—as in the standard normal distribution. In the lognormal distribution, the situation is very different: Due to the nonlinear nature of the log transformation, standard deviations and confidence

intervals are not symmetrical nor are they as easily interpretable as in the normal distribution model. This holds in particular when an offset time is added to the model to account for a minimum solution time. The mean then is neither in the "middle" of the distribution nor can the standard deviation, as an interval, be interpreted independently from its location on the time scale, which is partly determined by the offset time.

More recently, Rummel (2017) proposed using the Weibull distribution model for modeling task completion times. Its three-parameter form covers a wide range of task completion time datasets from usability tests. In addition, its parameters can be related in a straightforward manner to the dynamics of the task solution process. The Weibull distribution's model equation is

$$\text{Equation 1:} \; S(t) = 1 - e^{-(\frac{t-t_0}{\tau})^{\gamma}}$$

where $S(t)$ denotes the percentage, S, of users still working on the task (the "survival" function) at a given time (t). The percentage of users declines from 1 (100%) over time and reaches either 0 when all users solved the task or the eventual failure rate in case some users were unable to complete the task. The model parameters t_0, τ and γ can be estimated from the observed completion times (Rummel, 2017); once the parameters are determined, the analyst can use Equation 1 to estimate the expected task completion rate for any given time, and vice versa. In addition, the parameters can be related to different parts of the task solution process as follows.

The t_0 parameter is a constant *offset time.* It can be attributed to all time contributions that are basically constant, that is, have negligible variance across all test participants, such as system response time and the time needed to merely operate the UI on the shortest path taken by test participants (typically, this is the ideal solution path).

The scale parameter τ, also called *characteristic time*, describes the stochastic part of the process, that is, the more or less random process of users dealing with the various challenges present in the task and the user interface. The stochastic aspect here is introduced by the fact that not all users meet the same obstacles and have stochastically varying resources (and sometimes, sheer luck) to deal with them. The percentage S of users still working declines over time in an essentially exponential manner (see below). Around time $\tau + t_0$, when the exponential term in the model equation reaches -1, it reaches S = 37% (= 1/e).

The Weibull's *shape parameter* γ finally describes deviations from the exponential distribution model. If γ is 1, the distribution equals the exponential distribution—the Weibull distribution in fact is a generalization of this distribution. The exponential distribution can be found in numerous natural processes that are based on purely random events such as radioactive decay, time between calls in call centers, and so on. It is fully determined by the characteristic time τ, which readers may be familiar with the conceptually similar term *half-life* in radioactive decay. This makes the deviation term γ interesting, as it indicates whether a process evolves, compared to a purely random one, in an accelerated or decelerated way. If γ is smaller than 1, the process is decelerated: In particular, slow users take longer to solve the task than we would expect in a purely random (i.e., exponential-distributed) process. If γ > 1, the process is faster; therefore, γ denotes a factor that systematically either accelerates or inhibits user performance.

Rummel (2017) described how to estimate the Weibull model parameters from empirical data. Once parameters are determined, the analyst can use Equation 1 to estimate the expected task completion rate for any given time, and vice versa.

The Weibull model thus decomposes task completion time distributions into components that correspond to different parts of the task solution process. Colloquially speaking, one can interpret t_0 as "click time," τ as "think time," and γ as "acceleration."

Naturally, the question arises how each component of the Weibull model contributes to overall user satisfaction. What influences user's experience more: operating a slow and "clicky" user interface (t_0), one that poses a multitude of random little challenges (τ), or one that systematically boosts or slows down their performance (γ)?

Method

Rummel (2017) demonstrated the applicability of Weibull modeling in a real-world industrial setting, using data from a series of summative, quantitative usability tests of business software applications. Out of this data set, tasks from nine tests were selected where user satisfaction ratings had been systematically collected in the same, standardized manner. Immediately after attempting each task, test participants were asked to "please rate your satisfaction with the user interface, as it supported you in the task you just performed" using a rating sheet with a 7-point rating scale (1 = very dissatisfied to 7 = very satisfied). For each task, ratings from all test participants were averaged to a task satisfaction score (aggregation scheme TM according to Sauro & Lewis, 2009). This task satisfaction score then could be related to completion time distribution parameters of the same task.

