

California State University, San Bernardino

CSUSB ScholarWorks

Theses Digitization Project

John M. Pfau Library

1990

A comparison of perceived credibility of evaluation sources in university organizations

Justina Diaz Cruz

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/etd-project>



Part of the [Higher Education Commons](#)

Recommended Citation

Cruz, Justina Diaz, "A comparison of perceived credibility of evaluation sources in university organizations" (1990). *Theses Digitization Project*. 763.

<https://scholarworks.lib.csusb.edu/etd-project/763>

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

A COMPARISON OF PERCEIVED CREDIBILITY
OF EVALUATION SOURCES IN UNIVERSITY ORGANIZATIONS

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology

by
Justina Diaz Cruz
June 1990

A COMPARISON OF PERCEIVED CREDIBILITY
OF EVALUATION SOURCES IN UNIVERSITY ORGANIZATIONS

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

by
Justina Diaz Cruz

June 1990

Approved by:


Matt L. Riggs, Ph.D., Chair, Psychology

7-3-90
Date


Janet L. Kottke, Ph.D., Psychology


Clark Molstad, Ph.D., Management

ABSTRACT

Sources of performance evaluation was the focus of this study. Student and faculty evaluations of instructor performance, labelled source A and B respectively, were compared on the basis of the ratee's perception of each sources' level of credibility. It was hypothesized that source A would be perceived as being more familiar with performance than source B. It was also hypothesized that source B would be rated more trusting and expert than source A. Instructors from a local university were solicited to participate in the research. A total of 133 questionnaires were completed and returned to the researcher. Of these subjects who chose to identify their gender, 50 were male and 45 were female. Seventy percent of the respondents were full-time instructors and 13% were part-time. Support was found for the hypothesis predicting that faculty would be rated more expert than students, $t(121) = 11.23, p < .05$. However, students were not perceived as being more familiar with performance than faculty. A multiple regression analysis was used to determine which of the three variables would best predict overall credibility. Limitations of this study and implications for future research are discussed.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee Matt L. Riggs, Janet L. Kottke, and Clark Molstad for being there when I needed their help and guidance.

I would also like to thank my friends and family for their concern and support especially these past two years. I would have never come this far without them.

Special thanks go to my dearest and most cherished friend John C. Terlaje, who was always there to pick me up when I was ready to quit. You are the best!

Finally, to my mom and dad who permitted me to pursue an education here in the mainland many miles away from home. I thank them both for their unceasing love and guidance. I hope and pray that my accomplishments have made them proud.

TABLE OF CONTENTS

Title page.....	i
Signature.....	ii
Abstract.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
List of Tables.. ..	vii
Introduction.....	1
Sources' Level of Familiarity, Expertise, and Trust.....	10
Student Evaluations.....	11
Perceived Credibility of Student Evaluations.....	13
Comparison of Levels of Familiarity, Expertise, and Trust.....	15
Hypotheses.....	17
Pilot Study Method.....	19
Justification of Pilot Study.....	19
Subjects.....	19
Procedure.....	19
Analysis.....	19
Pilot Study Results.....	21
Pilot Study Summary.....	27
Thesis Study Method.....	28
Subjects.....	28
Procedure.....	28

Analysis.....	29
Thesis Study Results.....	30
Factor Analysis.....	30
Student's t-Test.....	36
Multiple Regression Analysis.....	37
Discussion.....	40
Appendix A. Thesis Study Questionnaire.....	45
Appendix B. Thesis Study Demographic Information.....	48
References.....	52

List of Tables

Table 1. Pilot Study Eigenvalues before Rotation.....	21
Table 2. Pilot Study Factor Loadings after Rotation.....	22
Table 3. Reliability Analysis of Pilot Study (Factor 1: Trust Scale).....	24
Table 4. Reliability Analysis of Pilot Study (Factor 2: Task Expertise).....	25
Table 5. Reliability Analysis of Pilot Study (Factor 3: Familiarity).....	26
Table 6. Thesis Study Eigenvalues before Rotation.....	31
Table 7. Thesis Study Factor Loadings after Rotation.....	32
Table 8. Reliability Analysis of Thesis Study (Factor 1: Ability to Rate).....	34
Table 9. Reliability Analysis of Thesis Study (Factor 2: Task Expertise).....	35
Table 10. Reliability Analysis of Thesis Study (Factor 3: Familiarity).....	36

INTRODUCTION

Measures of performance have long been a concern for I/O psychologists. Jewell (1985) has suggested that performance appraisal is one of the more salient activities of organizations. By definition, performance appraisal is "a process of evaluating performance and providing feedback on which performance adjustments can be made" (Schermerhorn, Hunt, and Osborn, 1988).

