Japanese Journal of Applied Psychology
2012, Vol. 38 (special edition), 101–108

Minor Study

Validity of Computer-Assisted Health Assessment Using Visual Analogue Scales and Face Scales:

A Comparison between Younger and Older Participants

Tomoaki KIMURA,*, ** Yasutami TSUDA,* Seiya UCHIDA,*
Akira EBOSHIDA,** and Kiyoshi YAMAOKA*

To verify the validity of the visual analogue scale (VAS) and the face scale (FS) administered on personal computers (PCs), 293 participants completed PC versions of the VAS, the FS, and the Japanese version of the Perceived Stress Scale (JPSS), as well as paper-and-pencil (P-P) versions of the VAS, the FS, and the 8-Item Short-Form Health Survey (SF-8). The VAS and FS measured generic quality of life, health, and stress. For the younger participants (under 60 years of age), the degree of agreement for all intraclass correlation coefficients (ICC) between PC and P-P versions exceeded 0.7; for the older group, the ICC for the stress item of the VAS was under 0.5. For the younger group, correlation coefficients between the JPSS and stress items of the VAS and FS exceeded 0.5; for the older group, the stress item of the VAS was only weakly correlated with the JPSS and not significantly correlated with the mental component summary (MCS) of the SF-8. These results suggest that the VAS did not have sufficient validity for older respondents.

Key words: computerized, elderly, face scale (FS), perceived stress, quality of life, visual analogue scale (VAS)

INTRODUCTION

Computer-assisted measurement has the advantages of preventing loss of data, allowing the display of results to respondents, and saving time and effort required to enter data. Moreover, the widespread availability of low-priced personal computers (PCs) equipped with touch screen monitors facilitates the use of computer-assisted health surveys, even for participants who are unfamiliar or inexperienced with computers.

The validity of results of many kinds of health assessment questionnaires developed for administration on PCs has been demonstrated. For example, the results of a questionnaire for a PC touch screen to measure quality of life (QOL) were compared with those of a traditional paper-and-pencil (P-P) version for cancer patients. The QOL scores of the two versions yielded a high agreement ratio (Velikova, Wright, Smith, Cull, Gould, Forman, Perren, Stead, Brown, & Selby, 1999). In another example, the Quality of Life in Reflux and Dyspepsia questionnaire yielded results that were equally reliable and valid, regardless of whether patients with gastroesophageal reflux disease responded to the questionnaire using a PC touch screen monitor or P-P (Kleinman, Leidy, Crawley, Bonomi, & Schoenfeld, 2001). Furthermore, re-

* MOA Health Science Foundation, Shizuoka, Japan
** Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan
Address for correspondence: Tomoaki Kimura, MOA Health Science Foundation, 1–1–60 Nishi-Atami, Atami, Shizuoka 413–0038, Japan
Tel. +81–557–86–0663 Fax. +81–557–86–0665 E-mail: t-kimura@mhs.or.jp

sults from the PC version of the Work Productivity and Activity Impairment instrument to measure the effects of specific allergies were consistent with those of the P-P version, but not with those from a Web-based version (Litaker, 2003). Nonetheless, although computer-assisted data collection has become popular, further studies are needed to assess the validity of such data in other settings and with other instruments.

Health assessments are conducted using Likert-type instruments and graphic format scales: a visual analogue scale (VAS) or a face scale (FS). VAS, originally developed at the Scott Paper Company in 1920, has traditionally been used to measure feelings (e.g., mood and pain) (Freyd, 1923). The test-retest reliability of this scale has been well documented (Hayes, Patterson, & The Scott Co., Laboratory, 1921). The VAS also offers statistical advantages compared to Likert-type questionnaires with a four-point scale (Joyce, Zutshi, Hrubes, & Mason, 1975). The VAS enables respondents, or raters, to make very fine discriminations. VAS scores have an interval scale; therefore, parametric statistical techniques, such as analysis of variance and linear regression analysis that are robust tests with considerable power, can be used to analyze VAS data. Moreover, if the frequency distribution of VAS scores is skewed, it can be improved toward normality by arcsine transformation (Aitken, 1969).

The FS was developed to measure mood among patients, using 20 faces illustrating various moods (Lorish & Maisiak, 1986). The FS for children, which was developed to assess the severity of pain (Bieri, Reeve, Champion, Addicoat, & Ziegler, 1990), is also beneficial for use among people who have difficulty with verbal communication. The validity and test-retest reliability of the FS to assess pain among the elderly has been confirmed (Herr, Mobily, Kohout, & Wagenaar, 1998).

