Contextual Analysis of Team Productivity in the R & D Industry

Yehouda A. Shenhav, Yitchak Haberfeld
Tel-Aviv University

and

Bernard P. Cohen
Stanford University

December 1987

Working Paper No. 87-5

We thank Frank M. Andrews for his comments and Shaul Gabay and Yasmin Alkalay for their assistance. The data were gathered for a project financed by National Science Foundation Grant ISI # 8304340 conducted at the Center for Sociological Research at Stanford University, Bernard P. Cohen, Principal Investigator. Ideas and results expressed in this paper do not necessarily reflect the opinions of the National Science Foundation. The paper was presented at the National Academy of Management Meeting, Anaheim-Los Angeles, California, August 1988.

Contextual Analysis of Team Productivity in the R & D Industry

Abstract

We argue that productivity is a phenomenon which takes on various meanings in different contexts. Reliability coefficients of six scales of productivity, four of which have been used by Andrews and/or by Pelz and Andrews, are estimated in 28 work contexts using data on 224 R & D teams. The results support the argument.

Scientific productivity has attracted the attention of those engaged in R & D management and in social studies of science. Most research, however, concentrates on the antecedents and the consequences of productivity rather than on its meaning as a theoretical construct (see for example: Fox, 1983). The concept is rarely defined, and its arbitrarily chosen and interchangeably used indicators are seldom validated before the (assumed) correlates of the phenomenon are studied. In a recent study (Shenhav and Haberfeld, 1988), we demonstrated that fluctuations in reliability coefficients of productivity indices are associated with work contexts. This is not merely a methodological issue since conceptualization, validity, and reliability are interconnected: without proper conceptualization, valid and reliable measures cannot- be developed and, in turn, failure to determine valid indicators results in inaccurate specifications of the construct of productivity.

In this paper we elaborate upon contextual conditions within which the productivity of R & D teams may be better defined. An understanding of such conditions should be useful to R & D managers faced with the need to assess the performance of their teams. For this purpose we utilize information on 224 R & D teams from the American R & D industry.

1. Conceptual Framework

Our model is based on three main interconnected assumptions: (a) Productivity is a context-specific concept. That is, its meaning varies from one work context to another. (b) Organizations reward members who conform to the local, context-specific definition of productivity. (c) Organizational members are aware of the prevailing reward structure in their specific context and develop strategies designed to promote rewarded behavior (Scott et al, 1967)- The examination of productive behavior thus reveals the contextual meaning of productivity.

These concepts suggest that productivity is a socially constructed phenomenon, differentially interpreted by people operating in dissimilar work contexts, who establish their own expectations (Shrum, 1985), codes and cultures of productivity (Akin and Hopelain, 1986). Scientists also differ from one another in their perception of productivity as a result of their specific training, socialization processes, and divergent career paths (Goldberg and Shenhav, 1984).

We therefore argue that a crucial step toward defining the concept and validating its indicators is to minimize differences in the participants' conception of desired products. This can be achieved by delineating homogeneous contexts and subgroups for which the concept has a shared meaning. Such contexts may be defined according to various parameters, such as sector of activity (academic, nonacademic), research characteristics (theoretical vs. experimental, externally vs. internally funded research), individual attributes (young vs. veteran researchers) and professional affiliation (scientists vs. engineers) (see: Shenhav and Haberfeld, 1988).

In the present study we examine how structural contexts within which industrial R & D teams operate affect the teams' productive behavior.

A fruitful first step derived from the above model is to distinguish between industries and organizations. It is reasonable to assume that in each of these frameworks a somewhat different concept of productivity evolves, linked directly to circumstances under which the industry (or the organization) operates (Dalton et al, 1980).

Within each organization research teams may be classified according to their social structure. We distinguish in particular between centralized and decentralized teams (Cohen et al, 1982). Other factors of significance for the specific definition of productivity include the nature of the research project (i.e., basic research, applied research, or development) (see: Kornhauser, 1962; Yankevich, 1982); team size (Quarshi, 1984; Gooding and Wagner, 1985); engineering projects as opposed to the natural sciences (Ritti, 1968); type of scientific discipline (Hagstrom, 1965) or technical field (Shrum, 1985). Finally, we suggest that the disciplinary composition of a team should affect the nature of productivity (Chubin et al, 1986).