In a formal performance appraisal situation, supervisory evaluation of the subordinate's performance is usually a paper and pencil rating scale. Such a scale requires the evaluator to rate the individual being appraised on the quality and/or quantity of work produced. This performance measurement provides data for determining salary increases, promotion, retention, and tenure. Performance evaluation plays an important role in any organization. Latham and Wexley (1982) suggested that performance evaluation becomes an integral part of "bringing about and maintaining satisfactory performance on the part of the individual employee". The appraisal attempts to measure the employee's effectiveness on the job. It also points out the areas in which the employee needs improvement or more training. In addition to these roles, performance evaluation strives to maintain high levels of motivation through feedback and goal setting based on this feedback

(Latham and Wexley, 1982). Performance evaluation can be conducted by multiple sources such as a subordinate, supervisor, or a peer. The use of multiple sources for performance appraisal has gained considerable acceptance in the past two decades according to Harris and Schaubrouk (1988). In previous research, supervisor, peer, and self evaluations were the primary focus of comparison in the performance evaluation literature (for example Baird, 1977; Herold, Liden and Leatherwood, 1987; and Larson, 1986).

This study focuses on the performance evaluation of the university instructor. Student evaluations of instructor performance will be the focus of this investigation and examined as subordinate evaluations in reference to peer and supervisor evaluations. The supervisor source will be the department head, while the peer source will be the colleague placed with the responsibility of evaluating the instructor's performance during a prearranged class visitation.

Performance Appraisal Feedback

Latham and Wexley (1982) stated that the appraisal process includes the evaluation of past performance of an employee, feedback that assists an employee in understanding how well he/she is doing, and goal setting that specifically points out those activities which the employee should be involved in, in order to be rewarded accordingly.

Latham and Wexley (1982) suggested six different sources of appraisal (a) the supervisor, (b) the employee, himself/herself, (c) peers, (d) subordinates, (e) a person or persons outside the employee's work environment, or (f) some combination of the above. Because this study is concerned only with the supervisor, peer, and subordinate sources of evaluation, this review of the literature will specifically address the three above-mentioned sources.

To qualify as a source Latham and Wexley (1982) offered some key criteria such as being familiar with the objectives of the employee's job, observing the employee on the job regularly, and being able to decide whether or not the behavior observed is satisfactory. These criteria should be met before a potential source of evaluation is chosen in order to ensure that the performance appraisal will be valid and effective.

Supervisor Evaluations. Larson (1989) has suggested that "in general, supervisors are an important source of performance information for their subordinates". Besides supervisors giving feedback on an employee's performance, they are usually the individuals responsible for administering the rewards for satisfactory performance (Latham and Wexley, 1982). Based on these premises, it would seem reasonable to deduce that employees will regard

their supervisors' evaluation of their performance as important. It must be realized though, that they are not always as reliable and valid as peer evaluations. Latham and Wexley (1982) noted that alternative sources to supervisory evaluations are necessary in the event supervisors are unable to observe the employee on the job.

Peer Evaluations. As an alternative or supplement to supervisory ratings, peer evaluations are considered highly reliable and valid. Korman (1968), as cited in Latham and Wexley (1982), concluded that peer evaluations are good predictors of performance. Such a conclusion is quite significant, primarily when the evaluation is utilized to determine promotions (Latham and Wexley, 1982).

A potential problem exists unfortunately with peer ratings. From an organizational standpoint, the issue of competition for the available rewards such as promotions, may in fact turn peer evaluations into a conflict among employees. Despite the potential competition, employees are sometimes reluctant to rate their peers believing that appraisals are a way in which the organization encourages "snitching" on each other (Roadman, 1964, as cited in Latham And Wexley, 1982).

Subordinate Evaluations. The subordinate source of

evaluation has been viewed as valuable to both the employee and the organization. Various positive results have evolved through this process. The subordinates are given the opportunity to observe problems as though they were supervisors, and in turn, the supervisors are given the opportunity to see the concerns of their employees through the eyes of a subordinate. One outcome of this process is increased productivity, namely group productivity, and another outcome is increased job satisfaction (Latham and Wexley, 1982).

Although subordinate evaluations are considered valuable to the organization as a whole, there is a problem with the employees perceiving the evaluation process as "threatening" (Latham and Wexley, 1982). Anonymity is therefore crucial in this case. In order to restore some ease in the subordinates who feel as though they may be chastised by their superiors for honest and unfavorable evaluations, no names are given on any forms included in the appraisal process (Latham and Wexley, 1982).

