

Accuracy of Self Assessment and its Relation with some Psychological and Biographical Factors

S. Chatterji and Manjula Mukerjee

Indian Statistical Institute, Calcutta, India

The present study is an investigation of the validity of self assessment against two different criteria viz., the examiner's assessment and the assessment by peers, in highly competitive selection situations. It also envisaged to study the effects of certain psychological and biographical factors upon the quality of self assessment. Two groups comprising 230 and 187 subjects were studied in this connection. It was observed that: (a) a strong tendency to over estimation dominated self-appraisal; (b) self assessment was rather unpredictable against the two validation criteria i.e. the examiners' and peers' assessments though the two criteria were highly correlated; (c) the moderating effect of the psychological and biographical factors had little effect on the self assessment of older and more qualified subjects. However, socio-economic status, academic achievement, favourable family background, good schooling and urban background had positive moderating effects upon the quality of self reporting of younger and less qualified subjects.

That bias influences human judgement or rating of others has been well documented in research studies. Factors such as the assessor's inability to appraise correctly, lack of adequate and accurate information about the assessee and rater-ratee interaction adversely affect the accuracy of assessment. But in certain contexts, the use of human judgement is unavoidable and the need to improve its quality has motivated many researchers to explore ways of achieving this goal. Attempts have been made to reduce inaccuracy in rating by training raters in types of error (Bernardin, 1978; Ivancevich, 1979). Behaviourally anchored rating scales have been devised incorporating rules to be followed in the course of judgement (Sharon & Bartlett, 1969; Borman, 1979). These attempts have been fruitful in some cases; but contradictory results have also been reported (Holzbach, 1978).

In recent years there has been increasing interest in the use of self assessment. Some are of the opinion that even if it is not possible to do away with the subjectivity involved in the supervisor's assessment, it is possible to mitigate its impact by inclusion of self appraisal. Others (Wagner, 1973, 1974; Edward, Abram & Ronald, 1977) are more optimistic about the use of self assessment in place of expensive written examinations even

in selection and placement settings. However, self rating is highly subjective and susceptible to faking, and, questions may be raised about the individual's capability of making accurate judgements. Even if an individual is in the best position to make assessment about him or her self, it is likely that he/she may not like to disclose that accurately. Moreover, there may exist both internal and external factors, which induce a great deal of unwanted variance in self assessment. These important issues can be empirically studied. The present study is a venture in this direction.

The reliability and validity of self assessments have been investigated through a number of studies. Some of these are field studies of job or selection situations whereas others are laboratory experiments; some deal with assessment of abilities, skill and knowledge whereas others are related to performance variables and personality traits. Sister Amatora (1956), Wagner (1973), Ference (1975), Edward, Abraham and Ronald (1977) reported high validity for self ratings of abilities, skill, knowledge and personality variables against criteria of examination marks, supervisors' and peer ratings. However, other researchers such as Heneman (1974), Kilmoski and London (1974), DeNisi and Shaw (1977) obtained different results

and concluded that self ratings were not reliable. Thornton (1968) studied the self appraisal of a group of executives on their job performances and noted that these executives rated themselves more highly than they were rated by their supervisors. Those who were considered least promotable on the basis of a criterion measure of success in the organisation over-rated themselves most. Bartlett (1959) compared self ratings with peer ratings on a leadership behaviour scale and concluded that peer ratings seemed to be a good measure of all areas of leadership but self ratings were not.

The presence of inaccuracy in self appraisal has also motivated many researchers to study the impact and moderating effect of different cognitive and non-cognitive factors. The results of such studies are varied and inconsistent. DeNisi and Shaw (1977) reported that sex, general intelligence, self esteem, social desirability had no significant effect on self assessment under a selection setting. Deaux and Farris (1977) observed that males evaluated their performance more favourably than females. Safin (1975) studied the adequacy and stability of self-evaluation and came to the conclusion that the accuracy of self-evaluation depended on whether an individual was oriented towards self-evaluation and also on success in the activity. Johnson (1975) reported that subjects with high achievement motivation would be more likely to overestimate their success. Cusin (1972) reported that school pupils with experience of repeated academic failure, overestimated their performance equally with those with a better academic background.

