Analyzing cognitive processes in CATI-Surveys with response latencies: an empirical evaluation of the consequences of using different baseline speed measures

Mayerl, Jochen; Sellke, Piet; Urban, Dieter

SISS:

Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart

No. 2 / 2005

Analyzing cognitive processes in CATI-Surveys with response latencies: An empirical evaluation of the consequences of using different baseline speed measures

Jochen Mayerl

Piet Sellke

Dieter Urban

Universität Stuttgart
Abteilung für Sozialwissenschaften IV

70174 Stuttgart

SOWI

ISSN 0945-9197

SISS:
Schriftenreihe
des Instituts für Sozialwissenschaften
der Universität Stuttgart: No. 2 / 2005

Analyzing cognitive processes in CATI-Surveys with response latencies: An empirical evaluation of the consequences of using different baseline speed measures

Jochen Mayerl

Piet Sellke

Dieter Urban

Abteilung für Sozialwissenschaften IV
Universität Stuttgart

70174 Stuttgart

A B S T R A C T: The study investigates the use of response latencies as a measure of attitude strength in survey research. It examines various possibilities of measuring personal reaction times in CATI-surveys and correcting these measurements for personal baseline speed. It also elucidates the empirical consequences of using different baseline speed measures. The study argues that a specific transformation index newly developed by the authors, the so-called ´Residual-Score-Index”, offers a better procedure for controlling baseline speed when measuring response latencies than the traditional transformation indices (Difference-Score-Index, Ratio-Score-Index and Z-Score Index). The advantages of the new Residual-Score-Index are shown in a statistical analysis testing its moderational power in a multiple regression estimation.

Z U S A M M E N F A S S U N G: Die Studie untersucht den Einsatz von Antwortlatenzen zur Messung der Stärke von Einstellungen im Bereich der Surveyforschung. Sie beschreibt verschiedene Möglichkeiten zur Messung von individuellen Reaktionszeiten in CATI-Surveys und zur Bereinigung dieser Messungen von Einflüssen, die von personalen, frageunabhängigen Antwortgeschwindigkeiten („individuelle Basisgeschwindigkeiten“) ausgehen. Dabei verdeutlicht sie auch die unterschiedlichen empirischen Konsequenzen, die mit der Nutzung verschiedener Maße von Basisgeschwindigkeit verbunden sind. Die Studie stellt zudem einen von den Autoren neu entwickelten Transformationsindex, den so genannten „Residual-Score-Index“, vor, der ein besseres Verfahren zur Kontrolle von Effekten frageunabhängiger Antwortgeschwindigkeiten bei der Messung von Antwortlatenzen ermöglicht als die Verfahren der klassischen Transformationsindizes (Difference-Score-Index, Ratio-Score-Index, Z-Score Index). Die Vorteile des neuen Residual-Score-Indexes werden in einer statistischen Analyse gezeigt, bei der die Moderatorfunktion des Indexes im Kontext einer multiplen Regressionsschätzung überprüft wird.

1 Introduction[1]

The measurement of reaction times in CATI-surveys (also known as “response latencies”) is supposed to reveal cognitive processes of individuals while they are responding to survey questions. Reaction time however, is a multidimensional concept and is influenced by a number of effects such as attitude accessibility, mode of processing, effects of the measurement instrument (for example the wording of questions) and also by individual baseline speed. Individual baseline speed can be understood as a respondent’s general speed of answering survey questions *independent* of the question content.

In this paper we argue that without controlling for baseline speed no adequate interpretation of reaction time is possible. However, different *concepts* of baseline speed are available and when applied, one can observe different effects on the measurement of baseline speed. For instance, we will show that influences of a certain educational background will only bias the measurement of baseline speed when reaction time is measured at ´difficult items´ (i.e. ratings of abstract concepts like ´need for cognition´). Furthermore, we show that depending on the concept of baseline speed applied, one can expect differences in the moderational power of reaction time in regard to the attitude–behavior relationship.

To date the literature on reaction times shows no clear distinction between different concepts of measuring baseline speed but rather seems to use various concepts interchangeably or to apply no explicit concept at all. In this paper we will discuss various traditional and one newly developed concept of response time measurement. Our starting point is to provide an analytical distinction between different concepts of baseline speed and their operationalization (2). Next we present some common transformation indices used to compute the baseline speed from reaction time (3). To test our hypotheses, we present an analysis of a nation-wide representative CATI-survey (N=2002) employing different transformation indices. We then test the moderational power of reaction time as an interaction variable regarding the explanatory power towards the attitude–behavioral intention relationship (4). Finally, we summarize our findings and make some suggestions for applying response latency measurements in actual survey research (5).

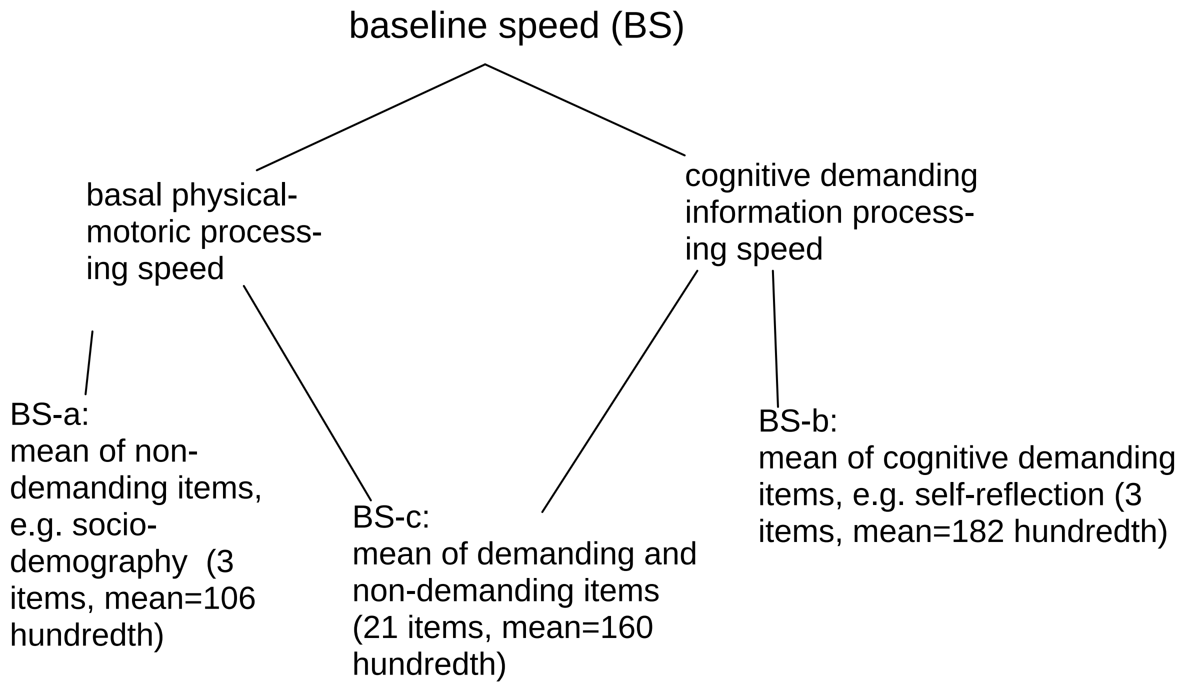
[1] This research was funded by a grant from the Deutsche Forschungsgemeinschaft (DFG), the German national science foundation.