These sources of evaluation have clear advantages and disadvantages. Yet, there exists no perfect system of performance evaluation. Latham and Wexley (1982) have illustrated the way in which the different sources of evaluation stand up against each other. But one important aspect that has not been addressed is the issue of

credibility. Herold, Liden, and Leatherwood (1987) suggest that although performance feedback has been previously researched, there is a lack of generalizable results due to the fact that researchers have not paid the attention needed to assess the differences in the sources of performance evaluation with regard to the recipient's perceived credibility and acceptance of performance feedback. If a source of performance appraisal feedback is not perceived as credible, then the feedback may be rendered ineffective.

Credibility of Source. As defined in various dictionaries, credibility is an adjective meaning believable. Credibility of source in the appraisal process has not been researched extensively. However, Ilgen, Fisher, and Taylor (1979) in their review of feedback, have contributed a substantial amount of information concerning source credibility.

Ilgen et. al. (1979) emphasized the aspects of feedback that are purported to influence its perception, its acceptance, and the recipient's intent to respond to the feedback. They introduced the idea that the different sources of evaluation carry varying levels of power with the recipient. Along the same lines, researchers such as Klimosky and London (1974) and Zammuto, London, and Rowland (1982) have asserted that the sources at different levels

weight performance dimensions dissimilarly, thus suggesting that these sources may disagree on the overall rating since their definition and measurement of performance will vary.

Recipients of feedback consider good performance evaluations as those which are done by individuals who possess a certain amount of expertise and trust. These two dimensions are extrapolated from Giffin's (1967) study where five dimensions of source credibility are identified--expertise, reliability, intentions toward the listener, dynamism, and personal attraction. Ilgen et. al. (1979) indicated that, in general, source characteristics which influence the acceptance of feedback do so by influencing the perceived credibility of the source. Furthermore, they stated that when the recipient considers the source to be credible there is an increased likelihood that the feedback will be accepted. The two dimensions--expertise and trustworthiness--influenced acceptance more than reliability, dynamism, and personal attraction (Ilgen et. al.).

Ilgen, Fisher, and Taylor (1979) posited that raters ought to be perceived by the recipients of the feedback as possessing the expertise necessary to judge their performance accurately. Their definition of expertise included task familiarity; the ability to supply pertinent information for the improvement of the feedback recipient's

performance on certain tasks being evaluated. Additionally, they stated that expertise encompasses the idea of familiarity with the recipient's performance on the tasks. Although this is indicated in Ilgen, Fisher, and Taylor's research, familiarity with performance in this study is suggested as a separate factor independent of expertise, affecting perceived credibility. It is suggested by Latham and Wexley (1982) that observer accuracy be increased in order to facilitate improved and effective evaluations. The rater who would be considered the primary observer of performance in the classroom, in this case, is the student. This concept is substantiated by Overall and Marsh (1982) who recommended that student evaluations not be overlooked since it is the student who has rare opportunities to observe teaching. It would seem safe to assume that this interaction between student and teacher will lead to the development of a certain amount of familiarity with the instructor's performance in class. The distinction between task familiarity (ie. expertise) and familiarity with performance on the task will hopefully give us two discrete factors that will affect perceived overall credibility. A pilot study on the questionnaire to be used in this research will determine whether or not it is possible to measure the two factors separately.

Additionally, the level of trust as perceived by the

recipient must also be taken into consideration as another factor that affects credibility (Ilgen et. al., 1979; O'Reilly and Anderson, 1980). The three factors being examined--familiarity with performance, expertise (task familiarity), and trustworthiness are addressed below.

Familiarity. For the purpose of this study, familiarity will be defined as the ability to evaluate performance based on direct observation of tasks involved in a specific job, preferably defined in a job analysis. Cusella (1982) alluded to the idea that those individuals far removed from the work setting, were unfamiliar with the relevant job or work unit. For example, supervisors and managers who have assistants, usually leave to their assistants the responsibility of handling employees on a daily basis. Supervisors and managers would at times rely on their assistants to relay information to them dealing with problems in the work place. Given this example, supervisors and managers would not be expected to be very familiar with the employee's performance on the job.