The present study had two objectives: (a) to investigate the validity of self assessment against the rating made by the examiners and by the peers under a strong inducement toward distortion as would be encountered in a selection setting. (b) To study the moderating effect of some psychological and biographical variables upon the accuracy of self assessment. The hypotheses investigated were as follows: (i) Self reporting would have poor reliability as well as poor validity; (ii) in a competitive selection situation the subjects, in general, would inflate their self ratings as few would be willing to accept that their own performance was worse than that

of other competitors; (iii) some psychological and biographical variables would have significant moderating effect upon the self-appraisal.

Method

Sample

Two groups of applicants A and B served as subjects for this study. Group A comprised of 230 Engineering graduates, all holding First Class degree, who had applied for position as Graduate Trainee Engineers in a large private sector organisation. There were about 80 vacant posts and the applicants were aged between 22 and 28 years. Group B consisted of 187 undergraduate students seeking admission to a professional course in Hotel Management. The duration of the course was 3 years and the number of positions available was around 80. Subjects in Group B were aged from 16 to 20 years.

The individuals belonging to Group A and Group B, however, had to pass a selection test and on the basis of test performance they belonged to the top 25 per cent of the original group of applicants.

Procedure

The subjects were assigned to different batches and completed Group Task and Group Discussion. There were 29 batches in Group A and 20 in Group B. The number of subjects per batch varied from 8 to 12. In Group Task, each batch had to solve a problem as a group in a period of one hour. In Group Discussion they discussed for half an hour a specific topic provided by the examiner.

The nature and complexity of the problems and the topic of discussion varied between Group A and Group B due to the different academic levels of the subjects, but for batches in the same group the problems used were more or less parallel to each other.

Three examiners observed the participants during the entire one and a half hour period and rated them on a five-point scale which was behaviourally anchored. For successful completion of the task, planning, cooperation, initiative etc., by the participants were called for. Analysing the interaction that took place among the participants during this period, the examiners assessed them on traits such as Leadership, Cooperativeness, Mental Alertness, Planning Ability, Clarity of Communication and Initiative. Based on the performance in Group Discussion, the examiners rated the participants on aspects like Address, Articulation,

Participation, Logical Presentation of Ideas, Breadth of Knowledge and Coordination.

After the exercises were completed, the independent ratings of the three examiners were added together to obtain consolidated rating for each participant on the traits, and also on overall rating. The participants were ranked on the basis of these ratings.

At the end of each exercise the following instruction was given to the participants

"Now you have gone through the Group Task/Group Discussion and naturally have gained some idea regarding your own performance as well as that of others. Considering the role played by each member in understanding of the problem/topic, planning of the steps and procedure, guiding others to arrive at the correct solution, rank each member of your group including yourself. The purpose of this is to find out how objectively you can judge your own performance and that of others. Of course the ratings that you would give here, will not affect the judgement of the examiners."

The participants however were not asked to provide separate assessments for different traits as had been done by the examiners; it was felt that they would not be able to evaluate the performance of others on a variety of distinct dimensions as they had no prior training or experience in this field. Studies which have used multitrait ratings report that considerable halo effect operates on such ratings even when experienced raters pass their judgement on different scales (Chatterji and Mukerjee, 1974; Holzbach, 1978).

In addition to this, each subject completed a questionnaire which provided the following information.

