



Original article

Scand J Work Environ Health [1999;25\(4\):361-367](#)

doi:10.5271/sjweh.446

Visual analogue scales for detection of changes in symptoms of the sick building syndrome in an intervention study

by [Kildesø J](#), [Wyon D](#), [Skov T](#), [Schneider T](#)

Key terms: [cleaning](#); [indoor air](#); [intervention](#); [questionnaire](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/10505662



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

Visual analogue scales for detecting changes in symptoms of the sick building syndrome in an intervention study

by Jan Kildesø, PhD,¹ David Wyon, PhD,² Torsten Skov, MD,³ Thomas Schneider, MSc¹

Kildesø J, Wyon D, Skov T, Schneider T. Visual analogue scales for detecting changes in symptoms of the sick building syndrome in an intervention study. *Scand J Work Environ Health* 1999;25(4):361—367.

Objectives This study tested questionnaires using visual analogue scales (VAS) in a cleaning intervention study and attempted to find a simple way of analyzing the replies to the questionnaires.

Methods A VAS questionnaire made up of 26 questions was developed and marked once a week for 28 weeks by the room occupants in 3 buildings. A total of 1248 questionnaires was used in the analysis of the results. A simple model based on the differences between a person's average responses during 2 different periods was used in the analysis.

Results No clear effect of the cleaning was found. Several significant correlations between different questions were established. Estimates for the design of future studies are given.

Conclusions The VAS questionnaire proved to be feasible for this type of study. It is suggested that each intervention period should last 4 weeks if the questionnaire is used once a week. However, the length of the period also depends on the expected latency of the symptoms, on how long it takes for environmental conditions to be affected by the intervention, and on how quickly conditions return to "normal" during control periods.

Key terms cleaning, indoor air, intervention, questionnaires.

Questionnaires using visual analogue scales offer the possibility of obtaining an indication of symptom intensity. On a linear scale, the person indicates how he or she is feeling at the moment by placing a mark between 2 statements concerning a specific condition. These questionnaires, giving a subjective evaluation of the person's condition, are easily administered by the person and quick to complete, and the result does not depend on the person's recollection of past symptoms.

Two employees of the Scott Paper Company (1) were apparently the first to record the use of a visual analogue scale (VAS). They reported high test-retest reliability ($r=0.76$) and high interrater reliability ($r=0.71$) when the VAS was used by superiors to rate subordinate workers, and they noted that the VAS freed the rater from using "direct quantitative terms" and allowed "as fine a discrimination of merit" as was desired. Freyd (2) reviewed this use of a VAS, which was known at the time as the "Graphic Rating Method". Other uses of the VAS approach began to appear in the literature with increasing frequency in the 1960s, initially for the assessment of environmental phenomena (3), but soon afterwards for

the assessment of pain (4). Wewers & Lowe (5) and Ahearn (6) provide useful reviews of the historical extension of the VAS approach to the assessment of mood, anxiety, alertness, craving for cigarettes, quality of sleep, and the like, but the VAS does not appear to have been used to assess the particular set of subclinical symptoms associated with the sick building syndrome until Wyon (7) performed intervention experiments in a Swedish hospital. Investigations concerning the sick building syndrome had previously been performed epidemiologically, using questionnaires to assess the frequency of occurrence of building-related symptoms over extended periods (8), rather than a VAS to assess current symptom intensity, which is clearly of more use in assessing the impact of short-term experimental intervention in occupied buildings (9). However, traditional questionnaire methods have been successfully used for daily and weekly symptom recordings in intervention studies (10, 11).

The characteristics of human responses obtained by the VAS have been well documented in connection with its use in clinical research. Bond et al (12) reported

¹ National Institute of Occupational Health, Copenhagen, Denmark.

² Johnson Controls Inc, Milwaukee, Wisconsin, United States.

³ Brønshøj Kirkevej, Brønshøj, Denmark.