Since we maintain that each context is characterized by its own definition of productivity, we assume that the behavior of teams working in these contexts reflects the appropriate definition. They would thus be expected to produce more of the product (or products) required by that definition and a smaller volume of other products. On the strength of these assumptions, we applied several widely used indicators of scientific productivity (including those used by Andrews (1979) and Pelz and Andrews (1966)) to various heterogeneous contexts, seeking to determine their relevance to each context.

1. Productivity Measurement

Our approach is based on measurement theory (Nunnally, 1978), which distinguishes between two concepts: measurement reliability and measurement validity. The former refers to the stability of an observed score over specially constructed parallel measures of the same attribute (Guilford, 1954), or to the stability of the score when using random sample measures of the same attribute (Lord and Novick, 1968). In its general form, the reliability coefficient ($r_{»_J »_J}$) can be expressed as:

y y " ^{s}y ' y

(The ratio of the true score variance to the observed—measured-score variance).

Of the several types of validity (Campbell, 1976) the most relevant to this study is construct validity (Schwab, 1980). While reliability refers only to the stability of the measure, construct validity refers to the relationship between this measure and the theoretical construct. If low reliablity coefficients are obtained we may conclude that the validity of the measures is commensurately low. Thus, every attempt at estimating the validity of mesures should begin with a reliability analysis. Our study seeks to identify conditions under which reliability varies. With this aim in mind, we consider the various combinations of productivity indices and alternative classifications of scientific contexts specified below.

Several studies have focused on measurement issues and interrelations between different indicators of productivity. An early comprehensive study of scientific productivity was conducted by Pelz and Andrews (1966). They identify several factors conducive to a productive scientific work climate, as measured by judged contribution, judged usefulness of products, output of technical reports, papers, and patents. While they find positive correlations between all the measures, these are far from perfect. The authors explain that the indicators are intended to measure different aspects of productivity and are therefore not expected to show perfect agreement (Pelz and Andrews, 1966). A subsequent study by Andrews and his colleagues (Andrews, 1979) also considers various productivity measures and arrives at a similar conclusion. Questions can be raised about this rationale, since different measures of the same construct should be highly correlated. Explanations proffering concept dimensionality thus raise a legitimate concern about the existence of multiple concepts rather than a single one. Other studies compare the reliability of diferent- measures of productivity (Anderson, Narin, and McAllister, 1978; Koenig, 1982; Koenig, 1983) but lack appropriate conceptualization.

Methods and Sample

It should be emphasized that our analyses serve illustrative purposes only. We do not make any inferences about the "appropriate" productivity measures for the various contexts. The main purpose of our analyses is to demonstrate possible differences in the construct of scientific productivity when applied to diverse situations.

The study utilizes reliability analysis. We employ Cronbach's $_{a}$, which is a function of the correlations among predictors and the number of predictors used.

The data for the study were collected during 1985-1986 as part of an NSF nationwide research project on the social structure of R & D teams conducted at Stanford University. The teams included in the sample were selected according to **prespecified guidelines, and drawn** from those companies which agreed to participate in the study. A team was defined as a group of people working on a common scientific or technical task and formally recognized as a unit within the company.

The sample of 224 different research teams was drawn from 29 American companies representing seven diferent lines of business (Cohen et al, 1986): automotive, chemical, electronics, biotechnology, aerospace, pharmaceutical, and oil. The 224 teams range in size from 3 to 34 members with an average of 10.2 members and with a standard deviation of 4.9. The data employed in this study are based on responses to three types of questionnaire. The major instrument (UM) was distributed to every team member including the team leader. The response rate was high: 91% of the members of the selected teams returned completed questionnaires. The team leader also completed a short supplementary questionnaire (UL). Finally, an external evaluator representing top level management answered a team evaluation questionnaire.

Several structural measures of the team are based on the UL questionnaire (e.g., tangible products produced by the team), while others are based on central statistical values (such as the mode or the mean) which were calculated across unit members within their original teams (from UM questionnaires).

Measures and Indices of Productivity

Both subjective and tangible measures of productivity are used in the present study (see appendix). The former comprises an overall rating of team productivity made by team members, team leader, and external evaluators on a seven-category scale. The latter is composed of ten possible products such as position papers, proposals, patents, or articles, as specified in the appendix. Team productivity on each item was rated by the leader on a four-category scale.