2 Baseline Speed

The time it takes a respondent to answer a survey question is – besides influences from attitude accessibility and mode of processing – dependent on his/her individual baseline speed. Baseline speed can be defined as the general speed of responding, regardless of question content. When measuring response latency, the necessity to control for baseline speed is well reported in literature. Fazio (1990) for instance reports about intra-individual covariances between the reaction times a respondent needs to answer different questions in a single research interview, which suggests that respondents are generally slower or faster when answering different questions in the same survey. Thus, the theoretical argument for controlling baseline speed is simple: there will always be differences in the response times of different respondents with some relatively slower or faster than others. Therefore, controlling for baseline speed means that the reaction time of every respondent is analyzed in comparison to his/her own individual standard, i.e. his/her individual baseline speed. Thus, it is necessary to control for baseline speed in order to achieve a solid interpretation of reaction times.

In the literature, however, many different concepts of baseline speed are applied or even no concept at all is used to correct reaction times. Fazio (1990) suggests using the mean of so-called filler latencies (reaction times to items that are neither substantially nor theoretically connected to the target latency) to calculate baseline speed. Furthermore, he suggests that questions at the beginning of the questionnaire used to introduce the respondent to different answering scales should not be used in calculating baseline speed because of learning effects on the respondent's side (Fazio 1990; Shrum/O'Guinn 1993, however: Hertel/Neuhof et al. 2000 did exactly this, and Bassili 2003 even proposed to compute baseline speed as the mean of all item-specific reaction times measured in a particular survey). A second consideration in computing baseline speed is the level of complexity of the items used. One possibility is to compute the baseline speed with simple factual questions concerning issues such as age or dietary preferences, and the idea is that this will measure the pure mental speed the respondent needs to process information (Bassili 1993; Johnson et al. 2002; Shrum/O'Guinn 1993). Another possibility to compute the baseline speed is to use more complex questions (e.g. questions concerning social judgments, values, attitudes) to measure effects on response time stemming from educational or psychological influences (i.e. ´need for cognition´). A third possibility to compute baseline speed is to use reaction times from both easy and difficult questions. The following figure 1 shows these variations.

Figure 1: Concepts of baseline speed (BS)



In figure 1, simple factual questions are labeled "non-demanding" items. In contrast, demanding items represent attitudinal and psychological questions and judgments. The figure also reports results of different models of BS-operationalization we applied in our own empirical study (see below).[2] In figure 1 and in our analysis, we excluded the specific form of baseline speed that is constructed with *all* items' reaction times. Because, if one wants to assess item-specific reaction times in separate analyses (like we want to do) it would be illogical to rely on these specific reaction times when computing a general baseline speed on reaction times. This would result in tautological computations.

In addition to differences in the items that can be included in computing baseline speed, there are also different transformation indices that can be applied to correct raw reaction time regarding baseline speed. These will be presented in the next section.

[2] The following items were used to measure baseline speed: BS-a) vegetarianism, age, confession; BS-b) self-efficacy, social desirability, tendency to use heuristic information processing (one item each); BS-c) items of BS-a, BS-b, religiousness, church attendance and additional items of self-efficacy, social desirability, and tendency to use heuristics.

3 Transformation Indices

The objective of all transformation indices is to transform reaction time so that there is no covariance between baseline speed and the transformed reaction time. Fazio (1990) suggests three different transformation indices to control for baseline speed: Difference Score, Z-Score, and Ratio-Score. We offer an additional index to these three: the Residual-Score Index. The following figure shows the calculation of the four indices.

Figure 2: Transformation indices

Difference-Score-Index:	$RT_{trans} = RT_{raw} - BS$
Ratio-Score-Index:	$RT_{trans} = RT_{raw} / (RT_{raw} + BS)$
Z-Score-Index:	$RT_{trans} = (RT_{raw} - BS) / stddev(BS)$
Residual-Score-Index:	$RT_{trans} = RT_{raw} - (a + b\ BS)$

with RT=reaction time; BS=baseline speed

The logic of the Difference-Score is pretty simple and is based on the notion of reaction time as an additive model[3]. Thus, the unwanted effect – the baseline speed – is simply subtracted from the raw reaction time. This score has been applied successfully in different studies (e.g. Bassili 1993; Fabrigar et al. 1998; Fazio et al. 1984; Fazio/Dunton 1997; Kokkinaki/Lunt 1999; Mayerl 2003; Sellke/Mayerl 2004; Stocké 2001, 2002c). The Difference-Score is scaled in hundredth. The Ratio-Score transforms reaction time to a scale from 0 to 1, with 0.5 indicating that target latency and baseline speed is equal to each other. Values above 0.5 indicate that the reaction time is slower than the baseline speed, values below 0.5 indicate the opposite. The important difference between the Ratio-Score and the Difference-Score is that in the case of the latter, two respondents with different baseline speeds can receive the same Difference-Score (e.g. 400-200=200 vs. 1000-800=200). However, the differences between these two respondents remain when using the Ratio-Score (e.g. 400/(400+200)=0.66 vs. 1000/(1000+800)=0.55).

In contrast to Difference- and Ratio-Score, the Z-Score also considers the standard deviation of the baseline speed. It has also been applied in several empirical studies (Fazio 1990; Fazio / Powell / Williams 1989; Fazio/Powell 1997; Mayerl 2003; Sellke/Mayerl 2004).