Reprint requests to: Dr Jan Kildesø, National Institute of Occupational Health, Lersø Parkallé 105, DK-2100 Copenhagen, Denmark. [E-mail: jki@ami.dk]

significant correlations ($P < 0.001$) between VAS ratings of anxiety and other established measures, not only between subjects but also when anxiety was experimentally manipulated by the blinded administration of caffeine. Miller & Ferris (13) report that the reliability of VAS assessments is better when the now standard 100-mm scale is used than with a 50-mm scale and that acceptable validity was achieved for the assessment of mood and pain in a comparison with other indices and verbal scales. The validity of the VAS was also investigated by Wyon (7): VAS ratings relating to symptoms of the mouth, lips, throat, skin, nails, eyes and headaches in relation to the sick building syndrome were found to be corroborated by objective indicators of these symptoms, the indicators including measurements of tear-film stability and inter-blink interval, time to swallow, reported use of palliatives, and objective observations of each patient as recorded by a nurse. VAS assessments are sensitive to small changes (14), more sensitive than category scales for changes over time (15), and stable even when used by different ethnic or cultural groups (16).

Nyren (17) has discussed the difference between the unipolar VAS (eg, as used in pain assessment), typically ranging from absence of the symptom to some level that is defined as maximum intensity, and the bipolar VAS (eg, as used in mood assessment), ranging from "depression" at one extreme to "elation" at the other. The center of a bipolar scale may be taken by some subjects to

represent their own habitual state, while others may take it to represent the "normal" state (for the population). These may or may not differ in a subject's estimation, and therefore additional variance is introduced. Wewers & Lowe (5) regard bipolar scales of this kind as more difficult for subjects to use, although subjects experience little difficulty in using bipolar scales to assess environmental dimensions such as hot and cold or dry and humid, which have a well-understood neutral point.

The distribution of the marks from, for example, the left border (scale recordings) often departs from normality and thus a nonparametric analysis is usually to be preferred. Nyren (17) points out that VAS data consistently violate the parametric assumption of constancy of "measurement error". This stricture applies only to the "error" introduced by using subjective methods of assessment. Manual measurement of the distance of VAS markings from one end, using a micrometer, was compared by Huang et al (18) with transcription of the same forms using a digitizer pad. Both methods yielded excellent accuracy, but the use of the digitizer introduced fewer errors. A digitizer was used to transcribe the VAS data obtained in the present experiment.

One practical advantage of the VAS questionnaire is that all questions fit into 1 sheet of paper, and this feature makes it easier to administer. It is only necessary to specify the time when the questionnaire has to be filled out and point out that the answers have to reflect how the person is feeling at that particular moment. While traditional questionnaires concerning the sick building syndrome typically ask for the frequency of occurrence, at any intensity, on a 3-point scale (often, sometimes, never), a VAS questionnaire offers the possibility of giving a more accurate answer to a given question and of recording slight variations in response to the same question when presented with the questionnaire on a regular basis. Furthermore, it is an advantage that the questionnaire does not require a recall period.

The purpose of this study was to evaluate the feasibility of using the VAS questionnaire in cleaning intervention studies. The background for the present study was that relationships between the composition of dust in indoor office environments and the occurrence of symptoms of the sick building syndrome had been reported by, for example, Raw et al (11), Gyntelberg et al (19), and Sundell et al (20). The purpose was to study whether improved cleaning could reduce symptom intensity.

Some preliminary results of this study have been presented by Kildesø & Tornvig (21).

Subjects and methods

The questionnaire shown in figure 1 was used. It consists of 26 different questions concerning general

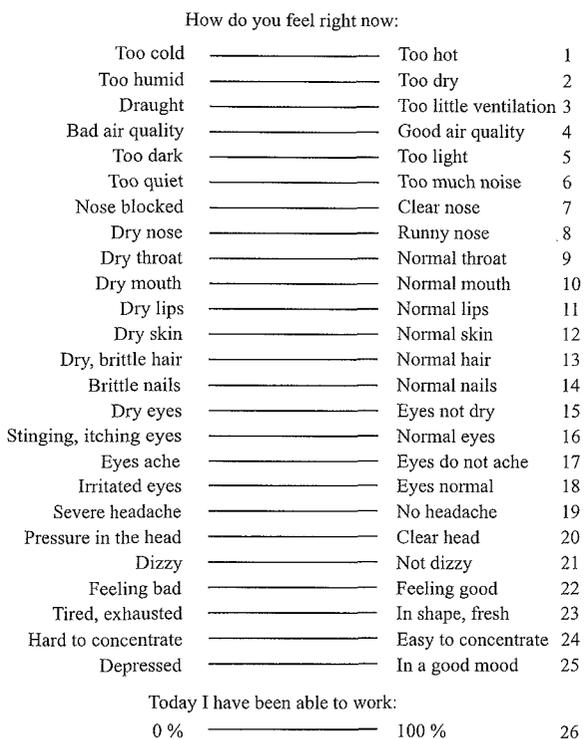


Figure 1. Questionnaire using a visual analogue scale. Not to scale. Each line has a length of 100 mm on the original questionnaire.