To ensure a sufficient number of data points for Weibull model estimation, only tasks with a completion rate > 50% were selected. From 73 tasks meeting this criterion, five were excluded where the Weibull model estimation procedure yielded a t_0 estimate of 0. For such cases, Rummel (2017) suggested that the t_0 estimate might not be actual click time (which realistically cannot be zero) but rather a distribution modeling artificiality. Eventually, 68 tasks remained for analysis, four of which had been run on a smartphone, all others on desktop systems. Participant numbers per task ranged from 14 to 18 with a median of 17. No individual participated in more than one of the nine tests, but within the same test, they attempted several (typically, all) tasks.

The Weibull modeling process for task completion times followed the procedure described in Rummel (2017), which is described only briefly here. Interested readers may want to refer to the original paper for details.
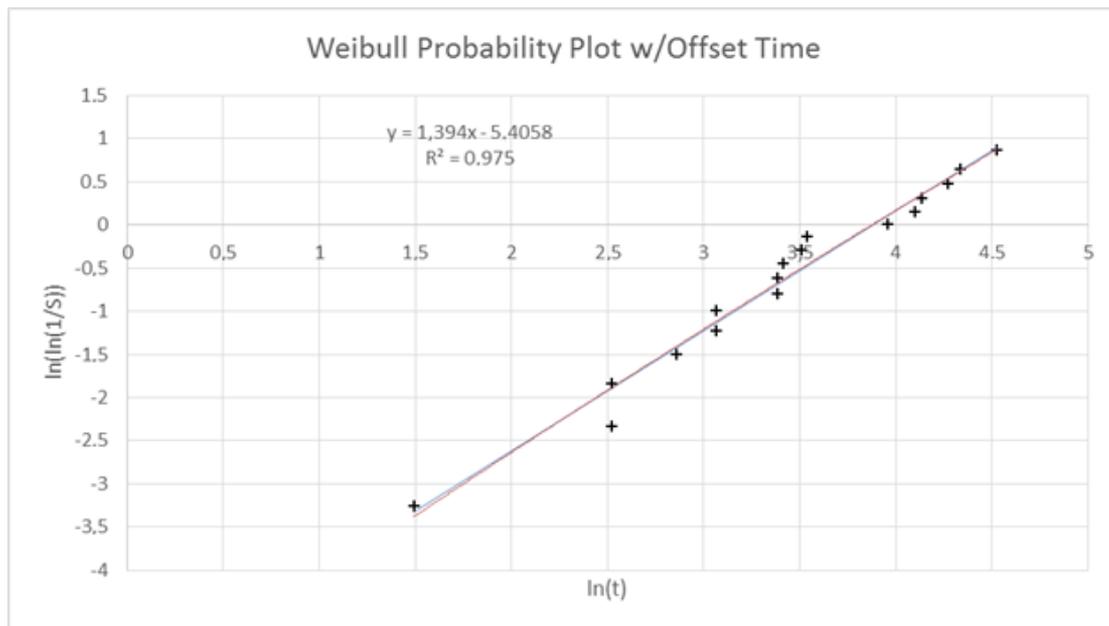

Figure 1. Example Weibull probability plot for a task in the present study, which 17/18 participants completed.

In the Figure 1 example, completion times for successful participants[1] are plotted logarithmically on the horizontal axis. The vertical axis shows the double logarithm of the survival function S.

---

[1] Times for unsuccessful participants are not plotted. The information of their failing the task is accounted for in the modified K-M estimate for the survival function S estimates for the successful participants.

The regression line represents the Weibull model of the task completion rate's progress over time. An overlaid quadratic regression line indicates that deviations from the linear model are unsystematic. The R² value of .975 indicates a good model fit for parameters $t_0 = 19.6s$, $\tau = 38.2s$, $\gamma = 1.394$. Note that τ is the exponential of the regression line's intersection with the time axis, γ its slope. The t_0 estimate is the value that maximizes R² when subtracting it from each individual task completion time.