Self

- (a) Selection test scores (mainly Aptitude Tests)
- (b) Past academic record (percentage of marks in the last University exam.)
- (c) Medium of instruction at school
- (d) Rural or Urban origin
- (e) Level of qualification (under or over-qualified etc., as per requirement)

Family

- (f) Educational level of the parents
- (g) Father's occupation
- (h) Economic status of the family (per capita income)
- (i) Number of brothers and sisters
- (j) Number of family members

The moderating effect of these factors upon self assessment was investigated. There are many more psychological or biographical fac-

tors whose moderating effect may be interesting to study. But to start with, two important psychological factors viz., Aptitude and Achievement were considered here. Information about these two aspects was already available. Similarly regarding biographical factors, those factors which were thought to have influence in character formation, in generating feeling of personal adequacy, in removing threat due to economic crisis etc., were taken into consideration.

Results and Discussion

As each subject ranked others in his batch, there were as many sets of ranks as the number of individuals in the batch. These ranks were then added to arrive at a consolidated peers' rank. Thus for each subject there were 14 ranks for 12 different traits and two overall assessments provided by the examiners; then there were the average of peers' ranks and finally self rank. The rank order correlations between all these ranks were calculated and the average values obtained for Groups A and B are presented in Table 1.

A high degree of agreement was observed between peers' ranking and examiners' ranking in both groups. However, it was somewhat higher in Group B (.85) than in Group A (.74). Self assessments, on the other hand, varied widely from these two sets of assessments. The size of correlations between self and examiners' overall rank varied in different batches from $-.15$ to $.81$ with a mean of $.33$ and that between self and peers' rank varied from $-.76$ to $.94$ with a mean of $.38$. This indicated relatively low validity of self assessment in relation to criterion measures.

The reliability of peers' and examiners' assessments was calculated separately for each batch using Ebels' method (1951) and it was observed that the average of the reliabilities of a single examiner's rank was of the order of $.80$, whereas that of the average of the three examiners' ranks was around $.95$. In the case of peers' assessments, the averages of the corresponding reliabilities were around $.40$ for a single peer and $.75$ for the summated peer rankings. This result indicated that too much reliance should not be placed on a single peer rank, but that as the average of peers' ranks possessed considerably high reliability, it may be profitably utilised. These results have shown a grati-

Table 1: *Average of Rank Order Correlations between Examiners' Rank, Peers' Overall Rank and Self Rank.*

Examiners' Rank on different traits	Group A		Group B	
	Average of Correlations for 29 batches with		Average of Correlations for 20 batches with	
	Self Rank	Peers' Rank	Self Rank	Peers' Rank
1. Mental Alertness	.29	.66	.30	.82
2. Cooperativeness	.31	.71	.28	.81
3. Ability to plan	.30	.70	.29	.81
4. Application	.33	.77	.33	.82
5. Leadership	.33	.74	.33	.84
6. Communication	.34	.73	.32	.81
7. Consolidated rank for Group Task	.32	.75	.34	.86
8. Address/Manners	.26	.72	.36	.80
9. Articulation	.27	.65	.40	.83
10. Participation	.29	.79	.38	.83
11. Logical presentation	.21	.57	.42	.78
12. Breadth of knowledge	.23	.66	.39	.79
13. Coordination	.28	.75	.38	.82
14. Consolidated rank for Group Discussion	.31	.72	.36	.84
Average of Rank Correlations between self rank and peers' rank in				
	Group A		Group B	
Group Task	.39		.41	
Group Discussion	.35		.38	

fyng similarity to those reported by Berkshire and Nelson (1958), Wherry and Fryer (1949) and Suci and Vallance (1954). In this connection Richard (1962) pointed out that "Peer ratings i.e., evaluation of individuals in a group by one or more individuals in that group, though made by untrained observers are good predictors of relative success or failure."

From the results it may be summarised that in the two exercises, assessors, whether participants or observers, could arrive at a common conclusion about participants' standing on relevant personality traits. However, self assessments were less accurate. Some candidates apparently appraised their own performance objectively and reported it without distortion; others either failed to evaluate their own performance accurately or reported inflated self assessments. It should be noted that the examiners' rankings and peers' rankings were based on a number of independent observations whereas self ranking was a single measurement and hence likely to have lower reliability.

