

## The Social Interview Schedule (SIS) – content, structure and reliability

Toni Faltermaier, Hans-Ulrich Wittchen, Rosemary Ellmann, and Reinhold Lüssle

Max Planck Institute of Psychiatry, Munich, Federal Republic of Germany

**Summary.** The English original form of a standardized social interview was presented by Clare and Cairns (1978). The German version is described with regard to concept, structure, and methodology. The interview is designed to assess social maladjustment by measuring three conceptual categories (“Objective Material Conditions,” “Social Management,” and “Satisfaction”) in eight role areas. The instrument is especially suited for outcome assessment in various clinical and non-clinical populations and takes an average of about 30–45 min to complete. The 39 items are rated on 4-point scales either by the interviewer with the help of an extensive rating manual, or by the subject himself (all “Satisfaction” items). The results of a reliability study are presented using both the test-retest and the interrater method. In general, the results show that the instrument is satisfactorily reliable, but some possible weaknesses are discussed. Finally, the structure of the instrument is examined by an analysis of interrelationships between the items in a normal population sample. The results are discussed with regard to questions of score construction.

---

In the realm of epidemiological psychiatry and outcome research, there is increasing interest in assessing not only psychiatric symptoms, but also aspects of social functioning or adjustment as distinct components of an individual's behavior. The value of making a distinction between symptoms and social adjustment and of an independent assessment of each is stressed by several investigators (Frank 1961; Williams et al. 1980; Wing 1981) and supported by a number of empirical findings (Freeman and Simons 1963; Cooper et al. 1970; Paykel and Weissman

1973; Kedward and Sylph 1974; Weissman et al. 1974; Tanner et al. 1975; Paykel et al. 1978; Mann et al. 1981). This distinction is not only of interest for outcome measurement but could also have quite practical implications for therapy. The question arises, for example, of whether one needs different interventions for social adjustment problems, on the one hand, and for psychiatric symptoms on the other. A number of instruments have been developed to evaluate the social adjustment of functioning of psychiatric (or other) patients (e.g., after discharge or after treatment programs), independently of the presence or absence of psychiatric symptoms. Reviews of assessment instruments (see Weissman 1975; Weissman et al. 1981; Donald et al. 1978) show that investigators have developed quite different approaches in conceptualizing and measuring social functioning and adjustment. For the purpose of assessment of social functioning as an outcome variable in a follow-up study of former psychiatric inpatients and a normal population sample (Wittchen et al. 1983; Faltermaier 1983), we chose to use the Social Interview Schedule (SIS).

### *Aims*

(1) This paper describes the German version of the SIS<sup>1</sup>, a standardized interview for assessment of social maladjustment and dysfunction, originally developed by Sylph and her colleagues and subsequently revised by Clare and Cairns (1978). The schedule can be used in various clinical populations (though not in the case of chronically hospitalized patients), although the English version has been applied primarily to general practice and social work

---

<sup>1</sup> The English version has recently been named “Social Maladjustment Schedule” (see Corney et al. 1982)

settings. We shall describe the concept, structure, and methodology of this instrument, noting also the modifications introduced into the German version.

(2) Furthermore, this paper presents the results of a reliability study using the German version. Instead of measuring the reliability in terms of interrater agreement alone (which was reported by Clare and Cairns (1978) to be generally high), we also included the test-retest reliability between different interviewers as a more adequate method.

(3) As proposed by Clare and Cairns (1978), the items of the SIS can be grouped into various categories (see below). An analysis of the interrelationships within and between these proposed categories was made in an attempt to confirm this categorization empirically. This is especially important with regard to scoring possibilities.

### Description of the SIS<sup>2</sup>

The SIS is a structured interview for assessing different aspects of social maladjustment and dysfunction. As in most other studies, adjustment is seen as an interplay between the individual and his social environment. The conceptual framework for assessment is role analysis. According to this concept, social adjustment can be defined as "the degree to which a person fulfils the normative social expectations of behavior that constitute his roles" (Barabee et al. 1955, p. 252). Most of the existing social adjustment instruments refer to this or similar definitions as their conceptual basis, but it has been questioned whether they really manage to operationalize what they claim to measure. (For a critical discussion of this topic, see Platt 1981.)

The role concept generally implies measurement of a person's "instrumental role performance" (Parsons and Bales 1954), his behavior in roles or in the number of socially recognized positions he occupies. In the SIS this aspect of social adjustment is called "*Social Management*" (M) and covers the social competence of the individual and his success in acting out different roles (see Table 1).

Another aspect of functioning, often included in more recent studies, the individual's "expressive" functioning (Parsons and Bales 1954), refers to the degree of satisfaction an individual is able to obtain from acting in his social roles. In the SIS this aspect is called "*Satisfaction*" (S) and includes the individual's subjective evaluation of the different aspects of his life situation, as represented by the different roles he fulfils.

Finally, unlike most other instruments, the SIS measures a third category in each role area: "*Objective Material Conditions*" (O). This is intended to serve as a kind of objective social framework against which the individual's functioning and satisfaction can be assessed; it provides information about the situation to which an individual has to adjust.

