

COOPERATION STRUCTURE, GROUP SIZE AND PRODUCTIVITY IN RESEARCH GROUPS

HILDRUN KRETSCHMER

*Institute of Theory, History and Organization of Science,
Academy of Sciences of the GDR, DDR-1100 Berlin,
Prenzlauer Promenade 149-152 (GDR)*

(Received August 9, 1983, in revised form April 12, 1984)

A research group is considered to be a system and the scientists the elements in this system. The degree of interaction among scientists is determined by means of a complex structure measure for groups. It is shown that optimum cooperation structures depend on group size. In addition, it was possible to determine an optimum group size. Various hypotheses have been verified employing the same data material by using several levels of the structure measure.

Introduction

For the purpose of organizing work in research groups it is important to know the productivity in dependence upon the group size. The results of studies on this topic are controversial:

1. *Cohen*¹ did not find any correlation between the *output per scientist* and group size, *Cohen* (Ref. 1). According to *Cohen* the number of publications of a laboratory in one year is proportional to the number of scientists in the laboratory during that year, so that "there is no gain in productivity to be sought by favouring the investment of resources differentially according to laboratory size per se" (Ref. 1, p. 49).

2. Some authors did find a correlation between *output per scientist/performance* and group size (Refs. 2, 3).

3. As a third alternative a correlation was found between the *output per scientist* and group size, yet only under specific conditions (e.g. leadership)⁴.

Several authors in groups 2 and 3 proceed from the hypothesis that the effectiveness of research groups is dependent upon the intellectual interaction among their members. This interaction is supposed to raise the creative potential and the productive capacity of group members. They conjectured that interaction would increase if the group size increased as well and, hence, correlations should exist between group size and productivity. This paper is an attempt to show that the

following two factors contribute to the actually conflicting views on the output per scientist and group size:

Interaction and group size are not exchangeable variables. The degree of interaction, in addition to group size, should be considered.

The degree of interaction can be determined with the help of the complex structure measure for groups.* In the calculations, cooperation was utilized as a characteristic of interaction. It is possible to consider a research group as a system and the scientists as the elements in this system. Structure can be described for systems, making it indispensable, however, to adapt the structure of the system to the expected performance of this system. If it were feasible to set exactly the various conditions of work in research groups and to derive an optimum cooperation structure it would be possible to increase the effectiveness in research groups.

Hypotheses

The structure measure allows to verify simultaneously various hypotheses derived from the same data material:

1st hypothesis. There is a correlation between the scope of cooperation of two scientists and the difference in the number of their publications;

2nd hypothesis. There is a correlation between the cooperation relations of a scientist and the number of his publications;

3rd hypothesis. There is a correlation between the cooperation structure of a research group and the output per scientist.

While the first hypothesis refers to measures on the first level and the second one to measures on the first and second levels, the third hypothesis refers to measures on all three levels of the complex structure measure.

Methods

The research activity in molecular biology done by some 450 scientists in 56 research groups was investigated. Technical staff was not included. The research groups were chosen by chance. The group size varied from 3 to 17 members. The group size was taken from the lists of the institutions on the day of the investigation. For this study the number of publications (output per scientist) based on a publication rate over five years was used as a rough measure of group performance.

*See Appendix A, Table 1 and Ref. 5 with explanations of the measure.

The group structure measure was applied to the cooperation structure of the groups. The application of the structure measure required a matrix, with the relations x_{ij} to be recorded between two group members each. The members of the research groups were unable to assess *directly*, their relations with the other members of the group so that the totality of work of a group member was divided into binary steps (Fig. 9 and questionnaire in Appendix B). The reliability of the questionnaire was determined by a retest $r_{tt} = 0.76$ with 50 scientists. Figure 9 shows that there is in fact no clear-cut division into binary steps. That is why it has to be tested empirically whether the error occurring lies within an admissible margin defined for psychological studies. Every relation between x_b and x_c was assessed twice, once from the scientist x_b and once from x_c . In comparing these pairs of values the coefficient $r = 0.65$ (Pearson's r) was obtained.

The validity of the results was tested by comparing some of the results with the findings made by other authors (cf. hypotheses 1 and 2). From this it was possible to conclude on the validity of further results on the third level; the calculation of the measures on the different hierarchical levels of the structure measure was carried out on the same basis.