Expertise. Previous research has shown that expertise may be affected by the source's level of training, experience, and ability. Heppner and Handley (1982) supported this notion in their study stating that those

trainees who perceived their supervisors as engaging in evaluative supervisory behavior, were more likely to perceive their supervisor as more expert, attractive, and trustworthy. Further support can be found in Holzbach's (1978) research in rater bias. He maintained that because supervisors possess more experience and responsibility in evaluating job performance, they would probably have greater sensitivity and awareness to particular job related behavior for the individual being rated. as well as, their subordinates in general.

Trustworthiness. Ilgen et. al. (1979) defined trust as the source's intentions toward the recipient, specifically speaking, it is the recipient's belief in his/her peers' motives as being consistent with the feedback peers offer. Trust in relationships, including professional ones, such as those found between academicians is important. An established trust between individuals leads to a more productive environment in an organization. O'Reilly and Anderson (1980) pointed out in their research, that trust serves as a moderator of the communication of information. Without trust, performance feedback would be distorted and likely to be taken offensively or perceived inaccurately.

Sources' Level of Familiarity, Expertise, and Trust

Keeley (1977) offered an explanation for the differences among subordinate, supervisor, and peer ratings. He indicated that "each rater occupies a different vantage point vis-a-vis the ratee", suggesting that the individual observations made by the sources will result in different evaluations made on the ratee's performance.

As indicated earlier, this study will be looking at the student as the subordinate source of evaluation, the department chairperson as the supervisory source, and the colleague responsible for evaluating the instructor during a class visitation as the peer source.

Further discussion on the sources' levels of the three factors affecting perceived credibility will come later in the review.

Student Evaluations

Previous research has looked at student evaluations in reference to their reliability and validity.

Validity. Whitely and Doyle (1979) in their article, examined the validity and generalizability of student ratings. They referred to several other studies which used the validity of student evaluations as indices of student learning such as Remmers, Martin, and Elliot (1949), Elliot (1950), Rodin and Rodin (1972), Frey (1973), Sullivan and Skanes (1974), and Doyle and Whitely (1974). Their findings

indicated high negative to high positive correlations, "with the majority indicating a statistically significant but very modest relationship between student ratings and tested student learning" (Whitely & Doyle).

Kurz, Mueller, Gibbons, and DiCataldo (1989) referred to a meta-analysis of multi-section validity studies done by Cohen (1981) which lended further support for the validity of student evaluations. Administrators in general, assume it is the best way to assess a teacher's performance in the class in order for them to decide promotion, retention, and tenure. Research done by Ulanski (1987) affirmed this assumption when he wrote "that with increasing fiscal belt-tightening in higher education and the decline in college-age students, tough decisions are being made with regard to faculty tenure, promotion and retention. He added that student evaluations often play a predominant role in faculty advancement.

Reliability. Kurz et. al. (1989) reported in their study that student evaluations are viewed as being reliable across items on evaluation forms, among multiple raters, and at different points in time in the same course or the same type of course. Unfortunately, these views are not shared by all who have researched student evaluations.

Reported inconsistencies in the student evaluation

literature suggest that they may be viewed continually as being unreliable and invalid on the basis of the reported low reliability and validity coefficients. Yet, there are other researchers such as Cohen (1981) and Gessner (1973) who posited that student ratings of instruction are a valid indication of instructor performance and effectiveness. There will be continued disagreement in the performance appraisal literature dealing with the various sources of evaluation.

In reviewing this aspect of student evaluations, the Rodin and Rodin study (1972) which indicated that students are unable to judge teaching effectiveness, was the single most negative article that implied using alternative methods of appraising instructors. Otherwise, researchers have challenged Rodin and Rodin's stand on this evaluation concern, therefore, suggesting that the use of student evaluations should continue in the university organization.

Perceived Credibility of Student Evaluations

There are currently a considerable number of studies which have looked at the reliability and validity issue of performance evaluations. Yet, there is one other issue needing more attention.

Perceived credibility of student evaluations has not been researched as extensively as the reliability and

validity issue. Therefore, the present study will concentrate on the perceived credibility of student evaluations in reference to supervisor and peer evaluations.

Feedback based on student evaluations should lead to the improvement of teaching quality as reflected in student evaluations of course performance (Morsch, Burgess, and Smith, 1956). Yet, if students are not perceived as credible sources of evaluation of teaching effectiveness, then the evaluation will be rendered useless to the instructor in terms of obtaining acceptable feedback deemed conducive to improving his/her performance.

