Visual analogue scales for mode-independent measurement in self-administered questionnaires

JOACHIM GERICH

Johannes Kepler University Linz, Linz, Austria

Previous research has shown evidence for mode differences between computer-assisted self-administered interviews (CASI) and paper-and-pencil interviews, especially in the case of sensitive questions. Some of these differences are explained by higher degrees of self-disclosure for CASI than for paper-and-pencil interviews, due to the more private situation with CASIs. This analysis examines the existence of different degrees of self-disclosure for CASI versus paper-and-pencil questionnaires and whether these differences can be reduced by the use of a specific response format. Judgments of items on a self-control scale with discrete 5-point (Likert-type) scales are compared with judgments on continuous visual analogue scales (VAS). Because a categorization effect for Likert-type items is assumed when pressure for social desirability is present, it is hypothesized that VAS compared to Likert-type response formats are less sensitive to mode differences.

Computer methods for data collection have increased rapidly in recent years. Techniques like CAPI (computer-assisted personal interviews), CATI (computer-assisted telephone interviews), and Internet surveys are widespread and commonly used. This article focuses on computer-assisted self-administered interviews (CASI),¹ which means that the computer is not used only by the interviewer or for the administration of the field work (as it is the case with CAPI and CATI), but is self-administered by the respondents (Couper & Nicholls, 1998; De Leeuw, Hox, & Kef, 2003; Ramos, Sedivi, & Sweet, 1998). The main difference to Web-based data collection is that CASI is offline as a rule. This means more controlled interview situations, as typically interviewers are present but respondents respond on their own.

CASI instruments in general can be used as a replacement for self-administered paper-and-pencil questionnaires. Although in most cases the traditional methods like paper-and-pencil will remain appropriate for self-administered data collection, there are some situations when CASI methods can help to extend the facilities in survey research. Reasons could be the implementation of more complex research designs like adaptive testing (Wainer, 1990), collection of network data (Gerich & Lehner, 2003), and the need for exceeding skip and branch structures in self-administered data collection (De Leeuw, Hox, Kef, & Van Hattum, 1997). Other situations are the inclusion of illiterate persons and persons with certain disabilities in survey research. With audio and video CASI and the utilization of touch screens and other devices, the inclusion of these groups is facilitated (Black & Ponirakis, 2000; De Leeuw et al., 2003; Fellingner et al., 2005; Parsons, Baum, & Johnson, 2000).

Another reason for CASI instruments could be that this mode of data administration seems to reduce the effects

of social desirability and to increase self-disclosure. Various studies conclude that CASI seems to offer a more private situation for the respondents than other forms of data collection, resulting in more self-disclosure (De Leeuw, 2002). Weisband and Kiesler (1996) conclude out of a meta-analysis that self-disclosure with CASI is higher than with face-to-face and even paper-and-pencil interviews (but with weaker effects). "Moreover, studies in which social desirability distortion was measured directly suggest that when computer instruments are administered to respondents who are alone and when respondents can respond fairly (e.g., backtrack to previous responses), they may feel particularly comfortable or less wary in giving socially undesirable answers" (Richman, Weisband, Kiesler, & Drasgow, 1999, p. 769). Especially with sensitive questions, there seem to be higher rates of self-disclosure with CASI and audio CASI (De Leeuw et al., 1997; Nicholls, Baker, & Martin, 1997; O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994; Schneider & Edwards, 2000; Turner et al., 1998).

Since this specific property is a desired mode difference for some applications, like for the study of drug abuse or illegal and risk seeking behavior (Tourangeau & Smith, 1996), different properties of CASI instruments compared to traditional methods could be undesired effects of measurement errors. With mixed-mode surveys (Dillman, 2000), where different modes of data administration are used (e.g., for the collection of the same data from different samples or comparative research with different modes) mode differences are a source of measurement bias which should be avoided. Different levels of self-disclosure could therefore be a source of measurement errors in situations of mixed-mode surveys.