Results and discussion

1st hypothesis

The scientists were divided into six groups according to the number of their publications. This choice of the groups is in accordance with the choice of the groups in Ref. 5 and allows the comparison of the results:

1st group: 0 (zero) and 1 (one) publication; 2nd group: 2–3 publications; 3rd group: 4–7 publications; 4th group: 8–15 publications; 5th group: 16–31 publications; 6th group: more than 31 publications.

The relative scope of cooperation between two scientists each ($S_{1ij} = x_{ij}$) was determined (cf. Table 1 and Fig. 9). The sum of these values between two groups each was ascertained. A chi-square-test was carried out with the values of the 6×6 matrix. It was possible to establish a statistically significant connexion ($\chi^2 > \chi^2_{0.001}$) between the differences in the number of publications of scientists and their cooperation relations.

The mean values of S_{1ij} between two groups each were determined and recorded in a 6×6 matrix. According to the chi-square-test the statistically expected values were calculated for these mean values. The frequency (F) of S_{1ij} above as well as the frequency (F') of S_{1ij} below the adequate expected value was determined for

each field of the matrix. The quotient F/F' was represented in dependence upon the difference D in the number of publications of scientists (D = difference between the ordinal numbers of two groups) (cf. Fig. 1).

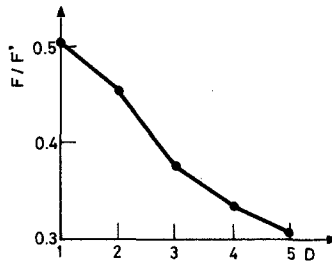


Fig. 1

The relative scope of cooperation decreases more and more as soon as the number of publications between two scientists differ. It was possible to prove this trend statistically ($p < 0.003$). This result agreed with the analysis obtained from the structure of citations⁵. Assuming that the eminence of a scientist increased proportionately with the logarithm of the number of publications, it had to be viewed as an expression of the stratification of scientists (see Ref. 6).

2nd hypothesis

On the basis of empirical findings by *Pelz* and *Andrews*⁷ prolific scientists had a more intense cooperation and a greater number of cooperation partners. However, it was not possible to prove this in the calculation on the basis of B_{1i} . The relative scope of cooperation B_{1i} of a scientist x_i shall be understood as part of his total work by the above method. But since the scope of total work of a productive scientist is larger than that of a less productive one, the same value for B_{1i} , taken in absolute terms, is larger for a productive scientist than that for a less productive one. In the transposed matrix this effect is largely eliminated. Next, B_{1ij} is assumed to be the reflection of the scope of cooperation of a scientist against the work of all other scientists of the research group. If this hypothesis were valid, the relationship of B_{1j} (in the transposed matrix) with B_{1i} should be greater for productive scientists than for less productive ones. The chi-square-test resulted in ($\chi^2 > \chi^2_{0.05}$). By inference, it was possible to prove a correlation between the number of publications of a scientist and the scope of his cooperation in the research group (B_{1j} in the transposed matrix) ($\chi^2 > \chi^2_{0.001}$), cp. Fig. 2.

The relative 'number' of cooperation partners of scientist x_i , B_{2i} , is independent of B_{1i} , thus it is not subject to the above restriction of B_{1i} either. Accordingly, both for B_{2i} and for B_{2j} (in the transposed matrix), it is possible to establish a significant correlation as to the number of publications. Hence, a productive scientist has more cooperation partners than a less productive one (Fig. 3). As expected from the results based on B_{1j} and B_{2j} there is a statistical connexion ($\chi^2 > \chi^2_{0.001}$) between the degree of integration of scientist x_i into the group, S_{2j} (in the transposed matrix) and the number of publications (Fig. 4).

3rd hypothesis

According to *Stankiewicz*⁴ there is a significant correlation between group size and the age of the group, i.e. the size increases with the period of existence of the group. It is to be tested whether the cooperation structure is changed with increasing group size. The results suggest an optimum group size in relation to cooperation. The relative scope of cooperation relations (B_1) is highest for groups with 6–12 members ($t > t_{0.001}$) (Fig. 5). The same result applies to the cooperation structure (S_3) ($t > t_{0.05}$) (Fig. 6). The relative 'number' of cooperation partners (B_2) decrease with increasing group size ($p < 0.001$), (Fig. 7).

The following hypothesis shall be tested here: "The adaptation of cooperation structure in research groups to the group size is an essential factor for raising the effectiveness of research groups."