Following along these lines, Shrauger and Lund (1974) stated that if the feedback received is not seen as credible, suggested adjustments to improve performance will be ignored. They stated that when the feedback received is not acceptable, it could be viewed in one of two ways. One way is to avoid using the feedback therefore not using it in any constructive critical manner. The second is to misinterpret the information received and use this knowledge in a way not intended by the rater. Shrauger and Lund (1974) went on to say that the validity of the feedback could possibly be undercut if the source was to be questioned. When the source of evaluation is perceived as credible, it is more likely that feedback based on the performance appraisal be accepted. This may hold true for

student and faculty evaluations.

Comparison of Levels of Familiarity, Expertise, and Trust

Students are probably most familiar with the instructor's performance since they interact with the instructor on a regular schedule. Dowell and Neal (1982) suggested that in some situations students may in fact be quite accurate in their ratings as proposed by other studies reporting high validity coefficients of student evaluations. Some instructors see student evaluations as being most informative of their performance in class. It is not dismissed as unnecessary to upcoming evaluations. Because student evaluations in some cases are considered accurate, it might follow then that the student source of appraisal would be perceived as being more credible than other sources of performance evaluation.

Besides familiarity, expertise is seen as necessary in order for the source to be perceived as credible. Ilgen et. al. (1979) redefined expertise as task familiarity. The department chairperson is considered to be the expert in this situation since most department chairs have served as teaching faculty prior to being administrators (Knight and Holen, 1985). Department chairs, as well as, peers are more aware of issues that may affect an instructor's teaching quality. Matters such as resource acquisition for teaching,

research, and/or service at the university, are issues better dealt with by those individuals such as the department chairperson and other faculty rather than by students.

Trustworthiness in the rater can be viewed as being more characteristic of peer sources of evaluation than of subordinate sources. Good performance evaluations should be based on the premise that all suggestions to adjust performance be unbiased, as well as, uncontaminated by stakes held by the different sources of evaluation. O'Reilly and Anderson (1980) stated that "if the relationship between the rater and the ratee is characterized by low trust, accurate feedback may not be perceived as accurate and therefore, not useful". Consequently, if the relationship is characterized by high trust, then the feedback source will be perceived as credible and acceptable.

Latham and Wexley (1982) have suggested that peers when compared to other sources of evaluation, have a more comprehensive picture of an employee's performance on the job. In line with this idea, instructors would see their colleagues as being knowledgeable of their responsibilities as an academician, and therefore trust their feedback more than they would other sources of evaluation. Yet, it must be noted that peers subsequently evaluate an instructor's

total performance and are not present in class where the teaching activity is primarily observed by students .

Student evaluations may be contaminated by the interest held in terms of the grades to be received in a particular class. If students are assigned an unsatisfactory grade, an unsatisfactory evaluation may result. On the other hand, if satisfactory grades are assigned, then a satisfactory evaluation may consequently result.

Despite the fact that faculty members possess the knowledge oftentimes unknown to students, there is the problem of their absence in the classroom. They do not have the luxury of spending time with their faculty peers as do students who interact with the instructor in a unique setting.

Hypotheses:

It was the purpose of this study to investigate these two sources of instructor evaluation: a) student, b) peers. These sources will be labelled source A and B respectively. The following hypotheses are generated:

Hypothesis 1: It is predicted that evaluation source B will be perceived as possessing higher levels of expertise than source A.

Hypothesis 2: It is predicted that evaluation source A will be perceived as being more familiar with the

instructor's performance than source B.

Hypothesis 3: It is predicted that evaluation source B will be perceived as being more trustworthy than source A.

Hypothesis 4: Finally, it is predicted that perceived familiarity with instructor performance will account for the most variance in perceived overall credibility followed by expertise and trustworthiness.

PILOT STUDY METHOD

Justification of Pilot Study

It was a necessary step to run the pilot study on the questionnaire since it was not an established measurement of perceived credibility. The internal consistency of the items was needed, as well as, the determination of the number of factors the items in the questionnaire were measuring.

Pilot Study Subjects

Employees from a southwestern medical university were solicited to be respondents in the pilot study. One hundred and forty questionnaires were distributed, the resulting $N = 92$. All subjects were treated in accordance with the ethical standards established by the American Psychological Association.

Pilot Study Procedure

Instructions were given to the pilot study subjects to complete the questionnaire. Anonymity was guaranteed to all participants in the study. Approximately ten minutes was needed by a respondent to complete the questionnaire items.

Pilot Study Analysis

A factor analysis was used to test the scale in order to determine whether or not the scale items were, in fact,

measuring the factors being investigated: familiarity,
expertise, and trustworthiness.