Task completion times were plotted against corresponding survival function *S* estimates in probability plots (Rummel, 2014, 2017). Figure 1 shows an example plot with explanations. Distribution parameters were estimated from linear regression equations derived from these plots. Varying task completion rates were accounted for by treating task times from failed users as "censored." Rummel (2014) provided a detailed discussion how to deal with censored task times in usability tests. The mathematical treatment of these data is greatly simplified by assuming that those participants who gave up or came to wrong solutions would have taken at least as long to solve the task correctly as the slowest successful test participant (for a discussion of this rationale see Rummel, 2014). Under this assumption, deemed legitimate in the given context, survival functions were estimated using the modified Kaplan-Meier (K-M) Product Limit recommended by the National Institute of Standards and Technology (NIST; 2012; see also Tobias & Trindade, 2012, p. 202) for small samples that include censored times.

Results

All Weibull distribution parameters in the present data set were found to follow lognormal distributions. Consequently, all subsequent analyses were conducted using natural logarithms of parameters, which were normal distributed. In order to visualize the respective predictive value of Weibull distribution parameters and task completion rate, Figure 2 shows scatterplots of task satisfaction scores by those predictors, respectively. Linear regression using characteristic time τ (logarithmized) alone explains 66%, task completion rate 35%, (log) offset time t_0 33%, and (log) shape parameter γ 10% of satisfaction variance.

In order to further analyze the respective contributions of t_0, τ, γ, and task completion time to user satisfaction, intercorrelations between these metrics need to be considered in more detail. In linear regression analysis, the sequence at which regressors are entered into the prediction model is crucial: Because shared variance can only be used once for explanation, its relative importance will only be attributed to the regressor entered first in the model equation.

Intercorrelations between the metrics we consider in this paper can be expected to be substantial: A longer click path (affecting t_0) will offer more opportunities to make and correct random sidesteps and mistakes (affecting τ) that may cause user fatigue (affecting γ), dissatisfaction, and the likelihood of task fails.

A Principal Components Analysis (PCA) reveals that the task completion rate and (log) Weibull parameters metrics are indeed highly intercorrelated. There is a strong first, generic component explaining 61.7% of common variance; the next two components explain 17.6% and 11.2%, respectively. Figure 3 shows a two-dimensional plot of the metrics in the first two PCA dimensions.

In this plot, Satisfaction and $\ln(\tau)$ are almost perfectly collinear, loading strongly on the first component. The second component is characterized by the shape parameter γ. Task Completion Rate (TCR) and Offset Time t_0 also contribute variance to this component; however, they share most variance with the Satisfaction and $\ln(\tau)$ dimension.