conceptions of indoor air quality and related parameters (questions 1–6), sensory irritation of the nose and throat (questions 7–10), sensory irritation of the eyes (questions 15–18), skin irritation (questions 11–14), and, finally, neurotoxic symptoms (questions 19–25). The neutral point differs for the various questions. The purpose of varying the neutral point was to make people consider their answer to each question carefully.

The cleaning intervention study was performed in 3 buildings operated by the municipality of Lyngby-Taarbæk north of Copenhagen, Denmark: a 6-story office building with a total floor area of 4800 m² built in 1973 with 240 employees, a kindergarten of 300 m² built in 1970 accommodating 10 employees and 53 children, and a school of 4500 m² built in 1958 with 22 teachers and 260 children. In the office building a ventilation system with heating and approximately 40% recirculation was in use. The floor coverings were tufted carpets in the administration building and linoleum in the school and the kindergarten. In the office building, intervention was performed on 2 floors, and 4 floors were used as a reference. For the school and kindergarten no reference was used due to limited resources. In each of the 3 buildings, 2 sampling rooms were selected for the measurement of dust.

The design of the intervention study and the measuring program has been described by Kildesø et al (22). In summary, the intervention was performed in different combinations: improvement of the vacuum cleaning by high-efficiency particulate-air (HEPA) filtering of the exhaust air and use of a brush mouthpiece, cleaning of the floors as the first step in the cleaning program, application of other cleaning agents, and cleaning of surfaces covered by papers, cables, and the like.

Every Friday, several people working in the buildings filled out a VAS questionnaire. The dust measurements and registration of other factors, such as activity, were also performed once a week but, for practical reasons, on Thursdays. It was found that the intervention had little effect on airborne dust levels, although several significant, but small effects on the levels of dust on surfaces could be shown (22).

The room occupants were asked to fill out the questionnaire half-way through their workday for practical reasons. They were instructed only to consider the way they felt at the time when they answered the questionnaire. Later, the marks on the questionnaires were transferred to a computer by means of a digitizer. A linear scale was used with a mark at the left border set at 0 and a mark at the right border set at 1. The value measured has been referred to as the scale recording in the following text.

All 22 teachers at the school were asked to fill out the VAS questionnaire at the end of the third teaching class Friday before leaving the classroom. In the

kindergarten, all 10 employees were asked to fill out the VAS questionnaire at the middle of the day on Friday. In the administration building, approximately 15 people from each of the 2 intervention floors and approximately 10 from each of the reference floors were asked to fill out the VAS questionnaire. People normally working in the dust sampling room on each floor were asked (2–3), and the rest were selected at random. People were asked to write the date, room number, and their name or some personal code on the questionnaires. Use of a personal code preserves confidentiality while permitting the use of each person as their own control.

The response rate of the questionnaires varied during the study period. Some people stopped answering, and some not initially asked in the administration building started to respond. Only employees who were initially asked and who had filled out at least 10 of the 28 questionnaires were included in the analysis. Therefore, approximately 20% of the questionnaires received were excluded.

Statistical methods

The Anderson-Darling normality test was used to test the data for normal distribution. The Kruskal-Wallis' and Mood's median nonparametric tests, both based on the median of the scale recordings, were applied for testing if the cleaning intervention had any effect on the responses to the questionnaires. The repeated measurements within an intervention period have been assumed to be repeated measurements of the same situation. The Spearman correlation matrix was calculated.

The VAS questionnaire was used to detect changes in the VAS responses between 2 periods (eg, a given intervention and the corresponding control). Thus a simple mathematical model could be used based on the difference between the mean scale recording for the individual person for 2 periods. This model can be applied for the evaluation of any of the 26 questions.