[3] Reaction time = (attitude accessibility) + (mode of processing) + (baseline speed) + (effects of measurement instrument) + (other biasing effects).

The Residual-Score is presented in this paper for the first time. The logic of this formula is based on procedures of regression analysis. If the raw reaction time is regressed on the baseline speed, the residuals will indicate the proportion of time that is independent of baseline speed. In commonly used software packages for statistical analysis (e.g. SPSS) the residuals of a regression estimation can be saved to a new variable and thus used for further analysis.

Figure 3: Logic of Residual-Score-Index

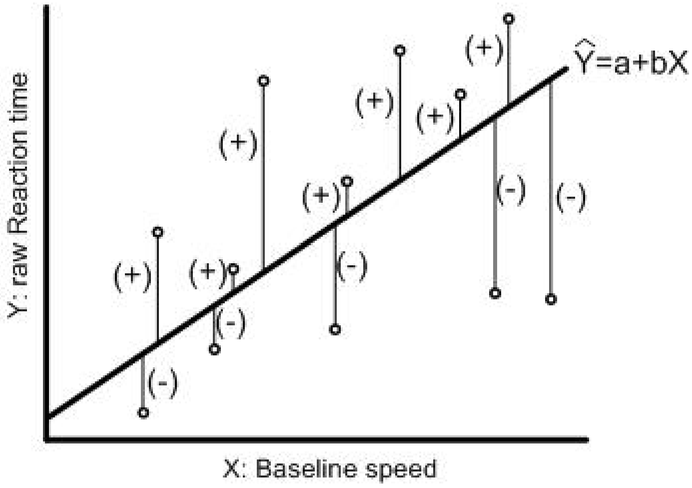


Figure 3 illustrates the logic of the Residual-Score. Every respondent is represented with his/her individual raw reaction time and his/her baseline speed by a single dot in the scatterplot. Both measures are needed for computing the Residual-Score. This is similar to the other transformation indices which also rely on individual baseline speed and raw reaction time. However, in addition to these two time measures another information can be detected in figure 3: a regression line. The regression line indicates a linear relationship between raw reaction time and baseline speed. When this linear function is estimated by statistical methods one can compute a reaction time for each respondent that is exclusively determined by his/her baseline speed (and, when graphically visualized, is located on the regression line). Based on this estimation, for each respondent the part of reaction time can be computed that is independent of his/her baseline speed. This we call "Residual-Score". The Residual-Score is the difference between the empirically observed reaction time (RT_{raw}) and the statistically expected reaction time (RT_{exp}) which is estimated as a linear function of baseline speed (BS). For each respondent then, there is the following Residual-Score:

$$\text{Residual-Score} = RT_{raw} - RT_{exp}$$
$$= RT_{raw} - (a + b\ BS)$$

The interpretation of the Residual-Score is straightforward. Positive residuals indicate that a respondent has a longer reaction time than predicted by the baseline speed. The higher the residual value, the slower the response compared to the predicted raw reaction time. Negative residuals, in turn, indicate a faster response than the predicted value. Smaller numbers, therefore, mean a faster response (e.g. interpreted theoretically as higher attitude accessibility). The difference between the Residual-Score and the other transformation indices is its dependency on the results of a regression estimation. Adding a regression estimation provides two major advantages: it allows to compute a personal reaction time that is relative to a group specific speed of all respondents in the sample; and it has the advantage of correcting the statistical relationship between raw reaction time and baseline speed almost completely. One disadvantage, however, is that the Residual-Score cannot be used to compare mean Residual-Scores when the reaction times concerning two or more questions are investigated. A comparison *between item-specific reaction times*, e.g. whether one attitude is more accessible than another, is not possible, because their mean values are always zero.

In the next section, we use survey data to compare the different baseline speed measures with each other and then use the different transformation reaction time indices to explore the attitude-behavioral intention relationship.

4 Empirical Analysis

The data we use for this analysis comes from a German nation-wide representative CATI-survey (N= 2002). In this study, active and passive reaction times were measured during the process of interviewing. Passive reaction times were measured by initiating the measurement automatically with the beginning of reading the question to the respondent. The interviewer stopped the measurement with typing in the appropriate answer. In contrast, active measurement of reaction times started after reading the question to the respondent. After having read the question, the interviewer immediately triggered the measurement by pressing a key and stopped it as soon as the respondent answered substantially. Thus, passive measurement includes the time it takes to read the question and the answering scale to the respondent, while active reaction times include only the time the respondent needs to provide an answer. In addition to these time measurements, the survey we conducted included questions about attitudes and behavioral intentions towards health conscious nutrition.

In the next sections of this paper we will present a statistical analysis of the measured reaction times. The analysis follows three steps. First, the validity of the measurement of baseline speed is evaluated. Second, we calculate the covariation of baseline speed and raw response latencies as well as its correction by transformation indices. Third, we test reaction times as a moderator of the predictive power of attitudes, controlled for different baseline speed measures.

4.1 Validity of the measurement of baseline speed

Baseline speed should be dependent on individual factors such as age, sex and education. Computing a regression with baseline speed as dependent variable and individual factors as independent variables can be seen as a test for criterion validity. First, however, we want to present some descriptive statistics regarding the different baseline speed measures.

Table 1: Descriptive statistics for different measures of baseline speed (corrected for outliers)[4]

Statistic	BS-a	BS-b	BS-c
sample size (N)	1849	1826	1326
mean	105.5	182.0	159,8
stddev	39	77.8	51.5
median	100.7	168.3	156.3
kurtosis	0.4	0.1	-0.1
skewness	0.7	0.7	0.3

It can be seen that all three baseline speeds have acceptable values of kurtosis and skewness. Furthermore, the mean time is, as expected, fastest with BS-a, slowest with BS-b, and with the mean of BS-c in between.

Applying a principal component analysis reveals that BS-a and BS-b have factor loadings of around 0.6 and load on a single factor. The results for BS-c are different. Because BS-c was constructed with 21 items and based on items of BS-a and BS-b, it is loading on three factors. However, because no target latency is included in BS-c, this result is still acceptable.

[4] Only data corrected for outliers is presented at this point. This is because with interviewer validation BS-c would drop down to approximately N=500 and lower, i.e. a reduction of about 75% of the original sample size. This is considered as unacceptable. The reaction times used to compute the baseline speed were corrected for outliers larger than two standard deviations above the mean, below respectively.