With such highly intercorrelated metrics, determining each one's relative importance for predicting user satisfaction is not trivial. The statistical phenomenon of predictors "stealing" each other's explainable variance requires additional considerations in modeling and interpretation of results.
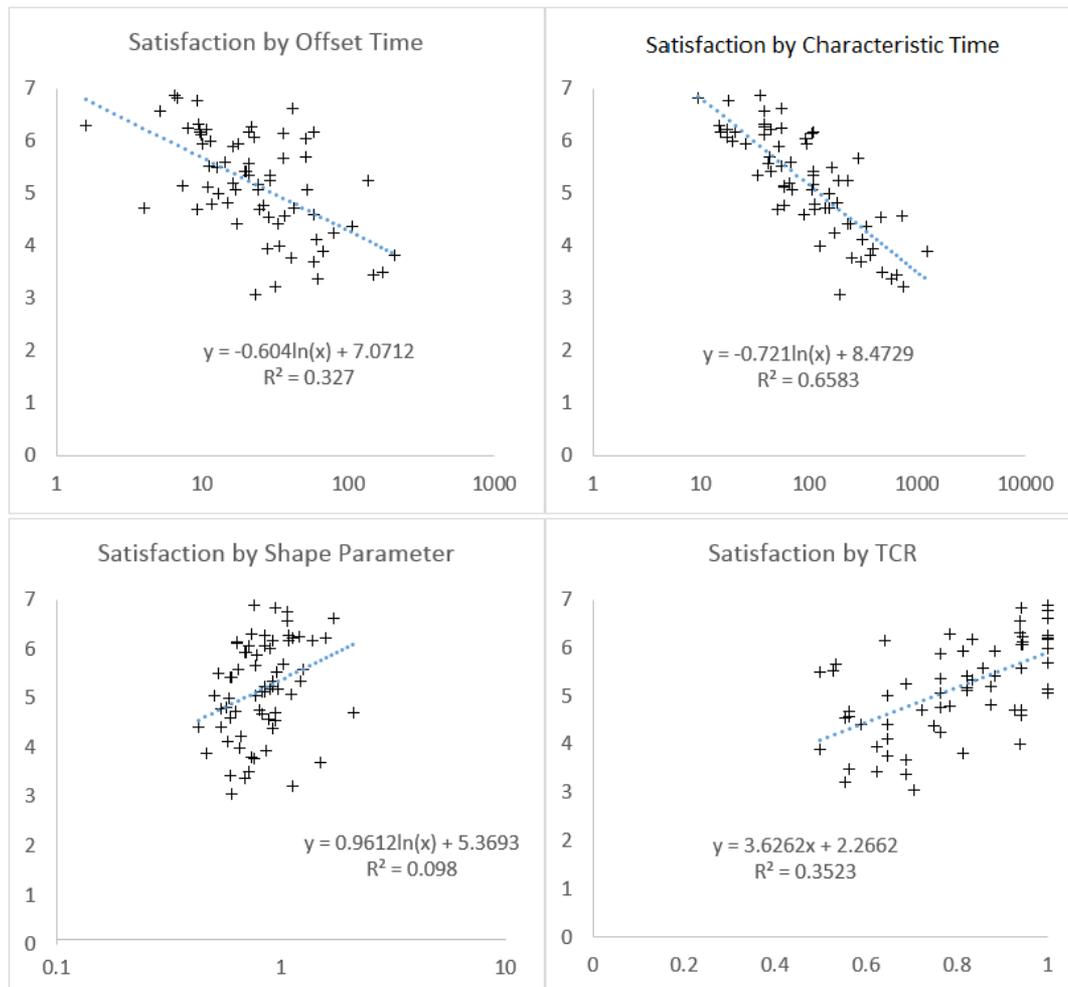
Figure 2. Scatterplots of task satisfaction scores by offset time t_0, characteristic time τ, shape parameter γ, and task completion rate (TCR). All times in seconds. Note log scales for all Weibull parameters.
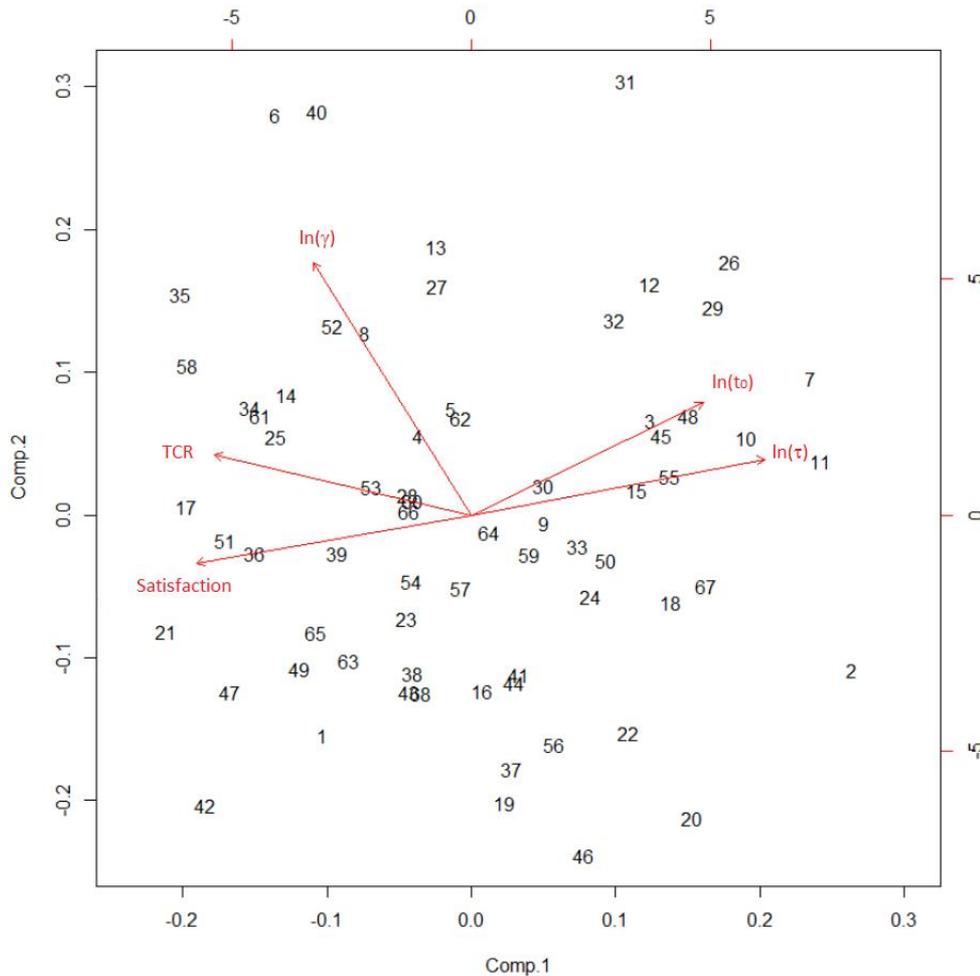
Figure 3. Principal Components Analysis result: bivariate plot of metrics in the first two dimensions resulting from PCA. Numbers indicate cases of tasks analyzed.