With a view to mode-independent measurement, two aspects will be focused on in this paper which can be examined with the help of a small experiment. The first aspect is concerned with measurement differences of self-administered paper-and-pencil and CASI questionnaires. Both modes are self-administered, with visual communication and controlled interview situations (no audio or video CASI, interviewers are present but responses are self-administered). The main difference that remains is the medium (paper versus computer). However, the questionnaire in the CASI situation is somewhat different, as the most common type of 'one screen one item' design is used. This means that each single item is presented separately on one screen and hence—in contrast to paper-and-pencil questionnaires—at the time of response only the actual item is visible to the respondent.² This is called the segmentation effect (House & Nicholls, 1988) and is seen as the source of a possible mode effect (Dillman & Christian, 2003).

If observable mode differences exist, in a second aspect of the research it should be determined if mode dependence is affected by different types of response formats. More precisely, this is the question, which response format is more suitable for unimode construction (Dillman, 2000; Dillman & Christian, 2003), that is, actions to improve the stability of measurement across different modes. In this regard, two response formats are tested: traditional (word-supported) 5-point Likert-type scales (LIK) and visual analogue scales (VAS).

The VAS scale is a graphic rating format in the form of a horizontal line and verbal anchors for the two extremes on either end of the line. Respondents have to mark the appropriate point along the line that most closely corresponds with their favored attribute. The answers are usually coded in millimeters from the origin. VAS scales are considered to be especially preferred by respondents when their intended answers do not fit with the categories of LIK items (Flynn, Schaik, & Wersch, 2004). Stubbs et al. (2000) describe VAS scales as "easy and quick to use, simple to interpret, they do not require the subject to invoke their own descriptive terms, they allow considerable discrimination and they are presented in a standardized format" (Stubbs et al., 2000, p. 407). Other authors, however, argue that respondents need more training for VAS scales compared to LIK items, that changes and distances, especially in clinical and diagnostical research, are difficult to interpret and that no evidence can be found that VAS scales provide more reliable or more valid measures (Brunier & Graydon, 1996; Flynn et al., 2004; McKelvie, 1978). Apart from general considerations about the measurement quality of VAS scales, however, the continuous type of response could possibly reduce effects of social desirability and support self-disclosure in a specific way. When a respondent's attitude on a specific item is extreme, but pressure through social desirability is present, subjects would tend to choose the category of a LIK scale that is next to the extreme. This would be the only way on a response format with discrete categories to express an extreme but not the most extreme attitude (categorization

effect). This is different on a VAS scale because even extreme locations on the spectrum can be chosen which are near the maximum (e.g., say only one mm from the end of the line) but not necessarily exactly at the maximum point.

If this hypothesis of a categorization effect is true, then it can be expected that VAS scales are less susceptible to social desirability compared to LIK scales. Further, if pressure for social desirability is mode-dependent, VAS scales should produce more mode-independent responses compared to LIK items.

To summarize, three main hypotheses are tested in this paper: The first hypothesis assumes higher self-disclosure in CASI compared to the paper-and-pencil mode. With the second hypothesis, it is expected that mode differences are weakened for VAS scales compared to LIK scales. With the third hypothesis, it is assumed that differences between VAS and LIK scales are caused by the "categorization" effect for LIK scales; that is, respondents with extreme attitudes tend to choose a "weaker" category under exposure of social desirability.

METHOD

An experiment with university students was conducted. A questionnaire containing ten items from the German Self-Control Scale (Seipel, 2004) was assembled with three items for impulsivity, three items for risk-seeking and four items for temper.³ All items were worded and coded in the same direction, i.e., higher values indicated lower self-control (higher impulsivity, risk-seeking or temper, respectively). It was assumed that low self-control (i.e., higher impulsivity), higher risk-seeking tendency, and higher tendency toward temper are undesired personal attributes. Hence, higher values on the ten items should indicate higher self-disclosure as a tendency. It has to be noted, however, that the items of the self-control scale do not represent extremely undesired stimuli. Hence, possible effects concerning desirability may be weak.