This hypothesis shows a way of intensifying research. According to *Haitun*⁸ an intensification of research means its optimization, and the science of science should use quantitative methods to solve this problem.

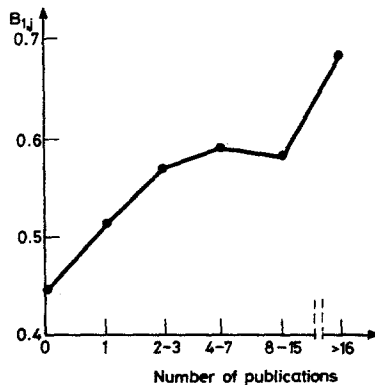


Fig. 2

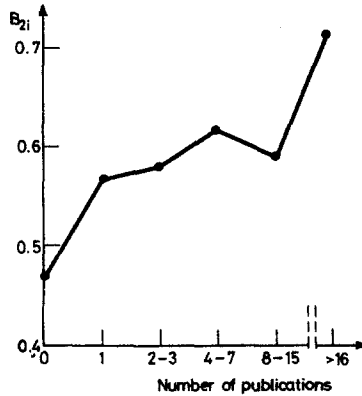


Fig. 3

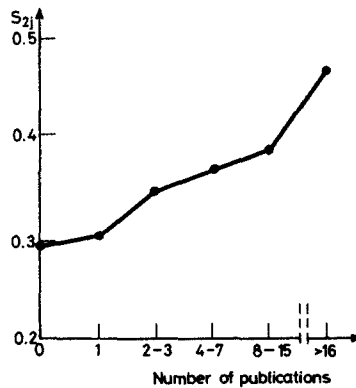


Fig. 4

The mean value m_1 of S_3 for groups with 6–12 members is significantly higher than the mean value m_2 of S_3 for groups with 3–5 and more than 12 members. If in this case it actually was a question of adaptation of structure to the group size, the mean values should represent the optimum and the effectiveness of groups should increase when approaching the optimum.

The circles show the empirically found averages of the number of publications over the intervals of S_3 (broken line = groups with 3–5 and more than 12 members; full line = groups with 6–12 members). The ordinate in Fig. 8 shows the average number of publications per group member. There is a statistically significant correlation ($p < 0.01$) for 56 research groups between the deviations of the structure of

HILDRUN KRETSCHMER: PRODUCTIVITY IN RESEARCH GROUPS

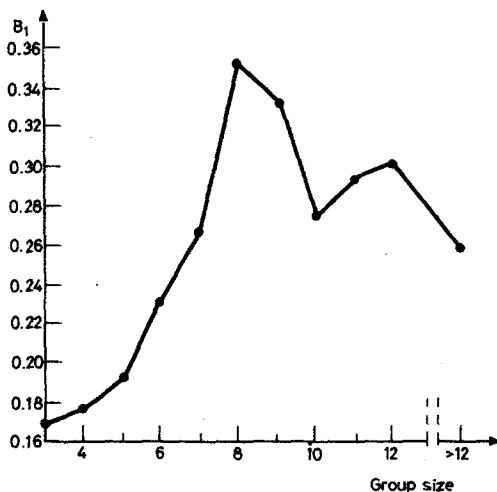


Fig. 5

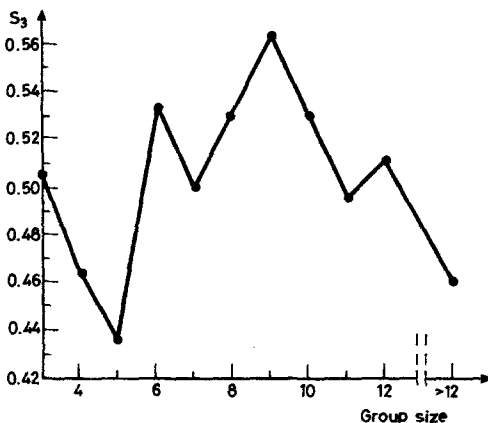


Fig. 6

research groups from the optimum* and their effectiveness (Spearman's correlation $R = -0.404$).

According to a theory by *Schroder, Driver and Streufert*⁹ on the connexion between the complexity of environment and the level of information processing, there was an optimum degree of complexity for information processing. The individuals differ from each other with regard to their optimum degree of complexity.