These indicators serve as a basis for constructing several indices. We began with four indices presented in the literature by Pelz and Andrews (1966) and Andrews (1979): (1) Pelz and Andrews (1966: 271) constructed an index comprising of written materials (patent applications, technical papers, books, and technical manuscripts). Andrews (1979: 36-37) proposed three additional and separate indices: (2) written output; (3) patents and prototypes; (4) reports and algorithms.

We then extended our list of indices to include the following two scales: (5) all tangible products; (6) a combination of all three subjective evaluations (members, leader, and evaluators) .

The subjective evaluations of members are represented by the average response within a team. The subjective evaluation of management is based on external evaluations. Since each team was evaluated by two different evaluators, the one who was more familiar with the team was included in the analyses, The **evaluation** is the score **standardized** across **all** teams **evaluated** by the same evaluator. Only those evaluators with at least two evaluated teams were included in the analyses.

Contexts of Analyses

The main purpose of the study was to analyze the relevance of productivity indices in different contexts. The organizational contexts described at the outset of this paper are as follows:

(a) Industry type (automotive, chemicals, electronics, aerospace, pharmaceutical, biotechnology, oil).

(b) Company (four different companies, each with at least ten teams).

(c) Social structure. The classification of team social structure is based on work done by Cohen et al (1982) who proposed a four item typology: (1) Leader centered (Type A): a highly centralized structure. This type is similar to the professor-student model. (2) Collegial (Type B): collective activity involving mutual agreement and sharing of responsibility among team members. This type is similar to the academic ideal of collegial interaction. (3) Autonomous (Type C): a highly decentralized structure with minimal exercise of internal control: "the team project is largely an umbrella over distinct and disparate subprojects..(Cohen et al, 1982: 211; Cohen et al, 1986). (4) Mixed (Type D): team members share the intellectual responsibility of the work through a collective planning process. However, the execution of the work is centralized.

(d) Type of research projects, representing the area of R & D: basic research, applied research, and development.

(e) Big versus small teams (where the median of the size distribution is the outpoint).

(f) Interdisciplinary versus monod iscipl inar y teams.

(g) Type of scientific discipline (four different disciplines each represented by at least ten teams: biology, physics, chemistry and engineering). Originally, ten different categories were used in characterizing individuals according to their disciplinary affiliation. A team was defined as affiliated to a particular discipline if all its members belonged to one discipline (monodisciplinary) . Alternatively, the modal response was used if <u>all</u> of the following selection criteria were met: (1) the team consists of at least five members. (2) no additional disciplinary category is greater than half of the modal response. (3) Blau's (1977) measure of heterogeneity does not exceed .50. This measure is calculated as follows:

$$H = 1 - \frac{EX.}{CSX.)}$$

where X is the number of members in each category and the sum is computed over all categories. If all members fall into one category, there is no heterogeneity (which is the desired outcome in our case). If all categories have the same size, heterogeneity approximates unity with increasing numbers of categories.

Results

Descriptive statistics for the productivity and the contextual variables are presented in Tables 1 and 2.

(Tables 1 and 2 About Here)

Table 3 presents the reliability coefficients of each of the six productivity scales in 28 different contexts. Summary statistics of means and standard deviations are provided for each column (scale). It is clear from the results that the scale which encompasses all tangible products has the highest- average coefficient (.69). Whereas in some contexts the coefficients are relatively high (for example, .81 in applied research), in others they are much lower (.57 in teams specializing in physics and an extreme outlier of -.19 in company 2).[1]

Low reliabilities characterize the scale that encompasses all subjective evaluations of productivity (team members, team leader, and management evaluation (X = .43; S.D. = .24)), indicating that the three evaluations fail to agree about the level of the teams' productivity. Andrews II (patents, experimental prototypes, experimental materials) is another scale with a low average reliability (.43), and an even smaller standard deviation (S.D. = .18) than that of the subjective

[1] We treat the obtained reliability coefficients as drawn from a population of reliability coefficients.

evaluations scale. The scale proposed by Pelz and Andrews (published articles, patents, books, and reports) has the smallest standard deviation (S.D. = .10). Andrews III (position papers, algorithms, blueprints, and reports), a scale with properties similar to those of Pelz and Andrews, has a mean reliability coefficient of .57 and a standard deviation of .14.