To compare VAS and LIK items, each questionnaire contained the ten items for self-control with VAS and LIK format as separated item sets (Figure 1). Between the two sets, some other questions, like demographic variables, were presented. The questionnaire was constructed in self-administered paper-and-pencil and self-administered computer versions. The VAS scale was constructed as a line of 66 mm length. Paper-and-pencil responses were coded in millimeters, hence responses in the range between 0 and 66 were possible. The positions of respondents' marks on the VAS scale in the CASI mode were also scaled as 67 distinct points, resulting in codes from 0 to 66.

To reduce a possible order effect, for each mode (paper-and-pencil and CASI) two versions were prepared. In version one, the item set with VAS scales and in version two the item set with LIK scales was presented as the first set. The scale means of the two versions for LIK and VAS were not significantly different ($p > .05$) for paper and pencil as well as CASI.

In all, 122 interviews were conducted in a public area of the main building of the university campus. Most of the lecture rooms and the cafeteria are situated there. In the fashion of a convenience sample, students—who were walking through—were asked to complete a questionnaire and were then randomly assigned to one of the four questionnaires (CASI, paper-and-pencil, with two versions each).⁴ Four interviewers were involved. They asked respondents for participation and assigned them to a consecutive number. With help of a table of random numbers for each consecutive respondent number, the respondents were assigned to one of the four versions of ques-

Paper and Pencil:

Sometimes I take a risk, only to have fun.

not at all |-----| completely right

	not at all	rather not right	sometimes	rather right	completely right
Sometimes I take a risk, only to have fun.	☐	☐	☐	☐	☐

CASI:

Computerfragebogen

Sometimes I take a risk, only to have fun.

Click on the line for an answer

not at all |-----| completely right

☐ no answer

back forward

Frage Nr. Beantwortet Fragen

Computerfragebogen

Sometimes I take a risk, only to have fun.

not at all rather not right sometimes rather right completely right no answer

☐ ☐ ☐ ☐ ☐ ☐

back forward

Frage Nr. Beantwortet Fragen

Figure 1. Presented item formats for LIK and VAS scales in CASI and paper-and-pencil modes.

tionnaires. After a short instruction,⁵ interviewees responded the questions on their own.

One respondent interrupted the interview (in paper-and-pencil mode). Accordingly, data of 121 respondents remain for the analysis. The mean age of the respondents was 23.2 years (SD 4.1); 51% were female and 49% male, with no significant differences ($p > .05$) between the experimental groups for gender and age.

On the basis of factor analysis (principal-component with Varimax rotation) the three subscales of the self control scale are reproduced for the VAS and the LIK version of the items. For the following analyses however, there is no specific interest on these sub dimensions, hence the whole scale of all ten items for self-control is examined. Internal consistency for all ten items of the self-control scale was $\alpha = .72$ for the LIK version and .71 for the VAS version (Table 1).

Alpha is somewhat lower for the paper-and-pencil mode compared to CASI for both LIK and VAS scales. The correlation between the sum-scores of the ten LIK and VAS items was .78 for the whole sample with small variation between the two modes (.82 for paper-and-pencil and .78 for CASI mode). This implies that only 61% of the VAS score can be explained by LIK scores (or vice versa), which

appears to be a low reproducibility. However, if the correlation is interpreted as test-retest reliability, it fairly reflects the value for the internal consistency of the scale.

The first (higher self-disclosure in CASI than in paper-and-pencil mode) and the second hypotheses (weaker mode effects for VAS than for LIK scales) were examined by means of simple t tests. Hypothesis 3 (categorization effect) was put to the test with the help of the conditional distributions of VAS judgments under a given LIK judgment.

RESULTS

Table 2 shows the mean scores for the ten items on the VAS and LIK scales and separated modes.