The correlations between the examiners' ranks on different traits and peers' overall

rank were relatively high whereas the corresponding values obtained with overall self assessment were low. Most of the correlations were of a similar order, indicating that neither self ranking nor peers' ranking was differentially related with the examiners' assessment on different attributes. This might have been due to the presence of halo error in examiners' assessment, or because different behaviours represented in the scales used by the examiners were strongly correlated.

Correlations between peer assessments and examiners' ratings showed some variation in Group A according to the attributes considered but in Group B the values remained more or less constant (Table 1). However, such variation was absent with respect to self assessment rejecting the possibility of the existence of differential relationships with the criteria used by the examiners.

Using the examiners' assessments as base points, the self rankings were classified into the following categories:

- (a) Under Ranking
- (b) Identical Ranking
- (c) Over Ranking

Table 2: *Percentage distributions of self assessment in different categories for Group A (N = 230) and Group B (N = 187).*

	Under Ranking		Identical Ranking		Over Ranking	
	Group A	Group B	Group A	Group B	Group A	Group B
Group Task	14	16	12	13	74	71
Group Discussion	14	16	10	10	76	74

The percentage distributions of self ranks in these three categories are presented in Table

2.

Almost identical distributions were obtained for Groups A and B. About 75% of the subjects over-ranked themselves when the corresponding values were 10 and 15 respectively for identical and under-ranking categories. Irrespective of actual performance in these exercises, the subjects tended to put themselves in the first or second or third position in the rank list provided by them.

In order to study the moderating effects of the factors such as level of academic achievement, level of aptitude as measured by selection tests, parents' educational level, father's occupation etc., upon the accuracy of self-reporting the following procedure was adopted. The subjects were divided into sub-

groups on the basis of each of the factors mentioned earlier. For example, the subjects were divided into equal High and Low sub-groups according to level of achievement in the last Degree Examination in case of Group A and in the last School Leaving Examination in case of Group B. Factors like "father's occupation", "rural or urban background" etc., were assessed on the basis of obtained response categories. "Medium of instruction at school" was classified into two: (a) English Medium, (b) Non-English Medium. These classifications could then be related to the nature of self-rankings i.e., under or identical or over-ranking. Chi-square test was applied and Coefficient of Contingency was calculated to find out the significance and size of the relationships. These results are presented in Table 3 and 4.

Table 3: *Chi-square values indicating relationship between the quality of self assessment and different psychological and biographical factors (Group A).*

Factors	Group Task			Group Discussion		
	χ^2 value	D.F	Contingency Coefficient	χ^2 value	D.F	Contingency Coefficient
(a) Level of Aptitude	7.44	8	0.18	9.54	8	0.20
(b) Level of Academic Achievement	0.51	4	0.05	6.45	8	0.17
(c) Medium of Education	6.96*	2	0.18	5.43	2	0.16
(d) Rural or Urban background	0.43	2	0.04	0.59	2	0.05
(e) **	—	—	—	—	—	—
(f) Level of Parents' Education	20.38	16	0.29	18.80	16	0.28
(g) Father's Occupation	3.68	8	0.13	6.88	8	0.18
(h) Level of Family Income	5.03	4	0.15	6.60	4	0.17
(i) Number of brothers and sisters	7.01	4	0.17	3.67	4	0.13
(j) Family members	2.17	4	0.10	4.05	4	0.13

* Significant at the 5% level.

** As only a few over-qualified subjects were in Group A, no analysis was performed on them.

Table 4: Chi-square values indicating relationship between the quality of self assessment and different psychological and biographical factors: (Group B).