One possible approach to this problem is to decide on a theoretical basis which predictor to use first in the regression model. A valid argument can be made for starting with the offset time t_0: Technical response time and UI operation time on the ideal path are basically user-independent properties of an interactive system, so t_0 is "built into the UI" before a test participant even starts interacting with the system. Next would be τ and γ because they are attributes of the task solution process. Task completion rate would be last because task success or failure is the final result of the process. Table 1 shows the variance analysis for a linear regression model using this sequence.

In this model, $\ln(t_0)$ and $\ln(\tau)$ both explain about one third of satisfaction variance each. Interestingly, γ and task completion rate in this model contribute not only not significantly, but not at all.

A linear regression analysis using only $\ln(t_0)$ and $\ln(\tau)$, in this sequence, yields the equation

$$\text{Satisfaction} = 8.52 - 0.076 \ln(t_0) - 0.679 \ln(\tau)$$

If we choose to refrain from making a priori assumptions, Grömping (2006) recommended a different approach for estimating the relative importance of predictors, in particular "when the focus of the research is more on causal than on predictive importance" (p. 12). If the sequence

of predictors in the analysis matters, and we don't want to make assumptions, why not calculate linear regression models with all possible sequences of regressors and, basically, average the respective variance contributions. The procedure, which contains further corrections in order to normalize variance contributions so they sum up to 100%, is implemented in the *R* package *relaimpo* (Grömping, 2006). The metric named *lmg* corresponds to the percentage of variance explained by the respective regressor. Table 2 lists the results; Figure 4 shows a corresponding column chart. In this analysis, $\ln(\tau)$ comes out as the most important predictor, explaining 37% of variance. TCR and $\ln(t_0)$ both explain around 13% each, $\ln(\gamma)$ 3%. The overall model explains slightly more variance (66.8%) than the sequenced regression model described above (66.5%).