For both response formats (LIK and VAS) higher mean scores (and also standard deviations) were observed for the CASI condition compared to paper-and-pencil. As higher self-disclosure is supposed to be associated with higher scores on each of the items, the hypothesis of higher self-disclosure under CASI condition (Hypothesis 1) is partly supported. Hypothesis 1 is only partly supported, because as shown in Table 2, mode differences appear to be significant only for LIK ($p < .05$). The effect sizes measured by Cohen's d and point-biserial correlation (interview mode with LIK and VAS) also show substantial differences. The effect size for VAS ($d = .19$) is beyond the border for "small" effect sizes (.2) following Cohen's rule of thumb (Cohen, 1977). The effect size for LIK ($d = .39$)

Table 1
Internal Consistency (Cronbach's Alpha) for the 10-Item Scale

	Total	Paper and Pencil	CASI
LIK	.72	.59	.78
VAS	.71	.67	.74
LIK/VAS correlation	.78	.82	.78

Table 2
Mode Differences for VAS and LIK Scores

	P&P		CASI		<i>t</i> Test <i>p</i>	Levene <i>p</i>	<i>d</i>	<i>r</i>	<i>n</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
LIK score	2.54	0.48	2.76	0.64	.041	.042	.39	.19	117
VAS score	24.51	8.48	26.21	9.46	.302	.521	.19	.10	120

Note—P&P, paper and pencil; *t* Test *p*, two-tailed significance of mean differences tested with *t* test (Welch); Levene *p*, Levene test for homogeneity of variance; *d*, Cohen's *d*; *r*, point-biserial correlation; *n*, number of cases.

is between the value of small (.2) and medium (.5) effects following Cohen's classification. This result supports the second hypothesis, that VAS scores are less sensitive to mode differences compared to LIK scores. Heterogeneity of variances was significant ($p \leq .05$) for LIK but not for VAS scales, which also indicates mode differences for LIK scales.

According to Hypothesis 3 (concerning the categorization effect for the LIK format), the different respondent behavior should be analyzed in more detail. For this analysis, the main focus was upon the single pairs of judgments on the LIK and VAS scales. Instead of analyzing single persons and their judgments on the 10 items, the single pairs of judgments on each item on VAS and LIK scales were objects for the analysis.

For this purpose, the original data matrix containing 121 rows of persons and 20 columns of judges (ten items with LIK and VAS) was rearranged. The data was transformed, so that each pair of judges (VAS and LIK) for each person on each of the items formed a case (judgment case). Hence the new data matrix contains 1,210 rows with two judgments (VAS and LIK) as columns. Of course, this rearrangement implies that the cases are not independent measures anymore, which limits the trustworthiness of tests for significance. Analysis of this data, however, allows comparison of VAS and LIK judgments in conjunction.

A possible categorization effect is examined by the inspection of conditional VAS judgments for a given category of a LIK judgment (Table 3). A categorization effect is considered to occur when a judgment on a high region of the VAS continuum, but not in the highest (especially not in the extreme) category of LIK format was chosen. If paper-and-pencil interview versus CASI leads to a higher

categorization effect, then it should be the case that a high LIK category in the paper-and-pencil condition corresponds to a higher VAS region than would appear in the CASI condition. The precondition for this kind of analysis is the general association of the single LIK-categories with a unique region on the VAS continuum.

To examine whether a unique location on the VAS scale exists for each location on the LIK scale, a test for multimodality was applied. Multimodality of conditional VAS judges would occur if the VAS scale was answered in a discrete (different single locations are preferred) rather than a continuous way, or if reproducibility of LIK and VAS is small. If the assumption of unimodality holds, it is possible to derive the location on VAS scales associated with a specific LIK category by the mean or median.

For the test of multimodality of conditional VAS judgments, the RUNT test suggested by Hartigan and Mohanty (1992) was applied. The test is based on the hierarchical agglomeration of single linkage cluster analysis for the VAS variable. The logic of the test is based on the idea that in the case of multimodality large clusters are merged by a single linkage algorithm, indicating more than one mode for a distribution. The RUNT test statistic is the maximum of the smaller size of merged clusters (called the "runt") on each step of hierarchical agglomeration. In the case of multimodality, the RUNT statistic exceeds $(n/2.1) + 0.5$ with $p \leq .05$ (Hartigan & Mohanty, 1992, p. 67). The RUNT statistic and the critical values for each conditional distribution of VAS judges are given in Table 3. Since none of the RUNT values exceed the critical RUNT, it can be assumed that unimodality for VAS distributions holds within each LIK category.