O, M, and S are assumed to be relatively independent factors contributing to the overall adjustment and role strain of an individual. The categories M and S, however, are not considered to be constant traits of an individual, but each is intended to describe similar aspects of his acting in different social roles. This three-way categorization in objective conditions for role acting, social management of the role demands, and satisfaction experienced within each role area is a unique characteristic of the SIS. Table 1 illustrates this structure by showing the distribution of the items across the different role areas and three rating categories. There are a total of 39 items in this form of the SIS, compared with 42 items in the original form of Clare and Cairns. Five new items were included in the modified version (marked by asterisks in Table 1), while eight items have been excluded in this version for various reasons, especially the need to reduce the length of the interview.

The SIS is aimed at assessing the current social situation of an individual: all questions refer to the respondent's typical situation over a period of 1 month before the interview. The trained interviewer is to ask certain standard questions for each item, but beyond that he is free to ask additional questions as appropriate. The standard questions are formulated in the interview schedule, additional "probes" being formulated as catchwords to remind the interviewer of information necessary for the rating. The interviewer must make written notes while interviewing because the rating procedure itself is too complicated to carry out during the interview. It takes an average of about 30-45 min to complete the interview.

As soon as possible after the interview, the interviewer should rate the recorded information with the help of an extensive rating manual, allowing about 30 min. Each item is rated on a 4-point scale; the scale point definitions for O and M items are as follows:

### Objective conditions (O)

- 1 = no restrictions/stressful conditions
- 2 = minor restrictions/stressful conditions
- 3 = marked restrictions/stressful conditions
- 4 = severe restrictions/stressful conditions

<sup>2</sup> For a more extensive conceptual and methodological discussion and description of the SIS, see Clare and Cairns (1978)

**Table 1.** Distribution of items rated on the SIS by rating category and role area

Role area	Rating category		
	O (Objective material conditions)	M (Social management)	S (Satisfaction)
Housing	Housing conditions	-	Satisfaction with housing
Occupation/housework/ other social roles	Work conditions <sup>a</sup>	Management of work/study <sup>a</sup>	Satisfaction with work/study
	Study conditions <sup>a</sup>		
	Opportunities for interaction with workcolleagues	Quality of personal interaction with workcolleagues	Satisfaction with personal interaction with workcolleagues
	Housework conditions <sup>a</sup>	Management of housework	Satisfaction with housework situation
	-	-	Satisfaction with other social role (includes unemployed, disabled, retired)
Economic situation	Household income	Management of income	Satisfaction with income
Leisure/social activities	Opportunities for leisure and social activities	Extent of leisure activities	Satisfaction with leisure activities
		Extent of social activities	Satisfaction with social activities
Interaction with relatives	Opportunities for interaction with relatives	Quality of interaction with relatives	Satisfaction with interaction with relatives
Domestic interaction	Opportunities for domestic interaction (with other adult household members)	Quality of domestic interaction (with other adult household members)	Satisfaction with domestic interaction (with other adult household members)
Living alone	-	Management of living alone	Satisfaction with living alone
Marital/partnership	-	Sharing of interests and activities	Satisfaction with marital relationship/partnership
	-	Sharing of responsibilities and decision-making	Satisfaction with sexual compatibility
	-	-	Satisfaction with having no partner <sup>a</sup>
	-	-	-
Parental/child management	Situational handicaps to child management	Child management	Satisfaction with parental role

<sup>a</sup> New items in the German SIS version

### Social management (M)

- 1 = no difficulties
- 2 = minor difficulties
- 3 = marked difficulties
- 4 = severe difficulties

The Rating Manual<sup>3</sup> defines rating rules and gives examples to illustrate typical situations for each point in each item's scale. In this manner each rating point is described as precisely as possible. Unlike Clare and Cairns (1978), we used a different procedure to rate the subjective aspects of functioning. Satisfaction, a clearly subjective measure, is rated by the subject himself with respect to each of the different roles he fulfils, in each instance in answer to a standard question and according to the following scheme:

<sup>3</sup> A revised version of the German SIS manual is now available (Faltermaier 1982) and can be obtained upon request from the first author

### Satisfaction (S)

- 1 = very satisfied
- 2 = satisfied
- 3 = dissatisfied
- 4 = very dissatisfied

The S scale measures variations in both a desirable and an undesirable direction, while the O and M categories are rated only in the direction of poor conditions or adjustment.

### Methods

#### 1. Assessment of reliability

In psychiatric research, reliability is often assessed by simply measuring *interrater reliability*, i.e., the agreement between two raters judging the same interview. Usually, one rater is the interviewer while the other is either present during the interview or

judges a videotape of it. This procedure, however, excludes important sources of variance, as both raters have the same information on which to base their ratings. The hereby neglected process of obtaining information is a main source of variance and deserves more attention. A more adequate but also more complicated and time-consuming method is to measure the agreement between two different raters making independent interviews with the same respondent (*test-retest reliability*). Both of these methods were used in the present study.