* (Standardized values $(\frac{x - \bar{x}}{s})$, x — S_3 , \bar{x} — optimum, s — standard deviation.)

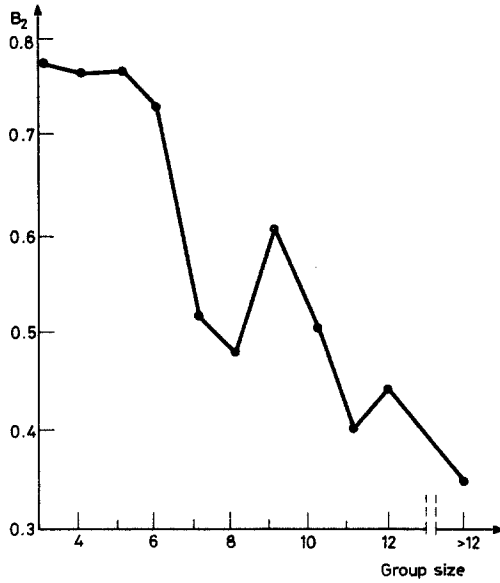


Fig. 7

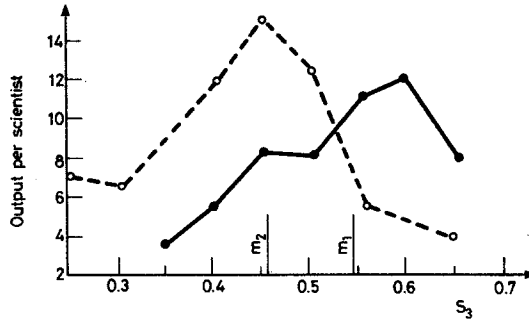


Fig. 8

The higher the optimum degree of complexity of an individual, the higher the maximum attainable level of information processing. By analogy, it is possible that this regularity is also applicable to research groups. In this case the structure is analogous to the complexity of environment, the publication rate analogous to the level of information processing and the members of a research group are analogous to the individuals.

For this reason two effects may arise:

1. The productivity of a group may increase with the approximation to its optimum structure, as shown here.

2. The maximum attainable level of information processing for groups with 6–12 members may be higher than the level for other groups, hence, a correlation between productivity and group size per se is possible, additionally to the first effect. If the second effect is weaker than the first one the second effect will not always be provable, so that the conflicting results on the correlation between group size and group performance will become apparent.

In this paper no relationship was found between group size per se and productivity. If the total random test was divided into two parts according to the deviation of structure from the optimum, it was found for groups with strong deviation (with $\chi^2 > \chi^2_{0.05}$) that groups with 6–12 members showed a higher productivity.

Under these aspects several considerations by *Cohen*^{1,10,11}, *Qurashi*^{3,12} and *Stankiewicz*⁴ shall be analyzed. According to *Cohen* the rate of publication was independent of the group size. By contrast, *Qurashi* asserted that there was one or several group sizes with maximum publication rate. *Qurashi*¹² analyzed *Cohen's* data^{1,10} from the National Cancer Institute (USA), N.C.I. as well as from the National Institute of Medical Research (U.K.), N.I.M.R. from 1976–77. This analysis was based on a subdivision of data into successive ranks of group size 1–3, 4–6, 7–9, . . . and on the calculation of the relevant publication rate per person (R) for each rank. As a result *Qurashi* found maximum publication rates for the groups sizes 6; 16 and 27 ±2 scientists. *Cohen*¹¹ has shown, however, that *Qurashi* did not provide any statistical proof and that R may accidentally assume these values for the indicated group size; nevertheless the ideas of *Qurashi* on the correlation between group size and output per scientist should not be rejected offhand. According to the theory outlined in Ref.⁹ and on account of the two effects to be derived in this paper it had to be expected that in several random tests optimum group sizes were provable, whereas they could not be fixed in others. The results of *Stankiewicz*⁴ allow to presume that there were also optimum group sizes. Hence, it shall be checked whether in the three institutions of *Cohen* – N.C.I., N.I.M.R. and Rockefeller University, RU – (the term 'number of publications', as designated by *Cohen*, is used) at least some tendencies have become discernible.

Several presuppositions shall be included in the analysis:

1. If studies are to be conducted on the behaviour of groups, the group should be considered a *unit*, the N.C.I. would be a random test with 46 cases, with each group being assigned the output per scientist.