Because unimodality was assumed, the location of LIK judges on the VAS continuum was inspected by lo-

Table 3
Conditional Distribution of VAS Judgments by LIK Categories

LIK	VAS			<i>n</i>	RUNT	Crit(RUNT)
	<i>M</i>	<i>SD</i>	Median			
1	6.25	11.22	3	227	71	109
2	16.40	12.62	13	378	139	181
3	30.69	12.65	32	262	77	125
4	41.72	13.70	45	237	58	113
5	52.05	14.66	55	93	30	45
Total	25.39	19.28	22	1,197		

Note—RUNT, $\max(c_i)$ with $c_i = \min(c_{ip}, c_{iq})$, where c_{ip}, c_{iq} are the cluster sizes of merge step *i* in a hierarchical single-linkage cluster algorithm; Crit(RUNT), critical cluster size for the RUNT test statistic for $p \leq .05$, given by $\text{Crit(RUNT)} = (n/2.1) + 0.5$, with *n* as the number of objects (Hartigan & Mohanty, 1992).

Figure 2. Conditional cumulative distributions of VAS judgments by LIK category.

cation parameters. As shown in Table 3, monotonic association between the two response formats was given which was indicated by increasing means and medians of VAS judges for higher categories of the LIK judges. The standard deviation of the conditional VAS judgments was also increasing with higher categories of LIK judgments. Inspecting the distances of the conditional location parameters showed for both—mean and median—that the distance between LIK Categories 2 and 3, in particular, was somewhat higher than the other distances.

Next, the location of LIK judgments on the VAS scales was examined on the basis of the conditional cumulative distributions (Figure 2).

Figure 2 replicates that the distance between LIK Categories 2 and 3 is larger than between the others. The figure shows further that the curves for LIK Categories 3, 4, and 5 have moderate slopes in comparison with Categories 2 and 1. This could indicate that the LIK Categories 3–5 are less accurately located in a specific region on the VAS scale, according to lower discrimination of VAS scores in this region. In Figure 3, the same conditional distributions are split by the mode of data collection (CASI and paper-and-pencil).

The figure shows largely the same distribution, except for LIK Category 5 with a greater distance between the curves in CASI and paper-and-pencil mode. In the paper-and-pencil mode, the conditional VAS judgments are situated in a higher region than for CASI. There are also some differences for LIK Categories 2 and 3. However, the total area between the curves of paper-and-pencil and CASI distributions for these categories is smaller than that for Category 5.⁶ The difference of the location is also illustrated by the Kolmogorov–Smirnov test which indicates significant differences of the distributions between CASI and paper-and-pencil only for LIK Category 5. Simple OLS regression for conditional VAS scores as dependent (and CASI versus paper-and-pencil as predictor) would lead to

the same results. However, due to the rearrangement of the data, these methods are formally not appropriate for the case of nested data. Therefore, a multilevel means-as-outcome model with CASI (versus paper-and-pencil) as covariate on Level 2 (person level) was applied.⁷ The results of these analyses also illustrate the larger mode effect with the same direction for LIK Category 5 (compared with the other LIK categories). The regression coefficient of the multilevel model, however is not significant ($p = .175$). The nonsignificant result of the multilevel analysis is likely due to the small sample size of persons.

Even though the results of the multilevel analysis are not significant, the results indicate that in the paper-and-pencil mode the LIK Category 5 is only chosen by those with especially high values for the VAS scale, than in CASI mode. It seems that in the paper-and-pencil mode, the maximum LIK category is only chosen if the attitude of respondents is very distinct.

However, as the cumulative VAS curve for LIK Category 5 in CASI mode is flatter than in the paper-and-pencil mode (which is also expressed by the higher standard deviation in Table 4), the discrimination of the VAS scale is possibly lower in the CASI mode.

DISCUSSION AND CONCLUSION

The experimental analysis examined answers on ten items for self-control. It was assumed for all items that higher values are associated with higher self-disclosure as a tendency, as low self-control is an undesired personal attribute. Based on the total mean score for the ten items, higher values for self-control for CASI compared to the paper-and-pencil mode were observed. It was also shown that the mode difference is stronger and significant for the LIK scales and substantially weakened—and not significant with the actual sampling size—for the VAS scales. Therefore, the hypothesis of higher self-disclosure for the

Figure 3. Conditional cumulative distributions of VAS judgments by LIK category for CASI (dashed lines) and paper and pencil (solid lines).