Let X_{ijk} denote the evaluation number k ($k = 1, 2, \dots, m_{ij}$) in period j ($j = 1, 2$) of person number i ($i = 1, 2, \dots, n$), where $E(X_{ijk}) = \mu_{ij}$ and $Var(X_{ijk}) = \sigma_{ij}^2$. σ_{ij}^2 is the variance for person number i in evaluating the symptom intensity during period number j , where conditions are assumed to be stationary. The change between the 2 periods for person number i is then $\delta_i = \mu_{i1} - \mu_{i2}$, which is estimated as

$$\hat{\delta}_i = \widehat{\mu}_{i1} - \widehat{\mu}_{i2} = \sum_{k=1}^{m_{i1}} \frac{X_{i1k}}{m_{i1}} - \sum_{k=1}^{m_{i2}} \frac{X_{i2k}}{m_{i2}}$$

This model is based on the assumption that the scale recording is an interval variable. People with only 1 or 0 response in 1 of the 2 intervention situations evaluated were excluded.

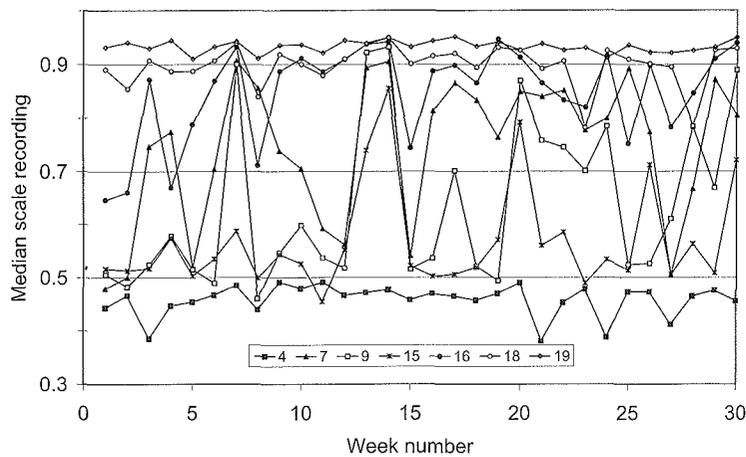


Figure 2. Mean values of the scale recordings for selected questions for all locations as a function of week number.

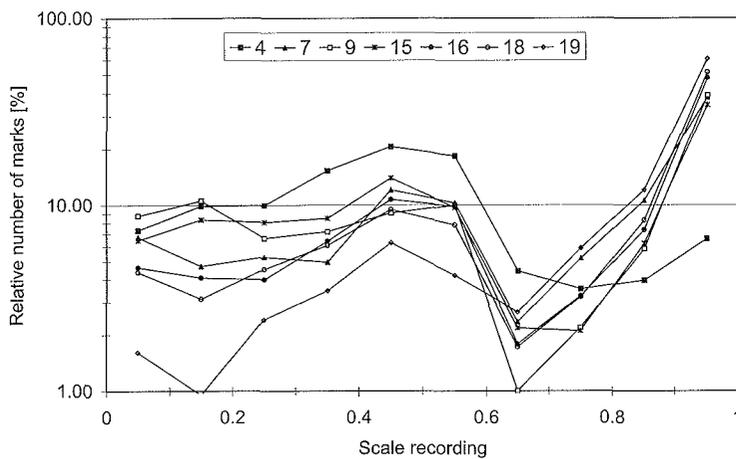


Figure 3. Distribution of the marks for different questions in the VAS questionnaire normalized to the percentage of the total number. Each dot plotted has been calculated from the number of marks within an interval of the scale.

questions. Figure 2 clearly illustrates the difference in variability over time for the questions. It appears that questions with a high variation as a function of time in figure 2 also tend to have a higher coefficient of variation (see table 2) (eg, questions 9 and 15). Some questions will obviously be influenced by, for example, changes in the weather and epidemics of common cold, while others obtain a very stable response (eg, question 19 on headache).

The tendency for people to place the marks in the questionnaire in the middle or close to one of the extremes results in the nonnormal distribution of the data, as illustrated in figure 3, which makes the application of nonparametric statistics necessary. However, the introduction of the model based on the difference of the means for each participating person turns out to be an efficient way for transforming the data set into normally distributed data.