Next we present the results of a regression analysis to assess the criterion validity of our baseline speed measures. The regression estimates are shown in Table 2.

BS-a, the baseline speed with non-demanding items, is highly significant depending on age and sex. Further, one can find 22 significant interviewer effects (out of 62 interviewers). Most importantly, there are no effects of education, need for cognition, or speed/accuracy instructions, respectively. This result can be interpreted that in fact, with BS-a, basal physical-motoric processes are measured as opposed to BS-b and BS-c. In the case of the latter two we can see significant effects of education, speed/accuracy-instructions and of need-for-cognition dispositions. Additionally, only 13 interviewers have a significant effect on the baseline speed b and c, as opposed to 22 interviewers in BS-a. Thus, obviously a baseline speed measure computed with cognitive non-demanding items is more biased by the interviewers and more dependent on factors like sex and age. This regression analysis shows that all three baseline speed indices are valid measures. They all are dependent on predictors according to the logic of their construction. For example, BS-b and BS-c (concerning reaction times of ´difficult questions´) are influenced by a high degree of education, whereas BS-a (concerning reaction times of ´easy questions´) is influenced by sex but not by level of education.

Table 2: Results of a regression analysis with baseline speed as dependent variable

X-variables	**Y: baseline speed (corrected for outliers)**		
	BS-a	**BS-b**	**BS-c**
age	0.195/0.078**	0.653/0.133**	0.413/0.125**
high education	n.s.	10.157/0.061*	7.360/0.067*
low education	n.s.	n.s.	n.s.
sex	-8.048/-0.102**	n.s.	n.s.
interview instruction: speed & accuracy	n.s.	-18.854/-0.118**	-21.102/-0.201**
interview instruction: speed	n.s.	-24.263/-0.106**	-21.943/-0.143**
interview instruction: accuracy	n.s.	n.s.	n.s.
interviewer (dummies)	22 significant interviewer effects*	13 significant intervie-wer effects*	13 significant interviewer effects*
need for cognition	n.s.	4.194/0.057*	n.s.
R^2	*0.223*	*0.146*	*0.257*

unstandardized/standardized regression coefficients; *p<=0.05; **p<=0.01; n.s.: p>0.05

The next step in this empirical analysis is to assess the covariation of raw reaction times and baseline speed and its reduction by different transformation indices.

4.2 Correlation of raw reaction times with baseline speed and its reduction by transformation indices

Correcting empirical measurements of response time for individual baseline speed aims at 1) controlling time measurements for individual biases; and 2) making it possible to compare reaction times of different respondents in the same sample. In the respective procedure, we have to check for the linear correlations between raw reaction time and baseline speed. If there is no significant correlation, there is no need (and no possibility) to control for baseline speed.

Before computing a linear correlation, the assumption of linear relationship between baseline speed and raw response latency should be tested statistically. The following figure 4 shows three scatterplots representing the empirical data for computing correlation measures (each for every concept of baseline speed).

In all three scatterplots (with BS-a, BS-b and BS-c), the linear regression line departs from the lowess line, however the differences between the two lines is rather small. In the centre of each scatterplot, the lowess line and the linear regression line are very close. Some tendencies towards nonlinearity can be detected at the lower and upper ends of the distributions where the differences between the two lines increase. But these differences are rather small when we look at their absolute values considering the range of our measurements. Thus, in our view the relationship between baseline speed and raw response latency can be regarded as approximately linear. These findings may legitimize the use of linear correlation measures and linear transformation procedures to compute the respective indices (the same results are found when analysing the distribution of regression residuals).