Table 1. Analysis of Variance for Sequenced Linear Regression Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Signif.	%	Cum.%
1. $\ln(t_0)$	1	207.381	207.381	619.662	5.98E-08	***	0.329	0.329
2. $\ln(\tau)$	1	211.950	211.950	633.312	4.22E-08	***	0.336	0.665
3. $\ln(\gamma)$	1	0.365	0.365	10.906	0.3003	n.s.	5.79E-07	0.665
4. TCR	1	0.0306	0.0306	0.0914	0.7634	n.s.	4.86E-08	0.665
Residuals	63	210.841	0.3347				0.335	1
---	Total	630.172					1	
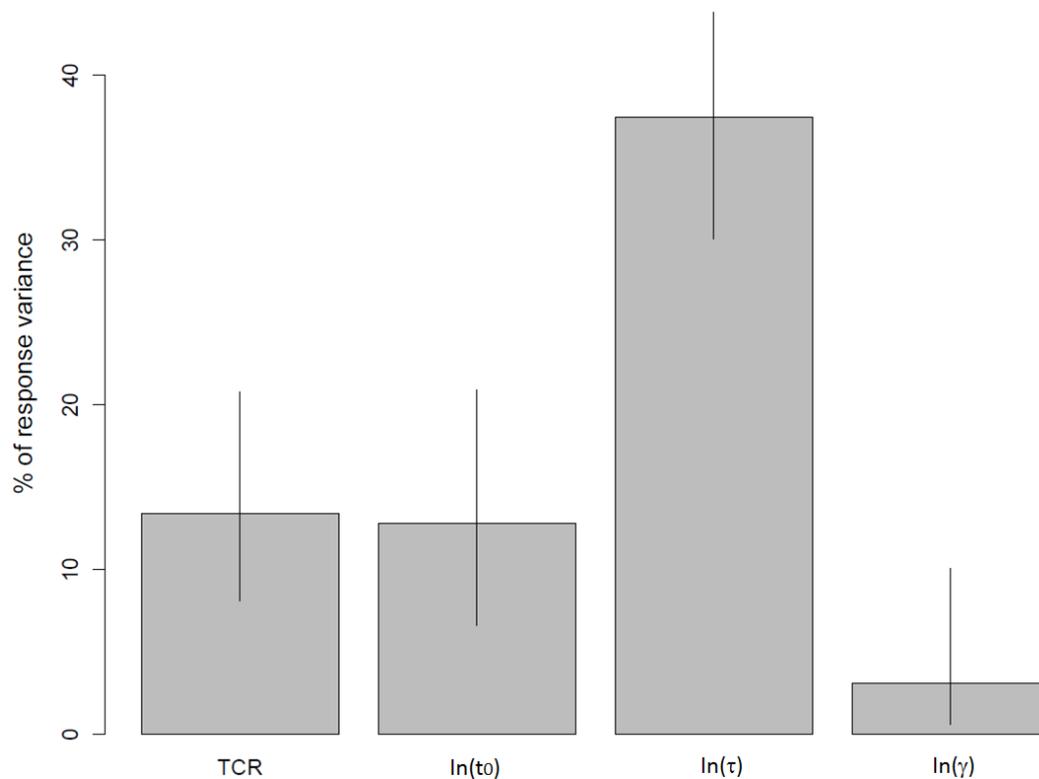
$R^2 = 66.75\%$, metrics are not normalized.

Figure 4. Relative importance for predicting satisfaction as assessed by metric *lmg*, *R* package *relaimpo* (Grömping, 2006). For numeric values see Table 2.

Table 2. Relative Importance for Predicting Satisfaction as Assessed by Metric *lmg*, *R* Package *relaimpo* (Grömping, 2006)

		90% Confidence Interval	
	lmg	**Lower Bound**	**Upper Bound**
$\ln(t_0)$	0.1280	0.0665	0.2089
$\ln(\tau)$	0.3742	0.3007	0.4381
$\ln(\gamma)$	0.0314	0.0063	0.1007
TCR	0.1339	0.0814	0.2077

Discussion

The findings of the present study underline the importance of pragmatic factors for post-task user satisfaction with business software. Time distribution parameters and task completion rate explain two-thirds of variance in post-task satisfaction ratings. Considering the obviously limited reliability of the one-item satisfaction rating instrument used here, this means there is not much explainable variance left. This finding corroborates the results reported by Xu and Mease (2009) and Strohmeier et al. (2013), as well as Hassenzahl's reasoning that pragmatic aspects are predominant for user satisfaction in quantitative tests on business applications (Hassenzahl,

2001)—in particular, if satisfaction ratings are collected right after task completion (Sauro & Lewis, 2009).