Each respondent was interviewed twice, each time by a different interviewer; the time lapse between the two interviews was in most cases 1 or 2 days. The sample consisted of 20 subjects, 10 inpatients of a psychiatric hospital, and 10 subjects without a psychiatric disorder. There was a total of three interviewers: one interviewer saw all 20 subjects; the two other interviewers each saw 10 different subjects. For our purposes here the latter two can be taken together as "one" interviewer seeing all 20 subjects. In this manner, the ratings for all subjects can be compared between the two interviewers. All interviews were recorded on videotapes and judged by a total of four raters. Each rater independently observed 20 videotapes: 10 of the first interviewer, 10 of the second one. Consequently, two other methods of assessing reliability were possible: the rating scores of each rater could be compared with those of the interviewer. This resulted in an additional four comparisons (rater - interviewer). Each rater could also be compared with each of the other raters in their ratings of all 20 subjects; this resulted in another six comparisons (rater-rater).

Altogether, three types of reliability assessment are reported: the test-retest reliability as the most important and, in addition, two ways of measuring inter rater reliability (rater - interviewer and rater - rater comparisons). The interviewers and raters were either clinical psychologists (4) or psychiatrists (3) of the Max Planck Institute of Psychiatry, all previously trained in use of the SIS.

We used the following *coefficients of agreement* as statistical measures of reliability (see Bartko and Carpenter 1976):

Per cent agreement. This measure is very often used, although it includes chance agreement. Thus, it has to be interpreted with caution. We chose it for reasons of comparison with other studies and because it is an easily interpretable measure, when taking into account the above restriction.

Intraclass correlation coefficient (ICC; Bartko 1966). This measure is applicable to data with interval quality. It is based on a one-way analysis of variance and can be interpreted as a correlation coefficient (Bartko 1966). It has a value of 1.0 when agreement is complete and of 0 when there is no agreement.

## 2. Analyses of interrelationships

The three main categories of the SIS ("Objective Conditions," "Social Management," and "Satisfaction"), as well as the grouping by role areas, are based on a theoretical concept. A common empirical approach to derive such categories is factor analysis.

However, this approach has some serious limitations when applied post hoc to existing social adjustment scales (see Clare and Cairns 1978). There is, for example, always a considerable proportion of items which have to be excluded from the analysis because they do not apply to the majority of subjects. In addition, there is the difficulty of choosing an appropriate factor analytic method to derive a meaningful number of factors. Factor solutions are also dependent on sample characteristics. Furthermore, in this study a factor analytic approach would seem inappropriate as we have not regarded a high proportion of common variance as a prerequisite for aggregating items.

An alternative approach, which we consider more appropriate here, would be first to accept the present grouping into three categories of adjustment and into the various role areas as theoretically meaningful and then to test if this classification is in accordance with empirical analyses of interrelationships between items. We would expect that relations between items of different categories (O, M, and S) or of different role areas are relatively small, but not necessarily absent. Bias in the scoring system can occur if some of the items which are aggregated measure the same aspect, at least to a certain degree, which would be reflected in a large proportion of common variance, while other items are relatively heterogeneous. In each aggregation score the overlapping items would be weighted higher than items that measure more independent aspects.

For the analyses of interrelationships we used data from a sample of the general population ( $N=480$ ). The sample was selected to be approximately representative of the adult population in the Federal Republic of Germany (age: 26-63; minimum IQ: 80; for further details of the sampling procedure see Wittchen and v. Zerssen, in preparation). According to a special selection process, however, it can be assumed that there is a slight overrepresentation of subjects with psychiatric symptoms. Of the sample 51% are females and 49% males; 55% are between 26 and 45 years old; 45% are older than 45 years. With regard to social class, 64% of the sample can be grouped into the upper or middle class (categories I-III of the Moore-Kleining Index) and 36% into the lower class (IV-V) (Moore and Kleining 1960). The lower class is thus somewhat underrepresented in the sample<sup>4</sup>.

Two stages of analysis have been undertaken:

(1) To analyze the strength of the *relationships between the three categories* O, M, and S, we examined for each role area the relationships between the three (or two) items representing the three main categories, e. g., for the work area, the relationships between the items "work conditions" (O), "management of work" (M), and "satisfaction with work" (S).

<sup>4</sup> The class distribution of the population in the FRG is as follows: 59% are in the upper or middle class (I-III) and 41% in the lower class (IV-V) (Kleining 1975)

(2) The strength of the *relationships between different role areas* was analyzed for each of the three categories, O, M, and S, by examining the relationships between all items falling into that category and representing the different role areas, e.g., for the category O (“Objective Conditions”), the relations between the items “work conditions” (O), “opportunities for interaction with workcolleagues” (O), “housework conditions” (O), “household income” (O), and so on (see Table 1 for the items in each category).

Log-linear analysis (Everitt 1977; Goodman 1978) was used to determine whether there is a relationship between the variables, and product moment correlations were used to assess the strength of the relationships<sup>5</sup>.

## Results

### 1. Reliability

One way to analyze SIS data is by the calculation of scores. While we will discuss the problems and possibilities of scoring later in detail, for the purposes of reliability analysis we used the relatively simple technique, as proposed by Clare and Cairns (1978), of counting the number of ratings of 3 or 4 (marked or severe difficulties, restrictions, etc.) for each subject, first for all items and then separately for the items of each of the categories O, M, and S. Thus four scores for each subject were obtained: a total score and three category scores (O score, M score, and S score).