The need for a context-specific definition of productivity may be demonstrated by examining the reliability coefficients of various scales across contexts. For example, scale Andrews I (project proposals, published articles, and books) is unreliable when used to measure productivity of R & D teams engaged in development projects, suggesting that project proposals, articles, and books are produced independently of one another by development teams, and hence cannot be regarded as stemming from the same domain of productivity. On the other hand, teams engaged in basic and in applied research show higher levels of uniformity with respect to proposals, articles, and books: those who produce more of one product produce more of the other two as well.

More striking differences are apparent when examining the reliability of this scale (Andrews I) across industries. Its coefficient ranges from .02 when used in the pharmaceutical industry to .72 in the aerospace industry. The obvious conclusion from this wide range is that Andrews I cannot be used as a productivity scale in certain industries (e.g., pharmaceutical, auto), whereas it may be a valid scale of productivity for R & D teams in other industries (e.g., aerospace, oil). Differences in reliability coefficients across industries are apparent on other scales as well: the reliability of Andrews II ranges from .02 in the oil industry, to .70 in the biotechnology industry. The scale labeled Andrews III yields a wide range of reliability coefficients when applied to different compan ies.

The scale composed of three subjective evaluations of productivity shows even larger variations across different contexts. Its reliability coefficient ranges from -.32 in biological sciences to .75 in company 2. The differences are particularly large across companies, industries, and scientific fields.

Discussion

This paper clearly demonstrates that scientific productivity is a context-specific construct which should be defined and redefined when shifting from one situation to another. Several measures of scientific productivity have been widely used in the study of R & D units and scientists. These measures are utilized across the board—regardless of the specific conditions under which scientific work is conducted.

The major conclusion to be drawn from the present study is that we cannot use the same scale of productivity in all situations. The variations in the reliability coefficients across industries, companies, disciplines, types of project, and structural compositions indicate that the validity of the productivity scales used varies as well.

The present study challenges in particular the indiscriminate use of scientific productivity scales. Furthermore, our findings raise questions as to the validity of the results regarding the correlates of productive scientific behavior obtained in empirical studies reviewed above.

The most unstable scale encountered was the combination of three evaluations of productivity (by manager, team leader, and team members). This is reflected by the relatively low mean and large standard deviation of its reliability coefficients. Lastly, the relatively high average reliability coefficient obtained for the scale composed of all products should be treated with caution, since such coefficients are partially determined by the number of items included in a scale.

We do not regard this study as a definitive tract. Rather, we propose to continue the search for appropriate measures of productivity in different contexts of activity.

Appendix
Indicators and Scales of Productivity

	Pelz and Andrews	Andrews I	Andrews II	Andrews III	All Tangible Products	All subjective Measures
Subjective measures:						
Members' evaluation*						
Leader's evaluation						
Management evaluation**						
Tangible products:***						
Position papers				x	x	
Project proposals		x			x	
Published articles	x	x			x	
Patents	x		x		x	
Books		x x			x	
Algorithms, blueprints				x	x	
Reports	x			x	x	
Experimental prototypes of devices			x		x	
Experimental materials			x		x	
Prototype computer programs					x	

* Members' evaluation is based on the average response of team members.

** A standardized score was calculated for each team within the scores assigned by the same evaluator.

*** The scores were reported by the team leader on a four category scale where:
"0" represents no product,
"1" represents one unit,
"2" represents 2-5 units of the product,
"3" represents more than five units of the product.

Table 1

Definitions, Means, and Standard Deviation of Productivity Indicators

Variable	Definition	Mean (S.D.)
Members' evaluation	A seven category scale where: 1 = lowest performance 7 = highest performance	5-3 (.70)
Leader's evaluation	A seven category scale where: 1 = lowest performance 7 = highest performance	5.7 (1.0)
Tangible products as specified below:	0 = none 1 = one 2 = 2-5 3 = more than 5	
Position papers		1.46 (1.22)
Project proposals		1.98 (1.08)
Published articles		1.41 (1.29)
Patents		1.50 (1.20)
Books		.13 (.47)
Algorithms		1.43 (1.43)
Reports		2.20 (.97)
Experimental prototypes		1.44 (1.33)
Experimental materials		1.12 (1.34)
Prototype computer programs		1.10 (1.27)

Note: Managers' evaluations of productivity are standardized within evaluators across teams and have a mean of zero.