2. The 'output per scientist' is often abnormally distributed, such as in this study as well as in the random test of the Rockefeller University, so that non-parametrical tests are used.

3. No separate 'peaks' shall be sought, but rather a continuous transition from minimum to maximum values and vice versa.

4. According to the 'peaks' indicated by *Qurashi*¹² an optimum group size (G_s) was to be expected in distances of 10 group sizes:

$$G_s = c + i \cdot 10 \quad (i = 0, 1, 2 \dots)$$

c = constant.

A minimum value of the output per scientist is to be assumed in the middle (g_s) between two optimum group sizes:

$$g_s = c + i \cdot 10 \pm 5$$

The course of S_3 is also indicative of the change between minimum and maximum, with the minimum lying with the group size 5 and the maximum lying with group size 9. The small extent of the random test might be the reason for the inaccuracy in the distance between minimum and maximum that is not precisely 5.

5. The constant c can vary according to the kind of random test. *Qurashi*¹² has shown that there are different optimum group sizes, e.g. a difference can be made between highly specialized and other research groups.

6. It is assumed that the output per scientist agrees with a sinus function

$$y = a + b \cdot \sin(0.2 \pi x + d) \quad \text{or}$$

$$y = A + B \cdot \sin(0.2 \pi x) + C \cdot \cos(0.2 \pi x)$$

y —output per scientist

x (in rad) —group size

in dependence upon the group size. c is analogous to the parameter d , $0.2 \pi x$ is determined by the assumption in point 4.

In the N.C.I. the output per scientist has a normal distribution. Hence, the analysis of regression resulted in

$$y = 1.2 + 0.3 \cdot \sin(0.2 \pi x - 2.8)$$

with the general correlation coefficient $R_{yx} = 0.37$. This correlation is significant

on the 5% level (46 value pairs). The correlation resulted in $R_{yx} = 0.21$ for

N.I.M.R. (21 value pairs — not significant). In the RU the output per scientist is

not normally distributed. According to *Moran*¹³ multiple correlation based on

Spearman's R can be applied.* On the basis of $y = A + B \cdot \sin(0.2 \pi x) + C \cdot \cos(0.2 \pi x)$

the multiple correlation resulted in $R_{1,2,3} = 0.25$ (57 value pairs — not significant).

The three cases with $x = 1$ were excluded in RU because a single scientist is not a

group of scientists. The weighted mean value of the three *Spearman* multiple correlation coefficients of the three institutions** resulted in $\bar{R}_{1,2,3} = 0.31$. The sum of the

*Corrected *Spearman's R* was utilized in case of identical rank for several x_i or for several y_i .

**N.C.I.: $R_{1,2,3} = 0.40$; N.I.M.R.: $R_{1,2,3} = 0.28$.

value pairs is 124. The question is whether $y = a+b \cdot \sin (0.2 \pi x+d)$ is valid in general or only in biomedical research institutions. In the latter case, it is assumed that the output per scientist agrees with a general sinus function

$$y = a+b \cdot \sin (g \cdot x+d)$$

in research groups in dependence upon the group size.

*Stankiewicz*⁴ reported on an optimum group size related to the performance of small groups, but did not find any correlation between output per scientist and group size. However, this result changed by incorporating the leadership (Ref. 4, p. 20): “. . . there exists a strong positive correlation between group size and output per scientist in the groups headed by scientists with more than 14 years of research experience . . .” The totality of results that have proved significant correlations between the output per scientist and group size, but also weak tendencies in this direction or even independence, was to be expected in conformity with the theory of the authors of Ref. 9.

The considerations presented here can be classified by a structural mode and further development could provide recommendations for practical application. At present it is an important concern to intensify scientific work with a view to raising the efficiency of research groups (Ref. 14).

*

I would like to thank my mother, *M. Bonitz*, *J. Fischer*, *U. Geissler*, *U. Goedecker*, *K. Gräf*, *H. Parthey* and *L. Römer* for their support and their helpful recommendations.

Appendix

A. The cooperation structure was quantitatively determined by means of the complex structure measure for groups. This measure determines the degree of interaction in a research group. Groupings of relationships on different hierarchical levels enter into the structure measure. These measures can be used for empirical research both independently of each other, and combined.

1. S_{1ij} is the structure measure on the 1st hierarchical level:

The smallest unit S_{1ij} of the complex structure measure is the relative scope of the cooperation between two scientists x_i and x_j .