Table 2 shows ICC values for the agreement between the two interviewers (I-I), between the interviewers and each of the 4 raters (I-R) (one value for each rater combined over all interviewers), and between each rater and each other rater (R-R).

The test - retest reliability (I-I) shows a good agreement in the total score with an ICC of 0.896. Similarly, the agreement coefficients in the category scores are high, with the exception of the category “Satisfaction” (0.638). The interrater reliability measures are also generally high, with the exception of the category “Management”, both for I-R and R-R comparisons. One rater has an especially low ICC value for I-R comparisons in the M category; he is probably to a large degree responsible for the low ICC M-value in the interrater comparison (R-R) as well.

Table 2 also shows the means of the overall percent agreement, calculated over all items and then

**Table 2.** Intraclass correlation coefficients (ICC) for agreement between the global scores of two interviewers (I-I), between the two interviewers and four raters (I-R and between the four raters (R-R). Means of percent agreement for different item groupings

Agreement coefficients	I-I	I-R <sup>a</sup>	R-R	
ICC for total score	0.896	0.880	0.875	0.895
Mean percent agreement for all items:				
- 4-point scale	78% (33)	81% (25)	81%	(32)
- dichotomized scale	90% (33)	93% (25)	92%	(32)
ICC for score O	0.888	0.761	0.838	0.823
Mean percent agreement for all O-items:		0.895	0.897	
- 4-point scale	84% (9)	71% (7)	73%	(9)
- dichotomized scale	90% (9)	92% (7)	89%	(9)
ICC for score M	0.897	0.147	0.537	0.310
Mean percent agreement for all M-items:		0.582	0.541	
- 4-point scale	82% (11)	69% (8)	68%	(11)
- dichotomized scale	91% (11)	81% (8)	87%	(11)
ICC for score S	0.638	0.994	1.000	0.981
Mean percent agreement for all S-items:		0.967	1.000	
- 4-point scale	70% (13)	99% (10)	98%	(12)
- dichotomized scale	89% (13)	100% (10)	99%	(12)

Parentheses: number of items on which calculation is based

<sup>a</sup> ICC values had to be calculated in pairs to treat the interviewer not the same as each of the raters. Thus, there are four ICC values in each instance

separately over the items of each of the categories O, M, and S. They are given both for the original 4-point scales and for the dichotomized scales (1/2 versus 3/4). The total value is satisfactory with a mean percent agreement between the two interviewers of 78% for the 4-point scales and 90% for the dichotomized scales. The percent agreement is high for the categories O and M, and lower for the category S using the 4-point scales, but just as high when using the dichotomized scales.

### 2. Interrelationships

(a) *Relationships between categories:* First, for each role area (or aspect of a role area) the associations between the O, M, and S item scores were tested

<sup>5</sup> Product moment correlations were calculated both for the original 4-point scales and for dichotomized scales. In this presentation we will concentrate on results based on the dichotomized scales. There was, however, in no case a marked difference from calculations based on the 4-point scales

**Table 3.** Measures of association between the O, M, and S item scores within the different role areas: Significant log-linear relationships, product moment correlation coefficients and percent of common variance (based on dichotomized scales)

Role area	Related categories	Log-linear analysis	Product moment correlation	Common variance (%)
Housing	O-S	+	0.23	5.29
Occupation	M-S	+	0.25	6.25
Interaction with workcolleagues	O-S	+	0.16	2.56
Interaction with workcolleagues	M-S	+	0.21	4.41
Housework	M-S	+	0.32	10.24
Income	M-S	+	0.20	4.00
Leisure	O-S	+	0.25	6.25
Leisure	M-S	+	0.24	5.76
Social contacts	M-S	+	0.35	12.25
Relatives	M-S	+	0.41	16.81
Domestic interaction	O-S	+	0.38	14.44
Domestic interaction	M-S	+	0.50	25.00
Living alone	M-S	+	0.60	36.00
Partnership (interests)	M-S	+	0.47	22.09
Partnership (decisions)	M-S	+	0.48	23.04
Children	O-M	+	0.25	6.25
Children	M-S	+	0.51	26.01

+ significant relationship ( $p < 0.05$ )

**Table 4.** Product moment correlation (for the dichotomized scales) between items of the categories "Objective Conditions" (O), "Management" (M), and "Satisfaction" (S) over all subjects in a normal population sample

	Objective	Management	Satisfaction
Objective	1.00 (3603)		
Management	0.13 (3092)	1.00 (4231)	
Satisfaction	0.17 (3529)	0.33 (4192)	1.00 (4695)

Common variance between categories:  
 O-M: 1.69%  
 O-S: 2.89%  
 M-S: 10.89%

Parentheses: number of cases  $\times$  number of role areas compared

(sometimes only two categories were comparable). Table 3 shows all significant associations in the log-linear analyses ( $p < 0.05$ ), the product moment correlations for the dichotomized scales, and the percentage of common variance.