Table 2

Definitions and Frequency Distribution of Contexts

Variable	Definition	N of Teams
Type of project	1 = basic research	46
	2 = applied research	58
	3 = development	116
Disciplinary composition	1 = monodisciplinary teams	58
	2 = interdisciplinary teams	165
Unit age	below or equal to 3 years	102
	above 3 years	89
Unit size	below or equal to 9 members	114
	above 9 members	109
Discipline	1 = biological sciences	12
	2 = chemistry	24
	3 = engineering	50
	4 = physics	26
Social structure	Type A = leader centered	24
	Type B = collegial	94
	Type C = autonomous	13
	Type D = mixed	55
Industry	1 = automotive	48
	2 = chemicals	32
	3 = electronics	46
	4 = aerospace	34
	5 = pharmaceutical	21
	6 = biotechnology	20
	7 = oil	22

Table 3

Reliability Coefficients Across Scales and Contexts
(number of teams in parentheses)

	All Tangible Items	All Subjective Measures	Andrews I	Andrews II	Andrews III	Pel and And:
Monodisciplinary teams	.71 57	.29 46	.53 57	.23 57	.59 57	.53 57
Interdisciplinary teams	.73 154	.53 118	.49 155	.49 156	.63 155	.60 155
Young units (3 years and below)	.71 100	.38 81	.39 10	.49 102	.63 101	.54 101
Old units (above 3 years)	.67 89	.43 64	.40 89	.33 89	.55 89	.43 89
Biological sciences	.68 12	-.32 10	.67 12	.72 12	.44 12	.73 12
Chemistry	.80 24	.58 17	.60 24	.44 24	.66 24	.66 24
Engineering	.73 48	.34 40	.52 49	.46 49	.61 48	.58 48
Physics	.57 25	.67 21	.54 25	.42 25	.56 25	.45 25
Small teams (9 members and below)	.75 104	.44 80	.61 104	.49 104	.63 104	.58 104
Large teams (above 9 members)	.67 107	.52 84	.37 108	.37 109	.53 108	.54 108
Type A	.75 24	.19 17	.55 24	.40 24	.65 24	.67 24
Type B	.74 91	.53 69	.47 92	.50 92	.61 91	.60 91
Type C	.82 11	.36 11	.71 11	.51 11	.51 11	.74 11
Type D	.59 50	.65 41	.45 50	.06 51	.49 51	.52 51
Basic research	.60 44	.51 34	.61 44	.08 44	.42 44	.45 44

Table 3 (continued)

	All Tangible Items	All Subjective Measures	Andrews I	Andrews II	Andrews III	Pelz and Andre
Applied research	.81	.47	.63	.55	.76	.70
	53	43	53	53	53	53
Development-	.74	.45	.27	.54	.62	.57
	111	84	112	113	112	112
Automotive industry	.73	.31	.35	.60	.62	.33
	43	37	43	43	43	43
Chemical industry	.78	.64	.46	.49	.68	.65
	31	19	31	31	31	31
Electronics industry	.70	.61	.43	.52	.66	.65
	13	37	43	43	43	43
Aerospace industry	.77	.45	.72	.54	.65	.64
	32	25	33	34	33	33
Pharmaceutical industry	.76	.57	.02	.54	.71	.64
	21	16	21	21	21	21
Biotechnology industry	.68	-.17	.56	.70	.37	.46
	20	14	20	20	20	20
Oil industry	.68	.59	.65	.02	.56	.71
	21	16	21	21	21	21
Company 1	.76	.45	.60	.32	.57	.58
	22	19	22	22	22	22
Company 2	-.19	.75	.54	*	.04	.34
	15	10	15		15	15
Company 3	.80	.10	.33	.45	.45	.58
	11	9	11	11	11	11
Company 4	.67	.70	.51	.25	.66	.54
	26	25	26	26	26	26
S	.69	.43	.50	.43	.57	.58
S.D.	(.18)	(.24)	(.15)	(.18)	(.14)	(.10)

* Data are not available for this cell.