2. S_{2j} is the structure measure on the 2nd hierarchical level:

The combination of the smallest units to establish the next higher unit S_{2j} serves the determination of the cooperation relationships which one of the scientists each has to all the other scientists – this means the degree of his integration into the group. Differing from the smallest unit, the next higher unit is determined by two components:

2.1 B_{1j} by the relative scope of cooperation relationships of a scientist x_i to all the other scientists, i.e. the average of the smallest units which are attached to the scientist x_i .

HILDRUN KRETSCHMER: PRODUCTIVITY IN RESEARCH GROUPS

2.2 B_{2i} by the relative 'number' of cooperation partners of the scientist x_i . The two components can vary independently of each other and therefore they are combined together for the determination of the degree of interaction of a scientist into the group.

3. S_3 is the structure measure on the 3rd hierarchical level.

The combination of the units – mentioned in 3rd – serves the determination of the structure measure for the cooperation structure of the whole group. Analogous to second, the highest unit S_3 is determined by three components:

3.1 B_1 by the relative scope of cooperation relationships, that is the average of all B_{1i}

3.2 B_2 by the relative 'number' of cooperation partners, i.e., the average of all B_{2i}

3.3 by the relative homogeneity of the degree of interaction of all scientists. (B_3)

Analogous to second, the three components can vary independently of each other and therefore they are combined together for the determination of the cooperation structure in the research group.

The formula in Table 1 is a derivation of the original formula (Ref. 5), in accordance with the method for the recording of values in the matrix in this paper.

Table 1
The formula of the structure measure

						Column 1	Column 2	Column 3
						$\frac{m}{100} \sum_{j=1}^m x_{ij}$	$\frac{{}_2H_i(x_j)}{m-1}$	S_{2i}
x_j	x_1	x_2	x_3	x_4	x_5	$= B_{1i}$	$= B_{2i}$	
x_1		12.8	25.6	25.6		0.64	0.72	0.68
x_2	30		15	7.5		0.53	0.65	0.59
x_3	6	10		12	12	0.4	0.97	0.62
x_4	16.25	8.12	4.06		4.06	0.33	0.84	0.52
x_5	1.5	5	1.5	2		0.1	0.86	0.29
						$\frac{m}{\sum_{j=1}^m} \sum_{i=1}^m x_{ij}$	$\frac{m}{\sum_{i=1}^m} \frac{{}_2H_i(x_j)}{m-1}$	$\frac{{}_2^H(S_{2i})}{m}$
						$= B_1 = 0.4$	$= B_2 = 0.81$	$= B_3 = 0.96$

HILDRUN KRETSCHMER: PRODUCTIVITY IN RESEARCH GROUPS

Calculation of H:

There is a series of n numbers K_i ($i = 1, 2 \dots n$) for instance 12.8; 25.6; 25.6 then the h_i are

$$h_i = \frac{K_i}{\sum_{i=1}^n K_i}$$

According to the entropy formula

$$H = - \sum_{i=1}^n h_i \text{L}h_i$$

$H = 2.87$ for the upper series of numbers,

$$B_{2,1} = \frac{2^{2.87}}{4} = 0.72$$

$$\text{If } \sum_{i=1}^n K_i = 0 \text{ then } 2^H \geq 0$$

$$S_3 = \sqrt[3]{B_3 \cdot B_2 \cdot B_1}$$

$$S_3 = \sqrt[3]{0.96 \cdot 0.81 \cdot 0.4} = 0.68$$

derivation of the original formula by

$$\frac{\sum_{j=1}^m \frac{x_{ij}}{x_{ij\max}}}{m-1} = \frac{\sum_{j=1}^m x_{ij}}{100}$$

$$S_3 = \sqrt[3]{\frac{2^{H(S_{2i})}}{m} \cdot \frac{\sum_{i=1}^m \frac{2^{H_i(x_j)}}{m-1}}{m} \cdot \frac{\sum_{i=1}^m \frac{\sum_{j=1}^m x_{ij}}{100}}{m}}$$

$$S_{2i} = \sqrt{\frac{2^{H_i(x_j)}}{m-1} \cdot \frac{\sum_{j=1}^m x_{ij}}{100}}$$