The high correlation coefficients between responses to certain questions illustrates that the number of questions in the questionnaire could be reduced without losing too much information.

Table 3. Spearman correlation matrix for questions 4, 7, 9, 15, 16, 18, and 19.

	4	7	9	15	16	18
7	0.216					
9	0.379	0.476				
15	0.341	0.437	0.578			
16	0.320	0.534	0.583	0.715		
18	0.293	0.541	0.576	0.704	0.843	
19	0.304	0.397	0.412	0.362	0.555	0.532

The cleaning intervention did not affect the symptom intensity of the sick building syndrome of the room occupants, as measured by VAS questionnaires. This finding may not be surprising, as the effects of the cleaning intervention on dust levels, though significant, were small (22). A correlation, if any, between the symptoms of the sick building syndrome and the dust measurements may also have been reduced because there was a time difference between the dust sampling and the marking of the VAS.

An estimate of the dimensioning of future studies based on the variability in this study can be found by the

model based on the differences between the mean values of the scale markings in 2 different intervention periods (1 of them could be the control). As these differences were found to be normally distributed, the equation

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \sigma^2}{\delta^2}$$

can be used for finding the necessary number of persons *n* in each intervention (23). σ^2 is the standard deviation of the differences and δ is the average difference that needs to be detected for all people with the power $1 - \beta$ (being the probability of detecting an existing difference between 2 periods) and the level of significance α . $Z_{1-\alpha}$ and $Z_{1-\beta}$ are the percentiles in the standard normal distribution. Questions 4, 7, and 18 have been used as examples of estimations of the number of people required in future studies, depending on the duration of the study, to detect a given difference between 2 periods (eg, intervention and control). In the estimate of σ , it has been assumed that there are no systematic differences in the scale recordings. The variability is caused by factors that are not controlled in the experiment. Thus σ has been estimated for δ_i calculated for 2 consecutive periods of equal length T (T = 4, 6, 8, 14 weeks), starting from the beginning of the study period. T=14 weeks covers the entire study. The result is given in table 4 for a level of significance of $\alpha = 0.05$ and a power of $1 - \beta = 0.8$.

Table 4. Estimates for the design of future studies based on questions 4, 7 and 18. For different numbers of weeks in each of 2 periods (eg, intervention and control), the number of persons fulfilling the criteria of >1 response in each period is given. The total number of questionnaires on which each estimate has been based is also given. The necessary number of people in the study has been estimated for three different values of average change in the scale recordings δ which the study must be able to detect.

	Number of persons			
	σ^2	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$
14 weeks (70 persons, 1220 questionnaires)				
Question 4	0.124	614	153	38
Question 7	0.147	726	181	45
Question 18	0.149	740	185	46
8 weeks (67 persons, 675 questionnaires)				
Question 4	0.120	594	148	37
Question 7	0.157	775	194	48
Question 18	0.147	727	182	45
6 weeks (64 persons, 513 questionnaires)				
Question 4	0.142	703	176	44
Question 7	0.191	947	237	59
Question 18	0.150	741	185	46
4 weeks (51 persons, 312 questionnaires)				
Question 4	0.183	906	226	57
Question 7	0.212	1050	262	66
Question 18	0.174	861	215	54

It was found that, for periods (intervention and control) shorter than 4 weeks, more than half of the people will be excluded as a minimum of 2 responses in each period is required to include a person in the analysis. As table 4 indicates, there is no benefit in using intervention periods longer than 8 weeks. A likely explanation is that, as the period increases, more variability caused by uncontrolled factors will be introduced. Table 4 also indicates that the study design should take into account the expected ratio of missing data.

The estimates given in table 4 vary for other types of questions, but the approximate levels will not change considerably. It is of course important to note that the optimum length of an intervention period is not just a function of the statistical power of the VAS approach. It also depends on the expected latency of the symptoms (how quickly they appear and disappear), on how long it takes for environmental conditions (eg, dust) to be affected by the intervention, and on how quickly conditions (eg, dust levels) return to "normal" during control periods.

This study shows that, even though no effects of the cleaning intervention were found in the VAS responses, these questionnaires may be an expedient alternative to traditional questionnaires, for which repeated measurements are relevant and no recall time is required.