Figure 4: Scatterplots testing the assumption of linear dependency

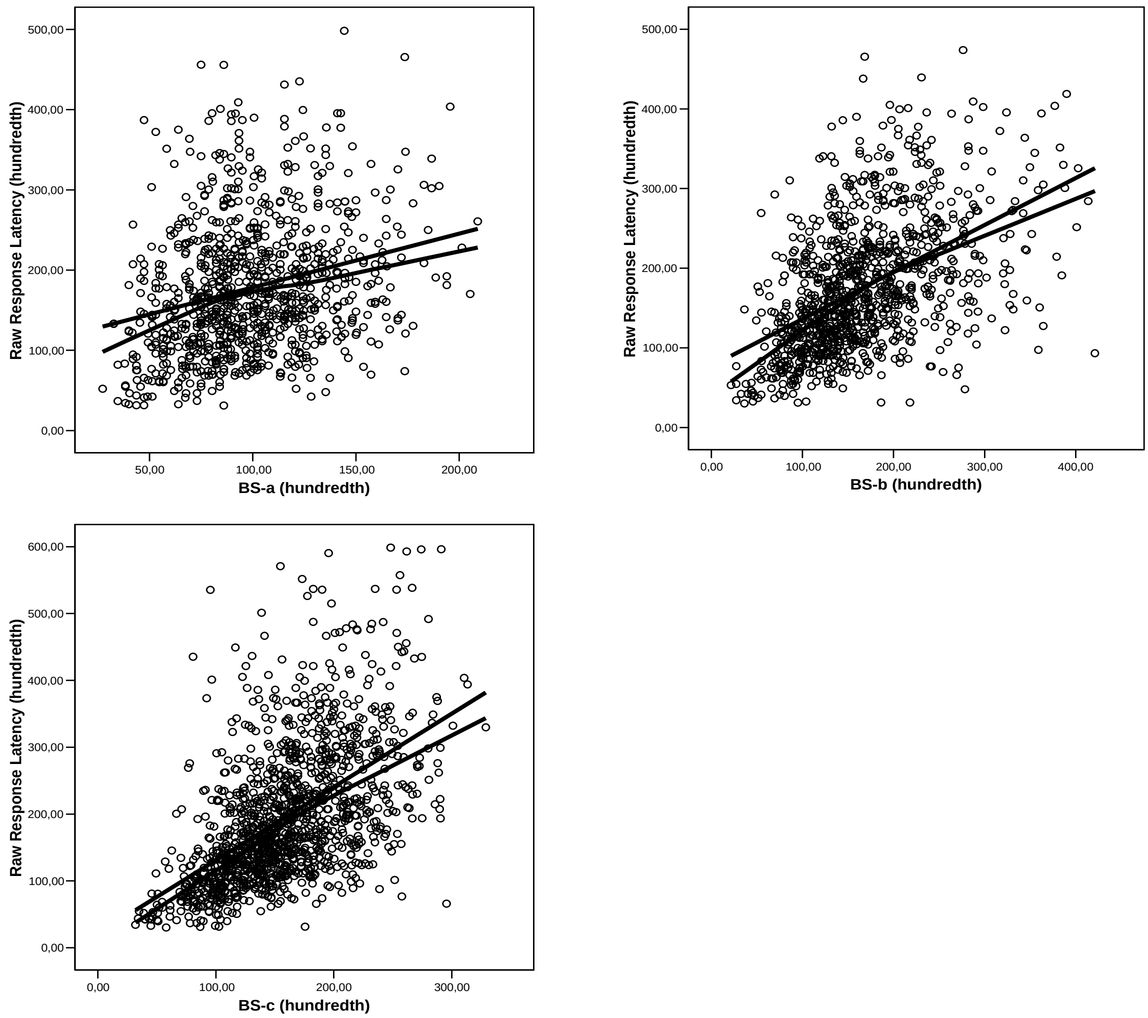


Table 3 presents the correlation values between raw reaction time and the different measures of baseline speed (see the first line in table 3). It also shows the modifications of these correlations when transformation indices are used instead of the raw measure of reaction time (raw RT).

Table 3: Correlations of measures of reaction time with three measures of baseline speed

		BS-a corrected for outliers	BS-b corrected for outliers	BS-c corrected for outliers
raw-RT	Pearsons r	0.245	0.466	0.544
	N	1631	1760	1297
Z-Score	Pearsons r	-0.311	-0.248	-0.123
	N	1625	1640	1242
Difference-Score	Pearsons r	-0.156	-0.345	-0.013 n.s.
	N	1631	1644	1242
Ratio-Index	Pearsons r	-0.408	-0.395	-0.108
	N	1631	1644	1242
Residual-Index	Pearsons r	-0.012 n.s.	-0.020 n.s.	-0.028 n.s.
	N	1631	1644	1242

all correlations $p<=0.01$; n.s.: $p>0.05$

The results of the correlation between raw reaction times and the baseline speed measures indicate a clear need to control for baseline speed. The baseline measure BS-a goes with a significant correlation of 0.245, and BS-b and BS-c correlate with raw-RT at even higher values (0.466, 0.544 respectively). However, looking at the modification of these correlations when using transformation indices instead of raw-RT, some results are rather surprising. In the case of BS-a, the Z-Score-Index has an even higher (negative) correlation, the same for the Ratio-Index. For BS-b and BS–c the modification results in a reduction (more or less well) with Z-Score-Index, Difference-Score-Index, and Ratio-Score-Index. The best reduction is achieved with the Residual-Index because for all three baseline speeds there are no significant effects left. This is not surprising, however, due to the computational logic of the Residual-Index (see section 3) that automatically leads to a non-significant correlation between the residuals and the independent variable (baseline speed) of the respective regression estimation. The only transformation index that also shows a non-significant correlation with baseline speed is the Difference-Score using BS-c for measuring baseline speed.

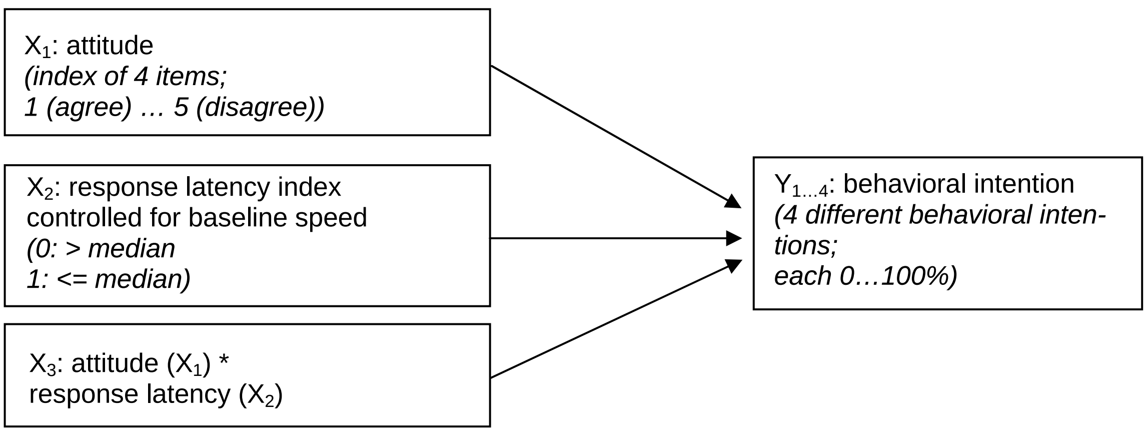
In the next section, we examine the moderational power of response latencies controlled for various baseline speed measures when used as additional predictors in a statistical model estimating the effects of attitude measures on various behavioral intentions.

4.3 Test of moderational power of reaction times controlled for different baseline speed measures

The final step in our analysis is a test of the moderational power of the Residual-Score-Index as a specific computation of reaction time relying on different baseline speed measures. This will be investigated in a statistical model estimating attitude effects on behavioral intentions. It is expected that the interaction effects between response latencies and attitude measures will be significant with transformed reaction times and will not be significant with raw reaction times. We examine the differences by using different baseline speed measures.

The design of the analysis is as follows. We specified an additive linear model directing the effects of three independet variables on one dependent variable. In a first analysis, the dependent variable was constructed as an index of three measures concerning three different behavioral intentions. In a second analysis, each measure of a specific behavioral intention served as a dependent variable on its own.[5] The independent variables consisted (1) of an index summarizing the evaluations of 4 attitude items (for example, an item concerning the importance of health conscious nutrition), (2) a binary measure of the raw BS and the Residual-Score of BS (based on the three different BS-measurements and computed by a median split into slow versus fast respondents), and (3) the interaction effect between attitude index and time measurement.[6] The following figure 4 gives an overview of the respective analytical model.

Figure 4: Model for testing the moderational power of response latency computed with different baseline speed measures



[5] See appendix for item wording.
[6] Due to multicollinearity, the independent variables were centred (results are replicated not centering X_2).