On the question, *which* pragmatic aspects are the most relevant, the detailed analysis of task performance parameters now sheds some light. It is not surprising that task completion is important for post-task user satisfaction, nor is it new that task completion time has an influence; in fact, the correlations found by Xu and Mease (2009) are in the same order of magnitude as the ones found here. The Weibull model however adds new means for understanding which components of the task solution process, as they affect completion time and as they are reflected in completion time distribution parameters, influence user satisfaction to exactly which extent. We now can investigate the effects of click time, think time, or acceleration on user satisfaction in detail, and separately.

This said, it is certainly surprising that the Weibull model's characteristic time τ is apparently more important than task completion rate, and time distribution parameters *alone* can predict user satisfaction at least to the same extent than when task completion is added to the picture. The quantitative assessment of this importance from the present data set needs to be taken with a grain of salt, as TCR and τ are correlated. In the absence of more detailed theoretical models on the processes how user satisfaction in business application is formed, it is difficult to determine which one is the truly leading parameter.

The importance of τ for user satisfaction, however, is actually quite plausible. As a crude model of thought, suppose each user randomly selected from a pool of usability issues present in the system, each with some cost in time and user satisfaction. Such a system and process setup would generate exponential-distributed task completion times, with τ directly reflecting the number and time costs of usability issues in the pool. Satisfaction costs would add up to a normal distribution, exactly as we typically see it in usability test data. The relationship would be explained *in toto* by the number of usability issues in the pool and the resulting likelihood of users "selecting" them.

For developing better causal models of user satisfaction, the offset time t_0 is also interesting. Compared to τ, the contribution of t_0 to the overall solution time is rather small (Rummel, 2014, 2017). Its contribution to user satisfaction however is substantial and in the order of magnitude of the task completion rate. Improving system performance and click count thus may have a small effect on actual efficiency but a substantial one on user satisfaction. It might very well be that users perceive time differently when waiting for system responses, when going through necessary operations, and when solving task and interaction problems. The former two are imposed by the UI, the latter involves their own activity. More experimental research is needed to investigate this further.

In this perspective, it is counter-intuitive that the Weibull shape parameter γ appears to be relatively unimportant for post-task user satisfaction. Small values of γ indicate that something slowed down the solution process systematically, beyond the random contribution of micro-usability issues. It is a bit surprising that the impact of such influence on satisfaction is rather small. Because the Principal Component Analysis reveals that γ is indeed a metric independent from others, further research on its practical importance, beyond its contribution to the numerical modeling of task completion rates over time (Rummel, 2017, see also Equation 1), might provide interesting insights.

Conclusions

Weibull distribution model parameters of task completion times have high predictive value for post-task user satisfaction, at least in the order of magnitude of the task completion rate. When characteristic time τ and the offset time t_0 are considered, the task completion rate does not add further predictive value. For the domain of business applications and task-based usability tests, this underlines the importance of pragmatic quality (Hassenzahl, 2001) for user satisfaction, with a particular emphasis on efficiency.

The amount of variance explained by Weibull model parameters τ and t_0 establishes them as key drivers for post-task user satisfaction. Further investigations into the detailed mechanisms, how exactly users' experiences in the time domain affect their satisfaction, therefore appear promising.

As discussed initially, these findings so far are restricted to the domain of task-focused business software applications, where task instructions put test participants clearly into a goal-oriented mode according to the taxonomy by Hassenzahl et al. (2002). This said, for this UI genre, they provide an interesting pathway for better understanding and improving user satisfaction. For other similar genres, such as web shops, where behavior tracking data may be more easily available than user satisfaction ratings, they may offer new pathways for analyzing and predicting user experiences.

Tips for Usability Practitioners

For Weibull-modeling task completion times, Rummel (2017) provided a detailed introduction and a calculation spreadsheet.