Factors	Group Task			Group Discussion		
	χ^2 value	D.F	Contingency Coefficient	χ^2 value	D.F	Contingency Coefficient
(a) Level of Aptitude	4.7	4	0.16	4.18	4	0.15
(b) Level of Academic Achievement	12.93**	2	0.25	12.96**	2	0.25
(c) Medium of Education	15.27**	2	0.28	10.88**	2	0.23
(d) Rural or Urban background	10.57**	2	0.23	8.32*	2	0.21
(e) Qualified or over-qualified	0.53	2	0.05	7.45*	2	0.20
(f) Level of Parents' Education	14.64	16	0.27	16.25	16	0.29
(g) Father's occupation	13.33	8	0.26	12.97	8	0.26
(h) Level of Family Income	14.48*	4	0.27	10.27*	4	0.23
(i) Number of brothers and sisters	10.67*	4	0.23	9.20*	2	0.22
(j) Family members	6.53	4	0.19	3.80	4	0.15

* Significant at the 5% level.
 ** Significant at the 1% level.

Table 5: Percentage distributions of Self ranks in Under, Identical, Over ranking categories and different psychological and biographical factors. (Group A and Group B).

Group A		Group Task			Group Discussion		
		Under Ranking	Identical Ranking	Over Ranking	Under Ranking	Identical Ranking	Over Ranking
Medium of Instruction	English	15	19	66	12	16	72
	Non-English	15	8	77	12	7	81
Group B							
I. Academic Level	High	50	17	33	50	11	39
	Low	16	12	72	15	9	76
II. Medium of Instruction	English	24	19	57	28	12	60
	Non-English	11	6	83	9	10	81
III. Original place of living	Rural	7	0	93	4	4	92
	Urban	22	15	63	24	11	65
IV. Level of qualification	With minimum required	19	12	69	25	7	68
	Over qualification	16	15	69	11	15	74
V. Income Group (per month)	Up to Rs.100/-	0	6	94	10	5	85
	Rs.101/- to 300/-	20	7	73	16	5	79
	Rs.301/- and above	26	19	55	24	17	59
VI. Number of brothers and sisters	Up to 2	22	20	58	24	10	66
	3 to 5	19	10	71	13	12	75
	More than 5	8	4	88	13	0	87

The degree of association between different factors and the nature of self assessment varied from group to group. In Group A only "medium of instruction at school" was found to be significantly related whereas in Group B significant relations were established for several factors: Academic Achievement, Economic Status of the Family, Number of brothers and sisters, Medium of instruction at school, Level of qualification, Rural or

Urban background. Percentage distributions of self rankings for factors which were found to be significantly related to nature of self assessment are presented in Table 5. In Group A fewer subjects trained in English Medium schools over-ranked themselves than those who studied in Non-English Medium schools. In Group B over-ranking was found to be greatest among low academic achievers, those from large families, those in non-

English medium schools, those from rural backgrounds and those of lower income groups.

This study suggests that a strong tendency to over-estimation dominates self-appraisal. This result is consistent with the findings of other research work and supports the second hypothesis of the study. In a selection setting it was observed that about 75% of the subjects over-estimated their performance. More or less identical results were obtained for both the groups though the group composition varied in both age and academic level. However, there were some persons in both the groups who evaluated their own performance as lower than did their examiners.

The relationships between self rank and examiner's rank or that between self rank and peers' rank was unpredictable or erratic in nature. It indicated the low validity of self appraisal and supported the first hypothesis of the study. The intercorrelations and reliabilities of the two validation criteria i.e., the examiner's ranks and peers' ranks were however relatively high.

The findings of the study partly confirm the third hypothesis. Variations in psychological and biographical factors had little effect on the self assessments of older, better educated subjects. However, socio-economic status, academic achievement, favourable family background, good schooling and urban background had positive moderating effects upon the quality of self reporting of younger, less qualified subjects.