However, further studies with this population are needed because the value of the VAS with the elderly has not been clarified. The results of a study of a community-based effort to gather data on the health status of elderly Japanese demonstrated that the VAS is a valid self-administered instrument; however, much data was missing (Kimura, Hayashida, Araki, Morita, Yamaguchi, & Eboshida, 2008). The previous study suggested that elderly participants could not understand the intention of the investigators or how to respond.

This study seeks to verify the validity of PC versions of the VAS and FS, and to compare the validity of these scales for younger participants with that for older participants. Our hypothesis was that VAS and FS scores for older people are less valid than those for younger people, and that the validity of the VAS is lower than that of the FS. Empirically, younger people are considered more familiar with the use of PCs and response to the scales than older people. Moreover, the FS is an easily understood format even for people with insufficient reading and writing skills.

METHODS

Participants and procedure

Data were collected at an event to promote food and health, which was held in Shizuoka, Japan, in November 2006. Visitors to the event were mainly healthy people interested in healthful living. Using a signboard and poster, we recruited participants from the visitors at a health-checkup booth. The participants agreed to respond to our questions after listening to well-trained receptionists at the booth explain the aims of the survey and the information to be collected.

Two hundred ninety-three participants (105 males and 188 females) completed the PC versions of the VAS and FS, as well as the Japanese version of the Perceived Stress Scale (JPSS). They then completed the P-P versions of the VAS and FS with the same format as the PC version, and the Japanese version of the 8-item Short-Form Health Survey (SF-8). Table 1 indicates the number of participants by gender and age group.

Laptop PCs equipped with touch screen monitors (12.1 in) were used in this study. The PC monitors displayed only one item. Participants pushed (clicked) a button of choice and then the next button, and the monitors displayed the next item. Participants unfamiliar with PCs were assisted by staff. A well-designed display with large characters and easy-to-see colors and layout were used; however, if necessary, staff read questions or

lent reading glasses to aged participants.

Measures

The VAS and FS consisted of three items: (1) generic QOL, (2) health status, and (3) stress. The VAS to measure generic QOL read as follows: "Please consider your life. Your life falls somewhere between the best state and the worst state. Please rate your life." Respondents could check (i.e., click) on the line between the ends of "the worst life" and "the best life" (Fig. 1). Similarly, the VAS about health status read: "Please consider your health status. Your health status falls somewhere between the best state and the worst state. Please rate your health status." Respondents rated the present state of their generic QOL and health status, but they rated their stress based on the previous month. Therefore, the VAS about stress read: "During the past month, to what extent have you been stressed? Your stress level falls somewhere between no stress and the strongest stress. Please rate your stress."

The FS to measure generic QOL asked, "Are you satisfied with your daily life?" To respond, the participant could push a button to select one of five faces that best represented his or her choice. Words associated with the faces to clarify their meaning were "very satisfied," "satisfied," "neutral," "dissatisfied," and "very dissatisfied." Responses on the FS were therefore a combination of faces and words (Fig. 2). Similarly, the FS about health status asked, "How do you feel about your health status?" The words associated with the five faces were "very good," "good," "neutral," "poor," and "very poor." The FS about stress asked, "During the past month, to what extent have you been stressed?" The words associated with the faces were "not at all," "not very strongly," "somewhat," "strongly," and "very strongly."

The Perceived Stress Scale (PSS) is a self-administered measure of the degree to which situations in a subject's life are appraised as stressful (Cohen, Kamarck, & Mermelstein, 1983). This scale was translated into Japanese (JPSS), and its validity has been verified (Iwahashi, Tanaka, Fukudo, & Hongo, 2002). The JPSS consists of 14 five-point Likert-type items; seven of the items are reverse scoring questions. Each item is scored from 0 to 4; the range of the total score is 0 to 56. A higher the score indicates greater perception of stress.

The SF-8 was developed to assess the health-related QOL in large-scale health surveys (Ware, Kosinski, Dewey, & Gandek, 2001). It is a shortened version (8 items) of the SF-36 (36 items), which is currently used internationally (Ware & Sherbourne, 1992). This measure yields two summary scores: a physical component summary (PCS) and a mental component summary (MCS), presented as deviation scores (the national standard score is 50). Higher scores mean better health status. The SF-8 was translated into Japanese, and national standard scores for the SF-8 have been established for Japanese people (Fukuhara & Suzukamo, 2004).