Table 4: Testing interaction effects with different baseline speed measures

	corrected for outliers								corrected for outliers and interviewer-validated						
	‚Raw'		Residual-Score with BS-a		Residual-Score with BS-b		Residual-Score with BS-c		'Raw'		Residual-Score with BS-a		Residual-Score with BS-b		*BS-c*
X_i	B	Beta	B	Beta	B	Beta	B	Beta	B	Beta	B	Beta	B	Beta	
dependent variable: behavioral intention (index)															
response latency	2.592	.057 *	2.628	.059 *	2.146	.048 *	10.704	.240 **	2.319	.051 ns	2.742	.062 $^{+}$	.376	.008 ns	*not estimated due to small sample size*
attitude (t_1)	-7.074	-.250 **	-7.016	-.247**	-7.035	-.248 **	-5.161	-.179 **	-7.693	-.279 **	-7.328	-.266 **	-7.547	-.272 **	
interaction effect	**-2.718**	**-.048 ***	-2.563	-.045 $^{+}$	-2.380	-.042 $^{+}$	**-3.484**	**-.196 ***	-2.192	-.039 ns	-3.009	-.055 $^{+}$	**-3.756**	**-.067 ***	
dependent variable: intention to eat chocolate															
response latency	4.500	.063 **	4.932	.071 **	3.628	.052 *	18.275	.261 **	3.912	.056 ns	5.802	.085 *	2.975	.043 ns	*not estimated due to small sample size*
attitude (t_1)	-3.141	-.071 **	-2.993	-.067 **	-3.566	-.079 **	1.195	.026 ns	-3.804	-.089 **	-3.379	-.080 *	-4.156	-.097 **	
interaction effect	-1.654	-.019 ns	.257	.003 ns	.890	.010 ns	**-5.439**	**-.195 ***	.728	.008 ns	-.761	-.009 ns	1.497	.017 ns	
dependent variable: intention to drink alcohol															
response latency	2.943	.037 ns	4.074	.053 *	2.701	.035 ns	11.847	.154 $^{+}$	3.645	.047 ns	3.880	.051 ns	-1.06	-.014 ns	*not estimated due to small sample size*
attitude (t_1)	-7.138	-.145 **	-7.024	-.143 **	-6.939	-.141 **	-5.515	-.111 **	-7.943	-.166 **	-7.010	-.149 **	-7.522	-.157 **	
interaction effect	-3.663	-.037 ns	**-5.832**	**-.059 ***	**-4.861**	**-.049 ***	-3.053	-.100 ns	-4.394	-.046 ns	-5.389	-.057 $^{+}$	**-7.748**	**-.080 ****	
dependent variable: intention to eat fruit and vegetables															
response latency	.221	.004 ns	-1.229	-.021 ns	.007	.000 ns	.985	.016 ns	-.789	-.013 ns	-1.650	-.028 ns	-.980	-.017 ns	*not estimated due to small sample size*
attitude (t_1)	-11.064	-.300 **	-11.172	-.300**	-10.742	-.289 **	-11.534	-.297 **	-11.508	-.314 **	-11.820	-318 **	-11.189	-.305 **	
Interaction effect	-2.591	-.035 ns	-1.834	-.025 ns	-2.886	-.039 ns	-1.587	-.066 ns	-2.553	-.035 ns	-2.436	-.033 ns	**-4.565**	**-.062 ***	

*** $p \leq 0.01$; * $p \leq 0.05$; $^{+}$ $p \leq 0.1$; "ns" not significant with $p>0.05$; Response Latency categorized as dummy variable (0: slower than median; 1: faster than median); N>1200*
"B": unstandardized regression coefficient; "Beta": standardized regression coefficient

Table 4 shows the results of the various model estimations based on several multiple regression analyses employing OLS-estimation method. On the left side of the table one can see the estimated effects for models with measures of reaction time corrected for outliers. The right side shows the estimation results for models with measures of reaction time corrected for outliers and interviewer-validation.[7]

Looking at the left side of the table (reaction times corrected for outliers), one can see that using BS-c (transformed with Residual-Score) the interaction effect becomes significant for the behavioral intention index as well as for the intention to eat chocolate. For BS-a and BS-b the interaction effect is significant in both cases with the behavioral intention to drink alcohol. Surprisingly, the interaction effect of raw reaction times and the behavioral index becomes significant as well. Possible explanations for these results have to consider the procedure of correcting for outliers and the validation of reaction times by interviewers. As said earlier on (cf. footnote 4), outlier correction was done by cutting off all reaction time measures slower than two standard deviations above the mean respectively faster than two standard deviations below the mean. However, this procedure does not distinguish between valid and invalid reaction time measures, only extreme values are considered. Thus, it is possible that 1) either a reaction time above this threshold will be excluded although it is a valid measurement; or 2) reaction times within this threshold will be included, although they are invalid accordingly to the interviewer. Thus, different results can also be expected based on interviewer validation and a higher data quality of reaction time is expected using interviewer validation.[8]

On the right side of the table one can see that solely the Residual-Score with BS-b (demanding items) shows significant interaction effects in three cases. Neither raw reaction time nor BS-a (easy items) have any significant interaction effects. Hence, looking at these results one can conclude that only transformed reaction times calculated with the Residual-Score Index controlled for a cognitive demanding baseline speed measure show interactional effects.

---

[7] Interviewer-validation means the evaluation of the time measurement procedure employed to every survey question when the evaluation is conducted by the interviewer himself. For example, measurements have to be judged invalid when a respondent had been asking for clarification or when other people had distracted the interview process. Measurements have also rated to be invalid when an interviewer had stopped his time measurement too early (or too lately).

[8] Empirically, 10-20% of the time measurements that were labelled as invalid by the interviewers are detected as invalid by the outlier correction, but 30-60% of the time measurements that are identified as outliers are also identified by the interviewer.

5 Summary

In this paper we argued that it is necessary to control for baseline speed in order to have an adequate interpretation of cognitive processes in CATI surveys. In our view, there is no consistent definition of baseline speed in the literature to date. Thus, we began by defining baseline speed and presenting different concepts of baseline speed in order to compare them analytically. We then described traditional computations of reaction time indices and presented an additional index: the Residual-Score Index. We argued that transformed reaction time measures should be used for a solid interpretation of response latencies.