Bibliography

Akin, G., and D. Hopelain (1986). "Finding the Culture of Productivity," Organizational Dynamics, Winter: 19-32.

Anderson, R. C., F. Narin, and P. McAllister (1978). "Publications Ratings Versus Peer Ratings of Universities," Journal of the American Society for Information Science.

Andrews, F. M. (ed.) (1979). Scientific Productivity: The Effectiveness of Research Groups in Six Countries, Cambridge, Cambridge University Press.

Blau, P. M. (1977). Inequality and Heterogeneity: A Primitive Theory of Social Structure, New York: The Free Press.

Campbell, J. P. (1976). "Psychometric Theory," in M. D. Dunnett (ed.), Handbook of Industrial and Organizational Psychology, Illinois: Rand McNally College Publishing Company.

Chubin, D. E. , A. L. Porter, F. A. Rossini, and T. Connolly (1986). Interdisciplinary Analysis and Research, Maryland: Lomond .

Cohen, B. P., V. R. Arechavala, D. Nobel, and Y. Shenhav (1 986). "Organization and Productivity in R & D Teams: A Report of Research Findings," Working Paper 86-4, The Center for Sociological Research, Stanford University, November.

Cohen, B. P., R. J. Kruse, and M. Anbar (1982). "The Social Structure of Scientific Research Teams," Pacific Sociological Review 25: 205-232.

Dalton, D. R., W. D. Todor, M. J. Spenolini, G. J. Fielding, and L. W. Porter (1980). "Organization Structure and Performance: A Critical Review," Academy of Management Review 5: 49-64.

Fox, M. F. (1983). "Publication Productivity Among Scientists: A Critical Review," Social Studies of Science 13: 285-306.

Goldberg, A., and Y. Shenhav (1984). "R & D Career Paths: Their Relations to Work Goals and Productivity," IEEE Transactions on Engineering Management 31: 111-117.

Gooding, R. Z., and J. A. Wagner (1985). "A Meta-Analytic Review of the Relationship Between Size and Performance: The Productivity and Efficiency of Organizations and Their Subunits," Administrative Science Quarterly 30: 462-481.

Guilford, J. P. (1954). Psychometric Methods, New York: McGraw Hill.

Koenig, M. E. (1982). "Determinants of Expert Judgment of Research Performance," Scientometrics 4: 361-378.

Koenig, M. E. (1983). "Bibliometric Indicators Versus Expert Opinion in Assessing Research Performance," Journal of the American Society for Information Science 34: 136-145.

Kornhauser, W. (1962). Scientists In Industry: Conflict and Accommodations, Berkeley: University of California Press.

Lord, F. M., and M. R. Novick (1968). Statistical Theories of Mental Test Scores, Massachusetts: Addison-Wesley.

Nunnally, J. C. (1978). Psychometric Theory (2nd ed.) , New York McGraw-Hill.

Pelz, D. C., and F. M. Andrews (1966). Scientists in Organizations, New York: John Wiley.

Quarshi, M. M. (1984). "Publication Rate as a Function of the Laboratory/Group Size," Scientometrics 6: 19-26.

Ritti, R. R. (1968). "Work Goals of Scientists and Engineers," Industrial Relations 7: 118-131.

Schwab, D. P. (1980). "Construct Validity in Organizational Behavior," pp. 3-43 in B. M. Staw and L. L. Cummings (eds.), Research in Organizational Behavior, vol. 2, Connecticut: JAI Press.

Scott, W. R., S. M. Dornbusc.h, B. C. Busching, and J. D. Laing (1967). "Organizational Evaluation and Authority," Administrative Science Quarterly 12: 92—117•

Shenhav, Y., and Y. Haberfeld (1988). "The Various Faces of Scientific Productivity: A Contingency Analysis," Quality and Quantity 22.

Shrum, W. (1985). "Quality Judgments of Technical Fields: Bias Marginality, and the Role of the Elite," Scientometrics 8: 35-57

Yankevich, W. F. (1982). "Analysis of Publication and Invention Productivity in Some Soviet Academic Institutions," Scientometrics 6: 431-437.