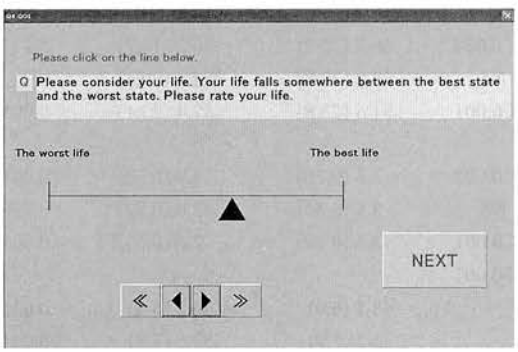

Fig. 1. A sample of the PC version of visual analogue scale for generic quality of life.

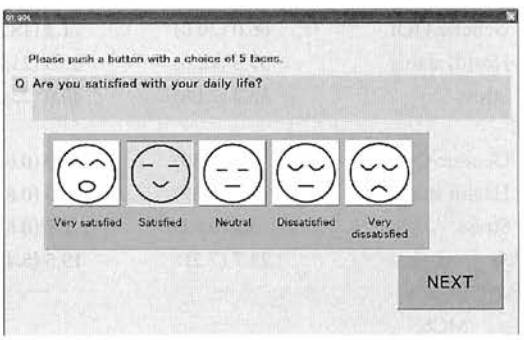

Fig. 2. A sample of the PC version of face scale for generic quality of life.

Statistical analyses

Two analyses were conducted according to the age of the participants: (a) those younger than 60 years of age and (b) those 60 years of age and older. If participants were divided at 70 years, the number of subjects in older group might have been too small for analysis. The means of the VAS, FS, JPSS, and SF-8 (PCS and MCS) were calculated and compared between age groups by the Mann–Whitney *U* test. The intraclass correlation coefficients (ICCs) were analyzed to confirm the degree of agreement between the PC versions and P-P versions of the VAS and FS; kappa values were analyzed for only the FS. Spearman's correlation coefficients were analyzed between the VAS and the FS, which have the same concepts; between the JPSS and VAS and FS; and between the SF-8 and VAS and FS. Statistical significance was set at $p<0.05$. These statistical analyses were conducted using SPSS for Windows, version 13.0 (SPSS Inc., 2004).

RESULTS

Both PC versions and P-P versions were completed by 293 participants. All data from the PC versions were available; but among the P-P versions, 279 data inputs were valid but the others included missing values (Table 1). Therefore, 279 subjects were used for statistical analyses: 180 younger subjects (<60 years of age) and 99 older subjects (>=60 years of age).

Table 2 indicates the means and standard deviations (*SD*s) of the scales by age group. The older subjects had a higher generic QOL as indicated in the PC version of the VAS and both versions of the FS. However, no significant difference between age groups was indicated in the P-P version of the VAS. The health status scores generally cor-

Table 1. The number of participants by gender and age group.

Age group	<30	30–39	40–49	50–59	60–69	70=<	Total
Male	9	22	14	26	18	16	105
Female	28	26	15	49	45	25	188
Total	37	48	29	75	63	41	293
Available*	37	46	28	69	63	36	279

*All data were effective without missing data. They were available for statistical analysis.

Table 2. Comparisons between age groups in the visual analogue scales (VAS), face scales (FS), the JPSS, and the SF-8.†

	PC version			P–P version *		
	Younger ($n=180$) mean (*SD*)	Older ($n=99$) mean (*SD*)	*p* Value	Younger ($n=180$) mean (*SD*)	Older ($n=99$) mean (*SD*)	*p* Value
VAS						
Generic QOL	66.0 (20.0)	71.2 (18.5)	0.024	66.7 (20.2)	70.2 (16.7)	NS
Health status	59.3 (22.0)	63.3 (21.2)	NS	58.8 (22.7)	64.4 (20.4)	NS
Stress	52.2 (24.8)	40.0 (25.2)	<0.001	51.6 (23.8)	37.4 (23.1)	<0.001
FS						
Generic QOL	3.6 (0.80)	3.9 (0.63)	0.002	3.6 (0.78)	3.9 (0.50)	<0.001
Health status	3.4 (0.87)	3.5 (0.81)	NS	3.4 (0.86)	3.6 (0.72)	0.036
Stress	3.2 (0.82)	2.7 (0.67)	<0.001	3.1 (0.79)	2.6 (0.63)	<0.001
JPSS	23.7 (7.2)	19.5 (5.4)	<0.001	—	—	
SF-8 PCS	—	—		48.3 (6.0)	46.5 (6.4)	0.035
MCS	—	—		46.2 (7.9)	50.5 (5.9)	<0.001

† Analysis was coducted using Mann–Whiteny *U* test.