As can be seen, significant associations were found most frequently between the categories "Management" (M) and "Satisfaction" (S), rarely between the categories "Objective Conditions" (O) and "Satisfaction" (S), and in only one case between "Objective Conditions" (O) and "Management" (M). In no case could an interaction effect between the three variables be found. The associations are generally

rather weak. With one exception ("domestic interaction": O-S), only relationships between M and S variables exceed 10% common variance; in five areas the common variance lies above 20%. The strongest association was found between the categories M and S in the area "living alone," with a common variance of 36%.

The associations between the categories O and M, as well as between O and S, are significant in only a few areas. When a product moment correlation is calculated over all persons and all role areas (see Table 4), it can be seen that the general overlap between these categories is minimal. The association between the categories M and S is modest and reflects only about 10% common variance.

(b) *Relationships within categories:* Next, the relationships between the items representing various role areas within each of the categories O, M, and S were tested. In Table 5 one sees all the associations between role areas, which are either significant in log-linear analysis ( $p < 0.05$ ) or have common variance of at least 10%, as measured by product moment correlation of the dichotomized scales.

In the *category O*, there are three significant relationships between item pairs. The strongest relationship exists between "housework conditions" and "opportunities for domestic interaction" with 34% common variance. The item "housework conditions" is also significantly related to the item "opportunities for leisure and social activities." Here, however, one has to consider that there is a significant interaction between these last two variables and the variable "work conditions." This interaction points to a complex relationship between the three variables, which cannot yet be clarified.

In the *category M*, five associations are significant, most of them having a rather large amount of common variance (about 30%).

Most significant associations appear in the *category S*. However, the common variance here rarely exceeds 10%<sup>6</sup>.

## Discussion

### 1. Reliability

In general, the reliability of the "Social Interview Schedule" can be regarded as satisfactory. Few instruments for the assessment of social adjustment

<sup>6</sup> Two item pairs show a discrepancy between a relatively high product moment correlation and a non-significant log-linear analysis; this is probably an artifact due to the relatively low number of cases for these pairs, which could be expected to have a greater effect on the log-linear analysis

**Table 5.** Measures of association between various role areas within each of the categories O, M, and S; significance in log-linear analysis (+ :  $p < 0.05$ ) or common variance of more than 10%, product moment correlation coefficients for dichotomized scales

Category	Related role areas	Log-linear analysis	Product moment correlation $r$	Common variance (%)
O	Work - leisure/contacts	-	0.32	10.24
	Housework - leisure/contacts	+	0.44	19.36
	Housework - domestic interaction	+	0.58	33.64
	Leisure/contacts - children	+	0.42	17.64
	Work - housework - leisure/contacts	+	-	-
M	Work - income	+	0.39	15.21
	Interaction with workcolleagues - housework	+	0.24	5.76
	Leisure - living alone	+	0.52	27.04
	Relatives - Living alone	+	0.55	30.25
	Partner interests - partner decisions	+	0.56	31.36
S	Work - leisure	+	0.19	3.61
	Work - housing	+	0.21	4.41
	Work - living alone	-	0.45	20.25
	Housework - leisure	+	0.28	7.84
	Housework - social contacts	+	0.27	7.29
	Leisure - social contacts	+	0.20	4.00
	Leisure - housing	+	0.13	1.69
	Leisure - living alone	-	0.42	17.64
	Leisure - partner	+	0.17	2.89
	Social contacts - partner	+	0.18	3.24

have been examined using the *test-retest reliability* method. Only one recently developed interview schedule, the "Social Functioning Schedule" (Remington and Tyrer 1979), offers results which can be directly compared with our reliability data. It was tested by a similar method (independent interviews conducted on the same day) and the same statistical measure (intraclass correlation coefficient). This schedule showed an overall ICC value of 0.62; in the different subsections, the ICC values ranged from 0.45 to 0.81. A somewhat dissimilar instrument, the "Social Stress and Functioning Inventory for Psychotic Disorders" (Serban 1978), was applied twice to the same subjects within a 6-month interval. The test-retest correlations (type unspecified) ranged from 0.43 to 0.77. Compared to these results, the SIS test-retest agreement appears remarkably good.

Surprisingly, our test-retest measures tend to be even higher than the interrater measures, at least for the categories "Objective Conditions" (O) and "Social Management" (M). This finding was contrary to expectations, as the interrater measures exclude the variance inherent in the interviewing process and thus in theory should produce higher agreement coefficients.

With regard to the lower test-retest values for the category "Satisfaction" (S), it should be noted that the discrepancies are not serious. If one looks at the mean percent of overall agreement for all S-items (Table 2), the value for the dichotomized scales is much higher than the value for the 4-point scales and

about the same as the values in the categories O and M. This means that variations between the independent interviews seldom occur from levels of satisfaction (1/2) to levels of dissatisfaction (3/4), or vice versa; variations are in most cases of relatively minor quality (e.g., from "very satisfied" to "satisfied").