References

- Bartlett, C. J. The relationship between self ratings on a leadership scale. *Personnel Psychology*, 1959, 12, 237-246.
- Berkshire, J. R. & Nelson, P. D. Leadership peer ratings related to subsequent proficiency in training and in the fleet. Special Report, 58-20, Pensacola, Fla: *Naval School of Aviation Medicine*, 1958.
- Bernardin, H. J. Effects of training on leniency and halo errors in student ratings of instructions. *Journal of Applied Psychology*, 1978, 63, 301-308.
- Borman, W. C. Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, 1979, 64, 410-421.
- Chatterji, S. & Mukerjee, M. Group task and its use as a measure of personality in selection situation. *Japanese Psychological Research*, 1974, 16, 22-28.
- Cusin, P. & Piolat, Michael. Self evaluation of performance: some scholastic and intellectual determinants. *Cahiers de Psychologie*, 1972, 15, 43-57.
- Deaux, K. & Farris, Elizabeth. Attributing causes for one's own performance. The effects of sex, norms and outcome. *Journal of Research in Personality*, 1977, 11, 59-72.
- DeNisi Angelo, S. & Shaw, J. B. Investigation of the users of self-reports of abilities. *Journal of Applied Psychology*, 1977, 62, 641-644.
- Ebel, R. L. Estimation of reliability of ratings. *Psychometrika*, 1951, 16, 407-424.
- Edward, L. L.; Abram, Flory III & Ronald, A. A. Self assessment in personnel selection. *Journal of Applied Psychology*, 1977, 62, 428-435.
- Ference, L. W. Self-reported scores as predictors of test results of clerk, stenographer I/clerk typist, Milwaukee Wise; *City of Milwaukee Civil Service Commission*, 1975.
- Heneman, H. G. Comparisons of self and superior ratings of managerial performance. *Journal of Applied Psychology*, 1974, 59, 638-642.
- Holzbach, R. L. Rater bias in performance ratings: superior, self and peer ratings. *Journal of Applied Psychology*, 1978, 63, 579-588.
- Ivancevich, J. M. Longitudinal study of the effects of rater training on psychometric error in ratings. *Journal of Applied Psychology* 1979, 64, 507-508.
- Johnson, P. B. Achievement motivation and self-reported grade point average. *Psychology in the schools*, 1975, 12, 402-404.
- Klimoski, R. J. & London, M. Role of the rater in performance appraisal. *Journal of Applied Psychology*, 1974, 59, 445-451.
- Parker, J. W.; Tayler, E. K.; Barrett, R. S. & Martens, L. Rating scale content III. Relationship between supervisory and self ratings. *Personnel Psychology*, 1959, 12, 49-63.
- Richard, E. Doll & Alexander, A. L. Improving the predictive effectiveness of peer ratings. *Personnel Psychology*, 1962, 15, 215-220.
- Safin, V. F. The stability of self-appraisal and the mechanism of its preservation. *Voprosy Psichologie*, 1975, 3, 62-72.
- Sharon, A. T. & Bartlett, C. J. Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology*, 1969, 22, 251-263.
- Sister Mary Amatora, O.S.F. Validity in self evaluation. *Educational Psychological Measurement*, 1956, 16, 119-126.
- Suci, G. J. & Vallance, T. R. An analysis of peer ratings: II: Their validity as predictors of military aptitude and other measures in the Naval Officer Candidate school. *U.S.N. Bureau of Naval Personnel Technical Bulletin*, No. 54-10, 1954.
- Thornton, G. C. The relationship between supervisory and self appraisal of executive performance. *Personnel Psychology*, 1968, 21, 441-456.

- Touchey, J. A symbolic interactionist approach to self referent behaviour. *Psychological Reports*, 1971, 29, 87-90.
- Wagner, D. Improving clerical selection. Washington D.C.: U.S. Civil Service Commission, *Personnel Research & Development Centre*, 1973.
- Wagner, D. Typing speed tests, Sacramento, Calif. *Higher Education Personnel Board*, 1974.
- Wherry, R. J. & Fryer, D. H. Buddy ratings: popularity contest or leadership criteria? *Personnel Psychology*, 1949, 2, 147-159.