* Paper and pencil version.

Table 3. Degree of agreement between the PC versions and paper-and-pencil versions in the visual analogue scales (VAS) and face scales (FS) by age group.†

	Younger ($n=180$)				Older ($n=99$)			
	ICC	95% confidence interval [lower]	[upper]	kappa Value	ICC	95% confidence interval [lower]	[upper]	kappa Value
VAS								
Generic QOL	0.865	0.823	0.898	—	0.755	0.655	0.828	—
Health status	0.849	0.802	0.885	—	0.793	0.706	0.856	—
Stress	0.773	0.707	0.826	—	0.447	0.274	0.591	—
FS								
Generic QOL	0.872	0.832	0.903	0.768	0.644	0.513	0.746	0.540
Health status	0.856	0.812	0.891	0.670	0.783	0.693	0.849	0.615
Stress	0.770	0.703	0.823	0.547	0.641	0.509	0.744	0.514

† Analysis was coducted using intraclass correlation coefficient (ICC) and kappa value.

Table 4. Correlations between the visual analogue scale (VAS) and the face scale (FS) with the same concepts.†

	Younger ($n=180$)		Older ($n=99$)	
	Correlation	p Value	Correlation	p Value
PC version				
Generic QOL	0.610	<0.001	0.378	<0.001
Health status	0.723	<0.001	0.707	<0.001
Stress	0.712	<0.001	0.440	<0.001
P–P version*				
Generic QOL	0.617	<0.001	0.318	0.001
Health status	0.701	<0.001	0.721	<0.001
Stress	0.727	<0.001	0.457	<0.001

†Analysis was coducted using Spearman's correlation coefficient.
*Paper and pencil version.

Table 5. Correlations of the JPSS with the visual analogue scale (VAS) and face scale (FS) in the PC version.†

	Younger ($n=180$)		Older ($n=99$)	
	Correlation	p Value	Correlation	p Value
VAS				
Generic QOL	−0.535	<0.001	−0.513	<0.001
Health status	−0.388	<0.001	−0.484	<0.001
Stress	0.564	<0.001	0.299	0.003
FS				
Generic QOL	−0.539	<0.001	−0.446	<0.001
Health status	−0.476	<0.001	−0.437	<0.001
Stress	0.555	<0.001	0.510	<0.001

†Analysis was coducted using Spearman's correlation coefficient.

responded; however, on the P-P version of the FS, scores of the older group indicated a better health status. The younger subjects felt stronger stress on all scales, including the JPSS. In the SF-8, the younger group had higher scores in PCS and lower scores in MCS.

Table 3 indicates the degree of agreement between the PC and P-P versions by age group. All ICCs exceeded 0.7 for the younger group, whereas for the older group, the ICC of the stress item of the VAS was under 0.5. All kappa values of the FS exceeded 0.5.

Correlations between the VAS and the FS with the same concepts are indicated by age group in Table 4. All correlation coefficients exceeded 0.6 for the younger group; however, for the older group, they were under 0.5, with the exception of health status.

Correlations of the JPSS with the VAS and FS in the PC version are indicated by age group in Table 5. Correlation coefficients between the JPSS and the items about stress in the VAS and FS exceeded 0.5 for the younger group; for the older group, there was a weak correlation between the JPSS and the stress item of the VAS. The JPSS was also correlated with the items about generic QOL and health status.