References

- Hayes MHS, Patterson DG Experimental development of the graphical rating method. *Psychol Bull* 1921;18:98—9.
- Freyd M. The graphic rating scale. *J Educ Psychol* 1923;14:83—102.
- Aitken RCB, Ferres HM, Gedye JL. Distraction from flashing lights. *Aerosp Med* 1963;34:302—6.
- Bond MR, Pilowsky I. Subjective assessment of pain and its relationship to the administration of analgesics in patients with advanced cancer. *J Psychosom Res* 1966;10:203—8.
- Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health* 1990;13(4):227—36.
- Ahearn EP. The use of visual analogue scales in mood disorders: a critical review. *J Psychiatr Res* 1997;31(5):569—79.
- Wyon DP. Sick buildings and the experimental approach. *Environ Technol* 1992;13(4):313—22.
- Andersson K. Epidemiological approach to indoor air problems. *Indoor Air* 1998;suppl 4:32—9.
- Wyon DP. Symptom Intensity Feedback Testing (SIFT): behavioural science may be able to provide the key to curing sick buildings. In: Banhidi L, Farkas I, Magyar Z, Rudnai P, editors. *Proceedings of healthy buildings '94*. Budapest: Technical University; vol 3. 1994:42—7.
- Reinikainen LM, Jaakkola JJ, Seppanen O. The effect of air humidification on symptoms and perception of indoor air quality in office workers: sixperiod cross-over trial. *Arch Environ Health* 1992;47(1):8—15.
- Raw GJ, Roms MS, Whitehead C. Sick Building Syndrome:

- cleanliness is next to healthiness. *Indoor Air* 1993;3:237—245.
12. Bond AJ, Shine P, Bruce M. Validation of visual analogue scales in anxiety. *Int J Method Psychiatr Res* 1995;5(1):1—9.
 13. Miller MD, Ferris DG. Measurement of subjective phenomena in primary care research: the visual analogue scale. *Fam Pract Res J* 1993;13(1):15—24.
 14. Maxwell C. Sensitivity and accuracy of the visual-analogue scale: a psycho-physical classroom experiment. *Br J Clin Pharmacol* 1978;6:15—24.
 15. Pfenning L, Cohen L, van der Ploeg H. Preconditions for sensitivity in measuring change: visual analogue scales compared to ratings scales in a Likert format. *Psychol Rep* 1995;77(2):475—80.
 16. Gaston-Johansson F, Albert M, Fagan E, Zimmerman L. Similarities in pain descriptions of four different ethnic-culture groups. *J Pain Symptom Manage* 1990;15(2):94—9.
 17. Nyren O. Visual analogue scale. In: Hersen M, Bellack AS, editors. *Dictionary of behavioral assessment techniques*. New York (NY): Pergamon 1988.
 18. Huang HY, Wilkie DJ, Berry DL. Use of a computerised digitizer tablet to score and enter visual analogue scale data. *Nurs Res* 1996;45(6):370—2.
 19. Gyntelberg F, Suadicani P, Nielsen JW, Skov P, Valbjørn O, Nielsen PA, et al. Dust and the sick building syndrome. *Indoor Air* 1994;4:223—38.
 20. Sundell J, Lindvall T, Stenberg B, Wall S. Sick building syndrome (SBS) in office workers and facial skin symptoms among VDT-workers in relation to building and room characteristics: two case-referent studies. *Indoor Air* 1994;4:83—94.
 21. Kildesø J, Tornvig L. The effect of improved cleaning methods on the composition of dust and the well-being of people. In: Yoshizawa S, Kimura K, Ikeda K, Tanabe S, Iwata T, editors. *The proceedings of the 7th international conference on indoor air quality and climate — indoor air '96*. Nagoya, Japan, 21—26 July; vol 2. 1996:923—8.
 22. Kildesø J, Tornvig L, Skov P, Schneider T. An intervention study of the effect of improved cleaning methods on dust concentration and on dust composition. *Indoor Air* 1998;8:12—22.
 23. Kleinbaum DG, Kupper LL, Muller KE. *Applied regression analysis and other multivariable methods*. 2nd ed. Belmont (CA): Duxbury Press, 1988.

Received for publication: 13 July 1998