In our empirical analysis, we confirmed the need to apply a baseline speed construct and that the Residual-Score transformation index offers some analytical advantages. First we found that it does indeed make a difference whether baseline speed is constructed out of demanding versus non-demanding items. Using only non-demanding items, effects such as level of education will not be controlled for. A baseline speed measure constructed out of demanding items, however, will have more moderational power concerning the attitude–behavioral intention relationship. Our analysis also suggested that raw reaction time should not be used for substantial interpretation because in this case too many effects would remain uncontrolled or could lead to wrong conclusions. For instance, no interaction effects between time measurement (as an indicator of attitude strength) and other attitudinal predictors for future behavior could be detected if the analysis would rely on raw reaction times only (when corrected for outliers and interviewer validation, cf. table 4). Finally, we showed that using transformation indices to control for baseline speed improves interpretability of reaction times and further improves the moderational power of response latency. In particular, we showed that the Residual-Score provided two advantages: 1) the possibility to compute a baseline speed that is relative to the other respondents in the sample; and 2) the advantage of correcting the statistical relationship between raw reaction time and baseline speed almost completely.

6 Appendix

a) Attitude toward health-conscious nutrition (1: agree … 5: disagree)

Item1: "I personally think it is very good to eat in a health-conscious way."

Item2: "It is more of a disadvantage for me if I eat in a health-conscious way."

Item3: "In everyday life I think it is especially good to eat in a health-conscious way."

Item4: "Ultimately, it doesn't make a lot of sense to eat exclusively in a health-conscious way."

b) Behavioral Intention toward health-conscious nutrition (0 … 100% likelihood)

"What do you think, how likely is it that you ..."

Item1: "... will eat meat no more than 2 times per week in the next 4 weeks?"

Item2: "... will eat chocolate no more than 2 times per week in the next 4 weeks?"

Item3: "... will drink no more alcohol than 2 glasses of wine or beer per week?"

Item4: "... will eat fresh fruit and vegetable every day?"

7 **Literature**

Bassili, J.N., 1993: Response latency versus certainty as indexes of the strength of voting intentions in a CATI survey. Public Opinion Quarterly 57: 54-61.

Bassili, J.N., 2003: The minority slowness effect: Subtle inhibitions in the expression of views not shared by others. Journal of Personality and Social Psychology 84(2): 261-276.

Fabrigar, L.R. / Priester, J.R. / Petty, R.E. / Wegener, D.T., 1998: The impact of attitude accessibility on elaboration of persuasive messages. Personality and Social Psychology Bulletin 24(4): 339-352.

Fazio, R.H., 1990: A Practical Guide to the Use of Response Latency in Social Psychological Research. S. 74-97. In: Hendrick, C. / Clark, M. S. (Hrsg.), Research methods in personality and social research Newbury: Sage.

Fazio, R.H. / Dunton, B.C., 1997: Categorization by Race: The Impact of Automatic and Controlled Components of Racial Prejudice. Journal of Experimental Social Psychology 33: 451-470.

Fazio, R.H. / Herr, P.M. / Olney, T.J., 1984: Attitude Accessibility Following a Self-Perception Process. Journal of Personality and Social Psychology 47(2): 277-286.

Fazio, R.H. / Powell, M.C. / Williams, C.J., 1989: The Role of Attitude Accessibility in the Attitude-to-Behavior Process. Journal of Consumer Research 16: 280-288.

Fazio, R.H. / Powell, M.C., 1997: On the value of knowing one's likes and dislikes: Attitude accessibility, stress, and health in college. Psychological Science 8(6): 430-436.

Hertel, G. / Neuhof, J. / Theuer, T. / Kerr, N.L., 2000: Mood effects on cooperation in small groups: Does positive mood simply lead to more cooperation? Cognition & emotion 14(4): 441-472.

Houston, D.A. / Fazio, R.H., 1989: Biased processing as a function of attitude accessibility: Making objective judgments subjectively. Social Cognition 7(1): 51-66.

Huckfeldt, R. / Sprague, J., 2000: Political Consequences of Inconsistency: The Accessibility and Stability of Abortion Attitudes. Political Psychology 21(1): 57-79.

Johnson, M. / Shively, W.P./ Stein, R.M., 2002: Contextual Data and the Study of Elections and Voting Behavior: Connecting Individuals to Environment. Electoral Studies 21(2): 219-233.

Kokkinaki, F. / Lunt, P., 1999: The effect of advertising message involvement on brand attitude accessibility. Journal of Economic Psychology 20(1): 41-51.

Mayerl, J., 2003: Können Nonattitudes durch die Messung von Antwortreaktionszeiten ermittelt werden? Eine empirische Analyse comuptergestützter Telefoninterviews. Schriftenreihe des Instituts für Sozialwissen-schaften (SISS) Nr. 2/2003: Stuttgart.

Sellke, P. / Mayerl, J. (2004): How Long Does It Take To Be Ignorant? Measuring Nonattitudes With Response Latencies In Surveys.
In: van Dijkum, C./ Blasius, J./ van Hilton, B. (eds.): Recent Developments and Applications in Social Research Methodology. Proceedings of the Sixth International Conference on Logic and Methodology. Amsterdam: SISWO.

Shrum, L.J. / O'Guinn, T.C., 1993: Processes and Effects in the Construction of Social Reality. Construct Accessibility as an Explanatory Variable. Communication Research 20(3): 436-471.

Shrum, L.J. / O'Guinn, T.C., 1993: Processes and Effects in the Construction of Social Reality. Construct Accessibility as an Explanatory Variable. Communication Research 20(3): 436-471.

Stocké, V., 2001: The Influence of Frequency Scales on the Response Behavior. A Theoretical Model and ist Empirical Examination. Forschungsbericht No.01-24, Sonderforschungsbereich 504, Universität Mannheim.

Stocké, V., 2002: Die Vorhersage von Fragenreihenfolgeeffekten durch Antwortlatenzen: Eine Validierungsstudie. ZUMA-Nachrichten 50.

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart bisher sind erschienen:

No. 1/1994 "Vertrauen" - soziologisch betrachtet. Ein Beitrag zur Analyse binärer Interaktionssysteme.
Peter Antfang, Dieter Urban

No. 2/1994 Report on the German Machine Tool Industry.
Frank C. Englmann, Christian Heyd, Daniel Köstler, Peter Paustian with the assistance of Susanne Baur and Peter Bergmann

No. 3/1994 Neue württembergische Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht.
Udo Kornblum

No. 4/1994 Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht aus den neuen Bundesländern.
Udo Kornblum

No. 1/1995 Die Bedeutung Neuronaler Netze in der Ökonomie.
Hermann Schnabl

No. 2/1995 Regionale Strukturprobleme.
Sammelband der Beiträge zum Symposium vom 13. und 14.Oktober 1994.
Frank C. Englmann (Hrsg.)