Correlations of the SF-8 with the VAS and FS in the PC version are indicated by age group in Table 6. The PCS of SF-8 was correlated with the items about health status. For the older group especially,

Table 6. Correlations of the SF-8 with the visual analogue scale (VAS) and face scale (FS) in the PC version.[†]

	Physical component summary (PCS)				Mental component summary (MCS)			
	Younger (n=180)		Older (n=99)		Younger (n=180)		Older (n=99)	
	Correlation	p Value	Correlation	p Value	Correlation	p Value	Correlation	p Value
VAS								
Generic QOL	—	NS	—	NS	0.516	<0.001	0.371	<0.001
Health status	0.353	<0.001	0.522	<0.001	0.242	0.001	0.245	0.015
Stress	—	NS	−0.303	0.002	−0.501	<0.001	—	NS
FS								
Generic QOL	0.169	0.024	—	NS	0.382	<0.001	0.213	0.034
Health status	0.455	<0.001	0.447	<0.001	0.264	<0.001	0.310	0.002
Stress	—	NS	—	NS	−0.527	<0.001	−0.450	<0.001

[†] Analysis was coducted using Spearman's correlation coefficient.

the correlation coefficient between the PCS and the health status item of the VAS exceeded 0.5. Correlation coefficients between the MCS of SF-8 and the stress items exceeded 0.5 for the younger group. For the older group, however, there was no significant correlation between the MCS and the stress item of the VAS.

DISCUSSION

The VAS and FS are frequently used in surveys and clinical trials because of their popularity and convenience. Moreover, PC-administered versions of these scales offer the possibility of even more convenience and sophistication. This study was conducted to verify the validity of the PC versions. Most validation studies of questionnaires are conducted among relatively young participants who comprehend the aims of the various studies and the meanings of the items in a questionnaire or scale. In contrast, this study sought to verify the validity of the PC versions for older people who are often unfamiliar with computers and scales.

The results of this study reveal the following.

1. For younger adults, the PC versions of the VAS and FS were strongly correlated with the P-P versions. Therefore, computer-assisted data collection was valid for this group.
2. The results for older people suggest that the PC version of the VAS and FS did not have sufficient validity, especially with regard to the stress items.
3. The conceptual validity of the PC versions of the VAS and FS was clarified for the younger group but not for the older group.

These findings corresponded with our hypothesis, mentioned in the introduction, that VAS and FS scores for the older group were less valid than those for the younger group, and that the validity of the VAS for the older group was lower than that of the FS. In fact, the differences between younger and older people were more remarkable than expected. With regard to the degree of agreement between PC and P-P versions, ICCs of the VAS varied from 0.865 to 0.773 for the younger group and from 0.793 to 0.447 for the older group (Table 2). An ICC of under 0.5 occurred for only the older participants' responses to the item about stress. ICCs of the FS in both groups, however, exceeded 0.6. Even the ICC of the stress item for the older group was 0.641, and the pertinent kappa value was 0.514. These results suggest that the older adults had difficulty responding to only the stress item of the VAS, although the validity of the scales for them might be less sufficient than that for younger adults.

With regard to comparisons of the VAS and FS, correlation coefficients of the PC version for the item about health status were 0.723 for the younger group and 0.707 for the older group. Thus, they approximately corresponded. For the older group,

however, correlation coefficients for generic QOL and stress items were considerably lower than those for the younger group. Since these results corresponded to those of the P-P versions, weak correlations for the older group suggest that the older participants might have more difficulty understanding the concepts of these items than in the PC-assisted measurement.

Comparison of the VAS, FS, and JPSS indicated that for the older group only the stress item of the VAS was weakly correlated with the JPSS. For the younger group, the stress and generic QOL items of the VAS and the FS were considerably correlated with the JPSS ($r>0.5$); however, for the older group, the correlation coefficient between the stress item of the FS and the JPSS was 0.510 ($p<0.001$), and the correlation coefficient between the stress item of the VAS and the JPSS was 0.299 ($p=0.003$). Since the JPSS, which contains 14 items, was sufficiently reliable and valid, the older participants were considered to have little ability to interpret the stress item of the VAS.

With regard to comparisons of the VAS, FS, and the PCS of SF-8, the correlation between the health status item of the FS and the PCS for the younger group ($r=0.455$) corresponded to that for the older group ($r=0.447$); however, the correlation between the health status item of the VAS and the PCS for the younger group ($r=0.353$) was lower than that for the older group ($r=0.522$). This result suggests that the item of the VAS may be valid for the older participants who could assess their own health status. However, the correlation between the stress item of the FS and the MCS for the younger group ($r=-0.527$) was comparatively close to that for the older group ($r=-0.450$), whereas the correlation between the stress item of the VAS and the MCS for the younger group ($r=-0.501$) contrasted with that for the older group, which was not significant. This finding corroborates the above result for the stress item of the VAS and the JPSS. Higher scores on the generic QOL and health status items of the VAS mean a better status for respondents, whereas higher scores on the stress item of the VAS means a worse status for them. This reversal of concepts about the items may have confused the older respondents.