Two reasons could be mainly responsible for discrepancies in the category S between test and retest. There could, on the one hand, be some variation in his actual degree of satisfaction between the first and the second interview. It could be that the subjective assessment of satisfaction was somewhat unstable in this sample and that at least a few people felt differently about the same situation a few days later. On the other hand, the variation in the subject's ratings could also be attributed to the measurement procedure and the interview situation, for example, to the form of questioning or the interviewer - subject interaction processes. It is impossible to decide which reason was more important in this study. One of the possible sources of variability, however, can be ruled out in future by paying more attention to the form of questioning. Although the questions to be asked were indeed already standardized, the taped interviews demonstrated that interviewers sometimes tended to modify the questions slightly, influenced by the relatively free interview style appropriate for the categories O and M. This sometimes resulted in suggestive formulations because the interviewer was familiar with the situation in question and probably had his own opinion about it. Consequently, this

point will be particularly stressed in the revised interviewer manual and in future interviewer training: all questions referring to the respondent's degree of satisfaction are to be asked precisely in the reformulated way.

Our results on *interrater agreement* are generally comparable to those of other social interview schedules (e.g., Clare and Cairns 1978; Remington and Tyrer 1979; Platt et al. 1980; Paykel et al. 1971; Gurland et al. 1972; Serban 1978). In the category O, the agreement coefficients are satisfactory, particularly when considering the dichotomized scale results. In the category M, however, coefficients proved to be generally lower, and unsatisfactory for at least one rater. The reasons for this are not quite clear. As to the ICC values, one must keep in mind that the variance between the subjects is one important component in computation (see Bartko and Carpenter 1976): for a given extent of disagreement, if variance is low, as it is in the case of the M-scores, the ICC value also tends to be low. Thus, low ICC values may be due in part to the fact that our sample had relatively few management problems and consequently low variance in the M-scores. Nevertheless, there are indications that the M category can give rise to problems and is the most difficult to rate, as noted by Clare and Cairns (1978). Although all raters were trained in the use of the SIS, some of them had more interview experience than others. This could have resulted in discrepant ratings. It is possible that these differences in experience between raters are responsible for the unexpectedly lower interrater reliabilities, compared to the test-retest results.

In summary, the results of this study, although admittedly based on a small sample, indicate that the modified version of the SIS can be used reliably. Some weaknesses were identified by this study and by more detailed item-by-item analyses not reported here. As a result of these analyses, and with the help of our experience with about 700 interviews, we have tried to improve some items by making the rating rules in the manual more precise. Nevertheless, considerable effort must still be made during training, especially with respect to the rating of the management items. The SIS is not an easily applicable research interview, and it requires some time to become familiar with the main rating principles and the critical points of the instrument.

## 2. Structure and interrelationships

Analyses of interrelationships provided us with the following information on which to base a scoring system. The three categories ("Objective Condi-

tions," "Social Management," and "Satisfaction") can be seen as *relatively independent dimensions*. The relationships between the categories O and M, as well as between O and S, are so slight that they can be disregarded. The relative independence of the categories O and M is especially desirable, as it indicates that the concept of the SIS was well transformed into the rating rules. A degree of overlap between the subjective evaluation of category S and the other two categories is to be expected. We found, in fact, a modest association between the categories M and S, but nearly no overlap between O and S. It has to be remembered, however, that these results might be different in other populations. Interestingly, in a sample of former psychiatric inpatients the relationship between the categories M and S was generally higher (about 25% common variance) than in the normal population sample of the present study (about 10% common variance). We assume that for these psychiatric inpatients the degree of satisfaction was influenced by social management difficulties more than in the case of the subjects from the general population. In the case of correlation between any two dimensions, of course, the possibility of causal effects in both directions must be considered. The satisfaction and dissatisfaction expressed by both samples was generally not strongly related to the objective conditions. It would be worthwhile to explore in more detail the possible varying degrees of relationships between categories in different samples.

The relationships between the three categories, O, M and S, of the SIS can be differentiated by looking at the associations within the various role areas. The highest associations are found between the categories M and S, with common variance ranging between 20% and 36%. These associations are not strong enough however, to restrict seriously the possibility of aggregating item scores across different categories; the different categories cover sufficiently heterogeneous aspects within each role area to justify their aggregation.

The same conclusion can be drawn from the analyses of relationships between the items of different role areas (within each of the categories O, M, and S). Although there are some significant correlations, common variance exceeds 20% in only a few cases. These overlapping items would slightly elevate the weights of some elements in the aggregation score, e.g., the common aspects of management in the areas of partnership, or restricted leisure time due to excessive housework tasks. Overlapping between items of the SIS is thus generally minimal and less than that found in similar instruments also attempting to aggregate over heterogeneous aspects (see, for example, Platt et al. 1980).



In summary, our analyses of interrelationships gave some support for the conceptual grouping of the items in the various role areas and the three rating categories. The relationships between items of different categories or of different role areas are small enough to treat the aggregated items as discrete components and to base the scoring system on this grouping.

### 3. Suggestions for scoring

In general, the method of scoring SIS data should primarily depend on the question to be examined. This paper, therefore, will not prescribe a fixed scoring system, but rather circumscribe a framework of scoring possibilities.