CASI instruments is partially supported, and the second hypothesis, that mode differences caused by different degrees of self-disclosure are weaker for VAS compared to LIK scales, is also supported.

To analyze the response difference in more detail, judgment cases containing VAS and corresponding LIK responses were examined. It was shown that VAS and LIK judgments are associated in a monotonic way. Further, the assumption of unimodality for conditional distributions of VAS judgments for each LIK category was confirmed. Together these results suggest that LIK judgments can be located on a specific area of the VAS scale, which can be expressed by location parameters like median or mean. The analysis showed that the location of LIK categories on the VAS continuum is approximately the same for CASI and the paper-and-pencil mode except for LIK Category 5. The inspection of the conditional distributions of VAS judgments under a given LIK category suggests that respondents with more extreme attitudes are able to express their position on a VAS scale by choosing high—but not maximum—values, even if the tendency for self-disclosure is low. On the other hand, with a 5-point LIK

response format, a low tendency for self-disclosure would cause the tendency to avoid the maximum but prefer the choice of Category 4 instead (categorization effect). However, as the results of the multilevel model are not significant, the hypothesis of a categorization effect was not supported with the sample size of the study at hand. Additional research with larger sample sizes would be necessary to strengthen the results.

There are, of course, some further limits for these findings. In general (for both CASI and paper and pencil) it was shown that the LIK locations on the VAS scales do not appear to be equidistant. This would be a limitation of the measurement quality of rating items. It cannot be clarified, however, whether one of the two response formats is superior in this regard.

It was also shown that the distances between the VAS locations of LIK Categories 4 and 5 and also the discrimination of VAS judgments for LIK Category 5 are reduced in CASI mode. This possibly indicates further mode differences that effect lower precision of the VAS judgments in CASI mode. Furthermore, a word-supported 5-point Likert-type format was used in the actual study. Hence,

Table 4
Conditional VAS Distribution in CASI Versus Paper-and-Pencil Mode

LIK	VAS, Paper and Pencil			VAS, CASI			<i>n</i> , CASI	<i>n</i> , P&P	<i>p</i> (K-S)	OLS		Multilevel	
	<i>M</i>	<i>SD</i>	Median	<i>M</i>	<i>SD</i>	Median				<i>B</i> _{CASI}	<i>p</i>	<i>B</i> _{CASI}	<i>p</i>
1	6.35	10.55	3.0	6.14	12.01	2.0	104	123	.321	-0.21	.891	0.15	.935
2	17.26	12.58	13.0	15.48	12.64	12.0	183	195	.074	-1.76	.172	-1.47	.414
3	31.36	12.05	32.0	29.97	13.26	31.0	127	135	.090	-1.39	.374	-1.55	.439
4	41.06	14.50	45.0	42.25	13.07	44.0	132	105	.249	1.19	.507	1.51	.506
5	56.72	12.09	59.5	49.61	15.37	53.0	61	32	.026	-7.11	.025	-4.84	.175

Note—*n*, CASI/P&P, number of cases for CASI and paper and pencil; *p* (K-S), two-tailed exact significance of a Kolmogorov-Smirnov test; OLS, unstandardized regression coefficient and *p* value for CASI mode, with VAS score as dependent variable (OLS regression); Multilevel, unstandardized regression coefficient and *p* value for CASI mode, with VAS score as dependent variable (multilevel means-as-outcome model).

it cannot be determined if other formats without wording support or formats with more than five categories would have changed the results as—for instance—a higher number of categories could reduce possible categorization effects.

Further, as already mentioned, the items that were used in this study are not supposed to induce extreme effects on social desirability. Therefore, it may be possible that mode-differences for the VAS format were not detected only due to small effects induced. Hence, it would be possible, that the differences that occurred between VAS and LIK formats would be reduced for items with higher sensitivity (e.g., sexual risk behavior or illegal drug use).