This study has several limitations. One is that the participants were not selected by random sampling and thus may not be representative of the general population of Japanese adults. Moreover, the pool of participants was made up of people who were presumably interested in health and who were likely to be of a mind to volunteer for health surveys. Since socioeconomic data were not collected, the authors could not determine to what extent, if any, such factors might have influenced the results. In addition, the numbers of participants in the two age groups were not equal, with fewer participants in the older group. Despite these limitations, the differences in the validity of the scales between the younger and older adults were clear.

This study represents a step to understanding the value and use of computer-assisted assessment to measure the health of elderly people. Further investigation is needed to improve the PC scale for aged respondents. It is also important to establish easy-to-understand explanations for the design, purpose, and usage of the scales.

ACKNOWLEDGMENTS

We are grateful to the participants of this study and to the staff of the event to promote food and health. We also express our sincere thanks to Dr. Shigetoshi Iwahashi for valuable suggestions about the JPSS. This study was partly funded by the MOA Health Science Foundation.

REFERENCES

Aitken, R. C. B. 1969 Measurement of feelings using visual analogue scale. *Proceedings of the Royal Society of Medicine*, **62**, 989–993.

Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B. 1990 The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. *Pain*, **41**, 139–150.

Cohen, S., Kamarck, T., & Mermelstein, R. 1983 A global measure of perceived stress. *Journal of Health and Social Behavior*, **24**, 385–396.

Freyd, M. 1923 The graphic rating scale. *Journal of Education Psychology*, **14**, 83–102.

Fukuhara, S. & Suzukamo, Y. 2004 *Manual of the SF-8 Japanese version*. Kyoto: Institute for Health Outcomes & Process Evaluation Research. (In Japanese)

Hayes, S. & Patterson, D. G. 1921 Experimental development of the graphic rating method. *Psychological Bulletin*, **18**, 98–99.

Herr, K. A., Mobily, P. R., Kohout, F. J., & Wagenaar, D. 1998 Evaluation of the faces pain scale for use with the elderly. *Clinical Journal of Pain*, **14**, 29–38.

Iwahashi, S., Tanaka, Y., Fukudo, S., & Hongo, M. 2002 The development of the Japanese version of the Perceived Stress Scale. *Japanese Journal of Psychosomatic Medicine*, **42**, 459–466. (In Japanese with English abstract)

Joyce, C. R. B., Zutshi, D. W., Hrubes, V., & Mason, R. M. 1975 Comparison of fixed interval and visual analogue scales for rating chronic pain. *European Journal of Clinical Pharmacology*, **8**, 415–420.

Kimura, T., Hayashida, K., Araki, Y., Morita, T., Yamaguchi, N., & Eboshida, A. 2008 How valid is the self-administered visual analogue scale for assessing the health status of elderly people? *Hiroshima Journal of Medical Sciences*, **57**, 85–92.

Kleinman, L., Leidy, N. K., Crawley, J., Bonomi, A., & Schoenfeld, P. 2001 A comparative trial of paper-and-pencil versus computer administration of the Quality of Life in Reflux and Dyspepsia (QOL-RAD) questionnaire. *Medical Care*, **39**, 181–189.

Litaker, D. 2003 New technology in quality of life research: Are all computer-assisted approaches created equal? *Quality of Life Research*, **12**, 387–393.

Lorish, C. D. & Maisiak, R. 1986 The face scale: A brief, nonverbal method for assessing patient mood. *Arthritis and Rheumatism*, **29**, 906–909.

SPSS Inc. 2004 *SPSS 13.0J Brief Guide*. Chicago, Illinois: SPSS Incorporated.

Velikova, G., Wright, E. P., Smith, A. B., Cull, A., Gould, A., Forman, D., Perren, T., Stead, M., Brown, J., & Selby, P. J. 1999 Automated collection of quality-of-life data: A comparison of paper and computer touch-screen questionnaires. *Journal of Clinical Oncology*, **17**, 998–1007.

Ware, J. E., Kosinski, M., Dewey, J. E., & Gandek, B. 2001 *How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8™ Health Survey*. Lincoln, Rhode Island: QualityMetric Incorporated.

Ware, J. E. & Sherbourne, C. D. 1992 The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, **30**, 473–483.