The relationship between Weibull parameters and user satisfaction is strong but not linear—in fact, the relationship is logarithmic. The curve has a steep decline at the beginning and a shallow tail: Long task durations hurt satisfaction, and eventually satisfaction hits something like a floor. For usability practitioners, this leads to a simple rule of thumb:

- Make core tasks fast to complete! If click time t_0 is greater than 1 minute, or if think time τ is greater than 5 minutes, good satisfaction ratings become very unlikely.
- Click time t_0 can be approximated by the minimum observed time (Rummel, 2017; Tobias & Trindade, 2012) or pragmatically estimated by having someone click through the task on the ideal path. The latter you can do even before a usability test.
- Characteristic (think) time τ can be understood as the "typical" time a real user would take. It is in the order of magnitude of the time when 50% of users solve the task. So, if half your test participants take longer than 5 minutes or fail the task, watch out.
- Saving click time t_0 is good, but mind that think time τ is twice as important for satisfaction. More but simpler screens are often the better solution.

References

Coursaris, C., & Kim, D. (2011). A meta-analytical review of empirical mobile usability. *Journal of Usability Studies, 6*(3), 117–171.

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software, 17(1).* doi: 10.18637/jss.v017.i01

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? *Proceedings of the SIGCHI conference on Human Factors in Computing Systems, CHI 2000* (pp. 345–352). New York, NY: ACM Press.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction 13*, 481–499.

Hassenzahl, M., Kekez, R., & Burmester, M. (2002). The importance of a software's pragmatic quality depends on usage modes. In H. Luczak, A. E. Cakir, & G. Cakir (Eds.), *Proceedings of the 6th international conference on Work With Display Units* (WWDU 2002; pp. 275–276). Berlin: ERGONOMIC Institut für Arbeits- und Sozialforschung.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*(2), 79–102. doi: 10.1016/j.ijhcs.2005.06.002

Hornbæk, K., & Law, E. L. (2007). Meta-analysis of correlations among usability measures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007* (pp. 617–626). New York, NY: ACM Press.

Molich, R., Chattratichart, J., Hinkle, V., Jensen, J. J., Kirakowski, J., Sauro, J., Sharon, T., Traynor, B. (2010). Rent a car in just 0, 60, 240 or 1,217 seconds?—Comparative usability measurement, CUE-8. *Journal of Usability Studies, 6*(1), 8–24.

NIST/SEMATECH (2012). Empirical model fitting—Distribution free (Kaplan-Meier) approach. In E-handbook of statistical methods. *National Institute of Standards and Technology*.

Retrieved August 02, 2016 from
http://www.itl.nist.gov/div898/handbook/apr/section2/apr215.htm#Modified K – M.

Rummel, B. (2014). Probability plotting: A tool for analyzing task completion times. *Journal of Usability Studies 9*(4)*,* 152–172*.*

Rummel, B. (2017). Beyond average: Weibull analysis of task completion times. *Journal of Usability Studies 12*(2), 56–72.

Sauro, J. (2011). 10 things to know about task times*. Measuring Usability*. Retrieved August 2016 from http://www.measuringusability.com/blog/task-times.php

Sauro, J., & Lewis, J. R. (2009). Correlations among prototypical usability metrics: Evidence for the construct of usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009* (pp. 1609–1618). New York, NY: ACM Press.

Sauro, J., & Lewis, J. R. (2012). *Quantifying the user experience.* Waltham, MA: Morgan Kaufmann.

Strohmeier, D., Mikkola, M., & Raake, A. (2013). The importance of task completion times for modeling Web-QoE of consecutive web page requests. *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE

Tobias, P. A., & Trindade, D. C. (2012). *Applied reliability* (3rd ed.). Boca Raton, FL: CRC Press.

Xu, Y., & Mease, D. (2009). *Evaluating web search using task completion time.* Retrieved August 02, 2016 from http://static.googleusercontent.com/media/research.google.com/de//archive/dmease-sigir09-full.pdf

About the Author

Bernard Rummel
Trained in experimental psychology, Mr. Rummel has been working in the Usability and UI Design field for over 20 years. After nine years at the German Naval Medical Institute, he joined SAP in 2000, where he is currently responsible for usability testing methodology and participating in the national standardization body DIN.