In sum, the results may provide first suggestions to use VAS scales for reduction of mode differences especially with mixed-mode surveys. Due to the exploratory nature of the study at hand, further research is needed to examine whether the results transfer to other measurement scales and populations. Especially further research with larger sample sizes is needed to clarify the hypothesis of categorization effects. For this purpose, future work should vary response formats (e.g., “ticks” for VAS, additional word support for VAS, different extent of word support for LIK, different numbers of response categories for LIK), scales and populations.

AUTHOR NOTE

Correspondence relating to this article may be sent to J. Gerich, Department of Sociology, Johannes Kepler Universität Linz, Altenbergerstr. 69, 4040 Linz, Austria (e-mail: joachim.gerich@jku.at).

REFERENCES

- BLACK, M. M., & PONIRAKIS, A. (2000). Computer administered interventions with children about maltreatment: Methodological, developmental, and ethical issues. *Journal of Interpersonal Violence*, **15**, 682-695.
- BRUNIER, G., & GRAYDON, J. (1996). A comparison of two methods of measuring fatigue in patients on chronic haemodialysis: Visual analogue vs. Likert scale. *International Journal of Nursing Studies*, **33**, 338-348.
- COHEN, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.
- COUPER, M. P., & NICHOLLS, W. L. (1998). The history and development of computer-assisted survey information collection methods. In M. P. Couper et al. (Eds.), *Computer-assisted survey information collection* (pp. 1-21). New York: Wiley.
- DE LEEUW, E. (2002). The effect of computer-assisted interviewing on data quality: A review of the evidence. In J. Blasius, J. Hox, E. De Leeuw, & P. Schmidt (Eds.), *Social science methodology in the new millennium* [CD-ROM]. Opladen, Germany: Leske & Budrich.
- DE LEEUW, E., HOX, J., & KEF, S. (2003). Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods*, **15**, 1-29.
- DE LEEUW, E., HOX, J., KEF, S., & VAN HATTUM, M. (1997). Overcoming the problems of special interviews on sensitive topics: Computer-assisted self-interviewing tailored for young children and adolescents. In *Sawtooth Software Conference Proceedings* (pp. 1-14). Sequim, WA: Sawtooth Software.
- DILLMAN, D. A. (2000). *Mail and Internet surveys: The tailored design method*. New York: Wiley.
- DILLMAN, D. A., & CHRISTIAN, L. M. (2003). Survey mode as a source of instability in responses across surveys. Retrieved 21 June, 2005, from survey.sesrc.wsu.edu/dillman/papers/Mixed%20Mode%20Submission%20to%20Field%20Methods.pdf.
- FELLINGER, J., HOLZINGER, D., DOBNER, U., GERICH, J., LEHNER, R., LENZ, G., & GOLDBERG, D. (2005). An innovative and reliable way of measuring health-related quality of life and mental distress in the deaf community. *Social Psychiatry & Psychiatric Epidemiology*, **40**, 245-250.
- FLYNN, D., SCHAICK, P., & WERSCH, A. (2004). A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*, **20**, 49-58.
- GERICH, J., & LEHNER, R. (2003). Egozentrierte Netzwerkerhebung mittels selbstadministrierter Computerinterviews. *Österreichische Zeitschrift für Soziologie*, **2003**(4), 45-70.
- HARTIGAN, J. A., & MOHANTY, S. (1992). The RUNT test for multimodality. *Journal of Classification*, **9**, 63-70.
- HOUSE, C. C., & NICHOLLS, W. L. (1988). Questionnaire design for CATI: Design objectives and methods. In R. M. Groves et al. (Eds.), *Telephone survey methodology* (pp. 421-436). New York: Wiley.
- McKELVIE, S. J. (1978). Graphic rating scales—How many categories? *British Journal of Psychology*, **69**, 185-202.
- NICHOLLS, W. L., II, BAKER, R. P., & MARTIN, J. (1997). The effect of new data collection technologies on survey data quality. In L. Lybert et al. (Eds.), *Survey measurement and process quality* (pp. 221-248). New York: Wiley.
- O'REILLY, J. M., HUBBARD, M. L., LESSLER, J. T., BIEMER, P. P., & TURNER, C. F. (1994). Audio and video computer-assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, **10**, 197-214.
- PARSONS, J. A., BAUM, S., & JOHNSON, T. P. (2000). *Inclusion of disabled populations in social surveys: Review and recommendations*. Paper prepared by the Survey Research Laboratory, University of Illinois at Chicago.
- RAMOS, M., SEDIVI, B. M., & SWEET, E. M. (1998). Computerized self-administered questionnaires. In M. P. Couper et al. (Eds.), *Computer-assisted survey information collection* (pp. 389-408). New York: Wiley.
- RICHMAN, W. L., WEISBAND, S., KIESLER, S., & DRASGOW, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, **84**, 754-775.
- SCHNEIDER, S. J., & EDWARDS, B. (2000). Developing usability guidelines for AudioCasi respondents with limited literacy skills. *Journal of Official Statistics*, **16**, 255-271.
- SEIPEL, C. (2004). Deutsche Version der Self-Control Skala von Grasmick et al. (1993). In A. Glöckner-Rist (Ed.), *ZUMA-Informationssystem: Elektronisches Handbuch sozialwissenschaftlicher Erhebungsinstrumente* (ZIS Version 8.00). Mannheim: ZUMA.
- STUBBS, R. J., HUGHES, D. A., JOHNSTONE, A. M., ROWLEY, E., REID, C., ELIA, M., ET AL. (2000). The use of visual analogue scales to assess motivation to eat in human subjects: A review of their reliability and validity with an evaluation of new hand-held computerized systems for temporal tracking of appetite ratings. *British Journal of Nutrition*, **84**, 405-415.
- TOURANGEAU, R., & SMITH, T. W. (1996). Asking sensitive questions: The impact of data collection, question format, and question context. *Public Opinion Quarterly*, **60**, 275-304.
- TURNER, C. F., KU, L., ROGERS, S. M., LINDBERG, L. D., PLECK, J. H., & SONENSTEIN, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, **280**, 867-873.
- WAINER, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Erlbaum.
- WEISBAND, S., & KIESLER, S. (1996, April). *Self-disclosure on computer forms: Meta-analysis and implications*. Paper presented at the Human Factors in Computing Systems conference, Vancouver. Retrieved 23 May, 2002, from www.acm.org/sigchi/chi96/proceedings/papers/Weisband/sw_txt.htm.