No. 3/1995 Latent Attitude Structures Directing the Perception of New Technologies. An Application of SEM-Methodology to the Construction of Attitude Measurement Models Related to Technologies of Prenatal Genetic Engineering and Testing.
Dieter Urban

No. 4/1995 Handbuch zur empirischen Erhebung von Einstellungen/Kognitionen zur Bio- und Gentechnologie (inklusive Diskette)
(zweite, überarbeitete und erweiterte Auflage)
Uwe Pfenning, Dieter Urban, Volker Weiss

No. 5/1995 Social Indicators in a Nonmetropolitan County: Testing the Representativness of a Regional Nonrandom Survey in Eastern Germany.
Dieter Urban, Joachim Singelmann

No. 1/1996 Jugend und Politik im Transformationsprozeß. Eine Fallstudie zur Stabilität und Veränderung von politischen Einstellungen bei ostdeutschen Jugendlichen zwischen 1992 und 1995.
Dieter Urban, Joachim Singelmann, Helmut Schröder

No. 2/1996 Einstellungsmessung oder Einstellungsgenerierung? Die Bedeutung der informationellen Basis bei Befragten für die empirische Rekonstruktion von Einstellungen zu gentechnischen Anwendungen.
Martin Slaby

No. 1/1997 Gentechnik: „Fluch oder Segen“ versus „Fluch und Segen“. Bilanzierende und differenzierende Bewertungen der Gentechnik in der öffentlichen Meinung.
Dieter Urban und Uwe Pfenning

(Fortsetzung ...)

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart bisher sind erschienen (Fortsetzung):

No.2/1997	Die soziale Vererbung von Ausländer"feindlichkeit". Eine empirische Längsschnittanalyse der intra- und intergenerativen Transmission von sozialen Einstellungen. Dieter Urban und Joachim Singelmann
No. 3/1997	Politische Sozialisation im Transformationsprozeß: Die Entwicklung demokratiebezogener Einstellungen von ostdeutschen Jugendlichen und deren Eltern zwischen 1992 und 1996. Barbara Schmidt, Dieter Urban, Joachim Singelmann
No.1/1998	Bewertende Einstellungen zur Gentechnik: ihre Form, ihre Inhalte und ihre Dynamik. Kurzbericht zu Ergebnissen des Forschungsprojektes „Einstellungen zur Gentechnik". Dieter Urban. Uwe Pfenning, Joachim Allhoff
No.2/1998	Technikeinstellungen: gibt es die überhaupt? Ergebnisse einer Längsschnittanalyse von Bewertungen der Gentechnik. Dieter Urban
No.3/1998	Zur Interaktion zwischen Befragten und Erhebungsinstrument. Eine Untersuchung zur Konstanz des Meinungsurteils von Befragten im Interviewverlauf. Martin Slaby
No.1/1999	Role Models and Trust in Socio-Political Institutions: A Case Study in Eastern Germany, 1992-96. Joachim Singelmann, Toby A. Ten Ayck, Dieter Urban
No.1/2000	Die Zufriedenheit von Stuttgarter Studierenden mit ihrer Lebens- und Wohnsituation. Erste deskriptive Ergebnisse einer sozialwissenschaftlichen Studie zu allgemeinen und bereichspezifischen Zufriedenheiten der Studierenden des Campus Vaihingen und des Campus Hohenheim. Projektgruppe Campus: Slaby, M.; Grund, R.; Mayerl, J.; Noak, T.; Payk, B.; Sellke, P.; Urban, D.; Zudrell, I.
No.2/2000	Längsschnittanalysen mit latenten Wachstumskurvenmodellen in der politischen Sozialisationsforschung. Dieter Urban
No.1/2001	Unser „wir" - ein systemtheoretisches Modell von Gruppenidentitäten. Jan A. Fuhse
No.2/2001	Differentielle Technikakzeptanz, oder: Nicht immer führt die Ablehnung einer Technik auch zur Ablehnung ihrer Anwendungen. Eine nutzentheoretische und modell-statistische Analyse. Martin Slaby, Dieter Urban
No.3/2001	Religiosität und Profession. Longitudinale Analysen zur Entwicklung des religiösen Selbstbildes bei Erzieherinnen. Heiko Lindhorst
No.4/2001	Ist Glück ein affektiver Sozialindikator subjektiven Wohlbefindens? Dimensionen des subjektiven Wohlbefindens und die Differenz zwischen Glück und Zufriedenheit. Jochen Mayerl

(Fortsetzung ...)

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart bisher sind erschienen (Fortsetzung):

No.1/2002	Risikoakzeptanz als individuelle Entscheidung. Zur Integration der Risikoanalyse in die nutzentheoretische Entscheidungs- und Einstellungsforschung. Martin Slaby, Dieter Urban
No.2/2002	Vertrauen und Risikoakzeptanz. Zur Relevanz von Vertrauen bei der Bewertung neuer Technologien. Martin Slaby, Dieter Urban
No.3/2002	Probleme bei der Messung individueller Veränderungsraten. 13 empirisch und methodisch induzierte Effekte, die es schwierig machen, Veränderungen von generalisierten Bewertungen zu ermitteln. Dieter Urban
No.1/2003	Systeme, Netzwerke, Identitäten. Die Konstitution sozialer Grenzziehungen am Beispiel amerikanischer Straßengangs. Jan A. Fuhse
No.2/2003	Können Nonattitudes durch die Messung von Antwortreaktionszeiten ermittelt werden? Eine empirische Analyse computergestützter Telefoninterviews. Jochen Mayerl
No.1/2004	Erhöht ein Opfer-Täter-Zyklus das Risiko, Sexualstraftaten als pädosexuelle Straftaten zu begehen? Ergebnisse einer ereignisanalytischen Pilotstudie Dieter Urban, Heiko Lindhorst
No.1/2005	Persönliche Netzwerke in der Systemtheorie Jan A. Fuhse
No.2/2005	Analyzing cognitive processes in CATI-Surveys with response latencies: An empirical evaluation of the consequences of using different baseline speed measures. Jochen Mayerl, Piet Sellke, Dieter Urban