Clare and Cairns (1978) propose an overall social maladjustment score, which is computed by counting the number of items with a rating of 3 or 4 (marked or severe difficulties or dissatisfaction) for each individual over all items. We have two objections to this suggestion. First, the meaning of this score is rather obscure. Management and satisfaction aspects are combined with aspects of the individual's objective conditions. While the first two categories can be treated as the behavioral and expressive aspects of social adjustment, the objective conditions are hardly an aspect of adjustment itself but, rather, as formulated in the theoretical concept of the SIS, a framework for adjustment. Aggregating the aspects of adjustment with its objective framework does not seem to make much sense. Second, as with nearly all adjustment scales, the number of items rated for an individual varies with the roles he fulfils. Thus, individuals with more items rated have a greater chance of getting a high score than individuals with fewer items rated. Individual scores unadjusted for the number of items, therefore, are not necessarily comparable.

Consequently, we first would suggest an adjusted score: the number of items with a rating of 3 or 4 should be corrected by dividing by the number of items applied. We further suggest not including the items of the category O in this score of overall social maladjustment, aggregating rather only M and S items. In addition, we would suggest, in accordance with Clare and Cairns, that scores be computed for each category O, M, and S by counting the number of ratings of 3 or 4 within the category O, M, and S and then dividing by the number of items applicable to the individual in this category. Thus, four scores would result:

Overall social maladjustment score:

$$\frac{\text{number of items in the categories M and S with a rating of 3 or 4}}{\text{number of applicable items in the categories M and S}}$$

Objective conditions score:

$$\frac{\text{number of items in the category O with a rating of 3 or 4}}{\text{number of applicable items in the category O}}$$

Social management score:

$$\frac{\text{number of items in the category M with a rating of 3 or 4}}{\text{number of applicable items in the category M}}$$

Satisfaction score:

$$\frac{\text{number of items in the category S with a rating of 3 or 4}}{\text{number of applicable items in the category S}}$$

Each of these scores varies from a minimum of 0 to a maximum of 1. They are based on a dichotomization of the original 4-point scales between the rating points 1/2 and 3/4; only marked or severe stressful conditions/restrictions, marked or severe management problems, and the self-rating of "dissatisfied" or "very dissatisfied" are counted for the score. One could also imagine dichotomizing the scale at another point, e.g., between the rating points 1 and 2/3/4. In this case, no problems would be opposed to the existence of some kind of problem (from minor to severe). This strategy would be more appropriate when examining differences between groups which are hypothesized to have only a moderately reduced level of adjustment (e.g., between groups within the normal population).

We cannot recommend that the 4-point scale be used in its full range by aggregating the original item scores. We cannot claim interval properties for these scales and the reliability of some items is questionable. We see, however, no problems in using the dichotomized scales.

Another scoring possibility would be the aggregation of items of various role areas, e.g., work, leisure, family (partnership, children, relatives). The items to be combined would be the result of theoretical considerations and the question of interest. The analyses of interrelationships have revealed no major obstacles to aggregating items across the categories, O, M and S, or across role areas.

*Acknowledgements.* This work was funded in part by the Robert Bosch Foundation. The development of this instrument was initiated at the General Practice Research Unit at the Institute of Psychiatry, directed by Professor M. Shepherd. We are indebted to all contributors to the English version of the SIS.

The reliability study was organized by Dipl. Psych. G. Semler and Dipl. Psych. I. Buschor (Design: Hansert and Mombour). We would like to thank them and the interviewers and raters of the reliability study: Dr. C. Krieg; H. Hecht, Dipl. Psych.; W. Maier-Diewald, Dipl. Psych.; Dr. U. Rupp; G. Semler, Dipl. Psych.; Dr. H.-U. Wittchen, Dipl. Psych.; and Dr. A. Ulrich. We would also like to thank Dipl. Inf. H. Pfister for computation of the reliability results.