NOTES

1. Sometimes the acronym CSAQ (computerized self-administered questionnaire) is used.
2. Respondents, of course, are allowed to navigate between the questions, but the whole questionnaire is not visible like it is with paper and pencil.

3. The questionnaire includes items like "I do not care much about future" (impulsivity), "Sometimes I will take a risk just for fun" (risk-seeking), or "I get mad easily" (temper).

4. Past experience of other research showed that this is the best place on the campus to recruit students from the full range of disciplines of the university. The aim in the actual study, however, was not to draw a representative sample, but rather to create a random allocation to the different modes of administration.

5. In the paper-and-pencil situation, the response formats were explained on the paper. In the CASI situation, this was done with help of a test screen.

6. The area between the lines is 109.3 for LIK Category 1, 189.0 for LIK Category 2, 231.1 for LIK Category 3, 185.9 for LIK Category 4, and 710.4 for LIK Category 5. Hence, the area between the curves for

LIK Category 5 is more than three times larger than the area between the curves for LIK Category 3.

7. The means-as-outcome model assumes group constants and residuals on Level 1 (which are the judgment cases) and covariates on Level 2 (persons), resulting in the following model:

$$\text{Level 1: } \text{VAS}_{ij} = \beta_{oj} + r_{ij}$$

$$\text{Level 2: } \beta_{oj} = \gamma_{00} + \gamma_{01} * \text{CASI} + u_{oj},$$

with i indicating judgment cases and j indicating persons. The model was estimated with HLM 6.

(Manuscript received November 14, 2006;
accepted for publication November 27, 2006.)