$$S_{1ij} = x_{ij}$$

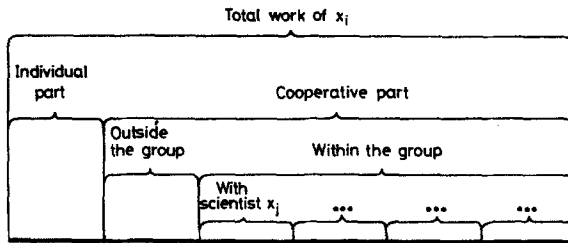


Fig. 9. Entries x_{ij} in the matrix

x_{ij} – relations between x_i and x_j

$$x_{ij} = S_{1ij} = [\text{cooperative part}] \cdot [\text{cooperative part within the group}] \cdot [\text{cooperative part with } x_j] \cdot 10^{-4}$$

Cf. questionnaire: cooperative part – question 3b; cooperative part within the group – question 2a; cooperative part with x_j – question 1 – the cooperative part (%) assigned the scientist x_j

B. Questionnaire

1. Rank the scientists in your research group. Put in the first place the colleague with whom you most closely cooperate. The cooperative part which can be assigned to the scientists of your research group amounts to 100%. Divide these 100% among the colleagues and first fill in the order and then the percentage.

	order	%
1.
2. etc.

2. 100% are assigned to your cooperative part in your work. Divide it into
 (a) a part which can be assigned to the cooperation relations with colleagues within your research group

(b) a part which can be assigned to the cooperation relations outside the research group
 Note that

$$a \dots \% + b \dots \% = 100\%$$

3. 100% are assigned to your total work. Divide your total work into

- (a) an individual part,
- (b) a cooperative part

Note that $a \dots \% + b \dots \% = 100\%$

(This questionnaire was given the definition of 'cooperation' according to *Winkler*.^{1 5}

References

1. J. E. COHEN, Publication rate as a function of laboratory size in a biomedical research institution. *Scientometrics*, 2 (1980) 35–52.
2. J. T. WALLMARK, S. ECKERSTEIN, B. LANGERED, H. E. S. HOLMQVIST, The increase in efficiency with size of research teams, in: *IEEE Trans, Eng. Manag.*, Vol. EM-20, No. 3, August, 1973, pp. 80–86. These authors define efficiency as the ratio of the number of references to the team size.
3. M. M. QURASHI, *The Mechanics of R & D and Technology Transfer, Pakistan Academy of Sciences*. No. 34, Street 32, Shalimar 7/1, Islamabad
4. R. STANKIEWICZ, *Leadership and the performance of research groups*. RPI Research Policy Institute, University of Lund, 1980.
5. H. KRETSCHMER, Representation of a Complex Structure Measure for Social Groups and its Application to the Structure of Citations in a Journal. *Scientometrics* 5 (1983) 5–30.
6. R. D. WHITLEY, Kommunikationsnetze in der Wissenschaft. Status und Zitierungsmuster in der Tierphysiologie in: P. WEINGART, Hrsg., *Wissenschaftssoziologie II*, Athenäum Fischer, Frankfurt/Main, 1974.
7. D. C. PELZ, F. M. ANDREWS, *Scientists in Organizations*, Rev. ed. Ann Arbor, Michigan, Institute for Social Research, The University of Michigan, 1976,
8. S. D. HAITUN, *Naukometrija. Sostoyanie i perspektivy*, Izdatelstvo Nauka, Moscow, 1983.
9. H. M. SCHRODER, M. J. DRIVER, S. STREUFERT, *Menschliche Informationsverarbeitung*, Weinheim, Basel, 1975.
10. J. E. COHEN, Publication rate as a function of laboratory size in three biomedical research institutions, *Scientometrics*, 3 (1981) 467–487.
11. J. E. COHEN, Statistical Theory Aids Inference in Scientometrics, *Scientometrics*, 6 (1984) 27.
12. M. M. QURASHI, Publication rate as a function of the laboratory/group size, *Scientometrics* 6 (1984) 19.
13. P. A. P. MORAN, Partial and multiple rank correlation, *Biometrika*, 38 (1951) 26–31.
14. P. B. SCHELISCHTSCH, *Dinamika nauki*, Leningrad' Nauka', Leningradskoye otdelenija, 1981.
15. R.–L. WINKLER, Zu Kooperationsformen in Forschungsgruppen, Akademie der Wissenschaften der DDR. Institut für Theorie, Geschichte und Organisation der Wissenschaft, *Studien und Forschungsberichte*. Teil I, 1977.