## References

- Barrabee P, Barrabee EL, Finesinger JE (1955) A normative social adjustment scale. *Am J Psychiatry* 112: 252-259
- Bartko JJ (1966) The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 19: 3-11
- Bartko JJ, Carpenter WT (1976) On the methods and theory of reliability. *J Nerv Ment Dis* 163: 307-317
- Clare AW, Cairns VE (1978) Design, development and use of a standardized interview to assess social maladjustment and dysfunction in community studies. *Psychol Med* 8: 589-604
- Cooper B, Eastwood MR, Sylph JA (1970) Psychiatric morbidity and social adjustment in a general practice population. In: Hare EH, Wing JK (eds) *Psychiatric Epidemiology*. Oxford University Press, London
- Corney RH, Clare AW, Fry J (1982) The development of a self-report questionnaire to identify social problems - a pilot study. *Psychol Med* 12: 903-909
- Donald CA, Ware JE, Brook RH, Davies-Avery A (1978) Conceptualization and measurement of health for adults in the Health Insurance Study, Vol. IV Social Health. Rand Corporation, Santa Monica
- Everitt BS (1977) *The analysis of contingency tables*. London: Chapman and Hall
- Faltermaier T (1983) Zusammenhänge zwischen sozialen Faktoren und psychiatrischer Morbidität. In: Kommer D, Röhrle B (Hrsg) *Gemeindepsychologische Perspektiven 3 - Ökologie und Lebenslagen*. Tübingen, Köln: DGVt/GWG
- Faltermaier T (1982) Manual "Social Interview Schedule": Deutsche revidierte Version (2. Fassung). Max-Planck-Institut für Psychiatrie, München (unpublished)
- Frank JK (1961) *Persuasion and healing*. Baltimore: John Hopkins Press
- Freeman HE, Simmons GG (1963) *The mental patient comes home*. John Wiley, New York
- Goodman LA (1978) *Analyzing Qualitative/Categorical Data*. Abt Books, Cambridge (Mass)
- Gurland BJ, Yorkston NJ, Goldberg K, Fleiss JL, Sloane RB, Cristol AH (1972) The structured and scaled interview to assess maladjustment (SSIAM). II. Factor analysis, reliability, and validity. *Arch Gen Psychiatry* 27: 264-267
- Katz MM, Lyerly SB (1963) Methods for measuring adjustment and social behavior in the community. I. Rationale, description, discriminative validity and scale development. *Psychol Rep* 13: 503-535
- Kedward HB, Sylph JA (1974) Social correlates of chronic neurotic disorder. *Soc Psychiatry* 9: 91-98
- Kleining G (1975) Soziale Mobilität in der Bundesrepublik Deutschland. II: Status- oder Prestige-Mobilität. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 27: 273-292
- Linn MW, Sculthorpe WB, Evje M, Slater PH (1969) A social dysfunction rating scale. *J Psychiatr Res* 6: 299-306
- Mann AH, Jenkins R, Belsey E (1981) The twelve-month outcome of patients with neurotic illness in general practice. *Psychol Med* 11: 535-550
- Moore H, Kleining G (1960) Das soziale Selbstbild der Gesellschaftsschichten in Deutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 12: 86-119
- Parsons T, Bales RF (1954) *Family socialization and interaction process*. Free Press, Glencoe, Illinois
- Paykel ES, Weissman MM (1973) Social adjustment and depression: a longitudinal study. *Arch Gen Psychiatry* 28: 659-663
- Paykel ES, Weissman MM, Prusoff BA (1978) Social maladjustment and severity of depression. *Compr Psychiatry* 19: 121-128
- Paykel ES, Weissman MM, Prusoff BA, Tonks CM (1971) Dimensions of social adjustment in depressed women. *J Nerv Ment Dis* 152: 158-172
- Platt S (1980) On establishing the validity of 'objective' data: can we rely on cross-interviewer agreement? *Psychol Med* 10: 573-581
- Platt S (1981) Social adjustment as a criterion of treatment success: just what are we measuring? *Psychiatry* 44: 95-112
- Platt S, Weyman A, Hirsch S, Hewett S (1980) The Social Behavior Assessment Schedule (SBAS): rationale, contents, scoring and reliability of a new interview schedule. *Soc Psychiatry* 15: 43-55
- Remington M, Tyrer P (1979) The Social Functioning Schedule - A brief semi-structured interview. *Soc Psychiatry* 14: 151-157
- Serban G (1978) Social stress and functioning inventory for psychotic disorders (SSFIPD): measurement and prediction of schizophrenics' community adjustment. *Comprehensive Psychiatry* 19: 337-346
- Tanner J, Weissman MM, Prusoff BA (1975) Social adjustment and clinical relapse in depressed outpatients. *Compr Psychiatr* 16: 547-556
- Weissman MM (1975) The assessment of social adjustment: a review of techniques. *Arch Gen Psychiatry* 32: 357-365
- Weissman MM, Klerman GL, Paykel ES, Prusoff BA, Hanson B (1974) Treatment effects on the social adjustment of depressed patients. *Arch Gen Psychiatry* 30: 771-778
- Weissman MM, Sholomskas D, John K (1981) The assessment of social adjustment: an update. *Arch Gen Psychiatry* 38: 1250-1258
- Williams P, Tarnopolsky A, Hand D (1980) Case definition and case identification in psychiatric epidemiology: review and assessment. *Psychol Med* 10: 101-114
- Wing JK, Bebbington P, Robins LN (Eds.) (1981) *What is a case? The problem of definition in psychiatric community surveys*. Grant McIntyre, London
- Wittchen H-U, v. Zerssen D (1985) Verlauf und Ausgang behandelter und unbehaunder psychischer Störungen. Springer, Heidelberg (In prep)
- Wittchen H-U, Ellmann R, v. Zerssen D, Möller H-J (1983) Der Verlauf behandelter und unbehaunder psychischer Störungen. In: Kommer D, Röhrle B (Hrsg) *Gemeindepsychologische Perspektiven 3 - Ökologie und Lebenslagen*. Tübingen, Köln, DGVt/GWG

Accepted 3 September 1984

Diplom Psychologe T. Faltermaier  
(Dr. H.-U. Wittchen) Max-Planck-Institut  
für Psychiatrie (Katamnese)  
Kraepelinstraße 10  
D-8000 München 40  
Federal Republic of Germany