PILOT STUDY RESULTS

Responses to the 20-item questionnaire were factor analyzed using the principal components technique and a varimax rotation. On the basis of the eigenvalues and the scree test, it was determined that three factors would be retained for further analysis. Eigenvalues can be found in Table 1.

Table 1. Pilot Study
Eigenvalues before Rotation

Factor	Eigenvalue	%of Var	Cum%
1.	7.93	49.5	49.5
2.	1.28	8.0	57.5
3.	1.12	7.0	64.5

After rotation, there were five items which loaded highly on Factor 1 that appeared to be items measuring the trust dimension. There were five items which loaded highly on Factor 2, but not on Factor 1 or Factor 3. These items seemed to be measuring the construct described in the literature review as expertise. The final construct defined by the items which loaded highly on Factor 3, appeared to be

familiarity. The first factor accounted for 49.5% of the variance, while Factor 2 and Factor 3 combined accounted for 15% of the total variance. Of the original 20 items from the pilot survey, 18 were used in the thesis experiment. Items 6 and 16 were discarded because they cross loaded on all three factors. Questionnaire items and the factor loadings can be found in Table 2. The corresponding alphas can be found in Tables 3, 4, and 5.

Table 2. PILOT STUDY
Factor Loadings after Rotation

<u>Scale:</u>	Factor 1	Factor 2	Factor 3
Item 1	.343	.721	.074
Item 2	<u>.831</u>	.273	.132
Item 3	.348	<u>.563</u>	-.041
Item 4	.092	.085	<u>.647</u>
Item 5	<u>.747</u>	.211	.103
Item 6	.547	.433	.199
Item 7	<u>.814</u>	.314	.176
Item 8	.323	<u>.789</u>	.149
Item 9	.304	.198	.368
Item 10	.105	.425	.702
Item 11	.278	.434	.503

Item 12	<u>.741</u>	.283	.205
Item 13	.168	.504	.255
Item 14	<u>.699</u>	.304	.209
Item 15	.130	.404	<u>.694</u>
Item 16	.356	.492	.319
Item 17	.308	-.066	<u>.649</u>
Item 18	<u>.628</u>	.407	.353
Item 19	.172	<u>.799</u>	.166
Item 20	<u>.766</u>	.369	.256

Table 3

Reliability Analysis of Pilot Study (Factor 1)
(Trust Scale)

<u>Items:</u>	Corrected Item- Total Correlation	Alpha if Item Deleted
2	.812	.891
5	.798	.893
7	.867	.939
12	.667	.911
18	.761	.898
<hr/>		
alpha = .914		
N = 92		

Table 4.

Reliability Analysis of Pilot Study (Factor 2)
(Task Expertise)

<u>Items:</u>	Corrected Item- Total Correlation	Alpha if Item Deleted
1	.582	.726
8	.735	.644
19	.683	.676

alpha = .778

N = 92

Table 5.

Reliability Analysis of Pilot Study (Factor 3)
(Familiarity)

<u>Items:</u>	Corrected Item- Total Correlation	Alpha if Item Deleted
4	.262	
15	.404	
17	.404	

alpha = .573

N = 92

PILOT STUDY SUMMARY

A principal axis factor technique was used to obtain evidence that the extraction of three scales from the principal components approach was appropriate for the major study. The use of three factors was supported by the principal axis approach. Items which cross-loaded were used in the main study. These items helped to establish the use of three factors. Five items were added to the questionnaire in the major study in order to ensure that the factors being observed were captured by these additional statements. Final questionnaire items can be found in Appendix A.

THESIS STUDY METHOD

Subjects

Subjects were 133 instructors from faculty at a southwestern university. Of those instructors who participated in the study and chose to identify their gender, 50 were male and 45 were female. Of the 133 respondents, 22% were professors, 26% were associate professors, 22% were assistant professors, and another 22% were classified as lecturers. Twenty-six percent of the respondents were tenured faculty members, 38% were probationary, and 20% were non-tenured track faculty. With regard to full-time and part-time faculty, 70% were full-time and 13% were part-time. Subjects were treated according to the ethical standards set forth by the American Psychological Association.

Procedure

The questionnaire developed by the researcher was used to assess the perceived credibility of the two different sources of evaluation. An even number of the available faculty were given questionnaires to rate one of two evaluation sources, either faculty or student. Use of faculty mail boxes was the primary means of distributing the questionnaires to the subjects. Respondents were asked to return completed questionnaires in the envelopes provided to the psychology office.

Analysis

Three t-tests were used to test hypotheses 1-3. The dependent measures were expertise, familiarity, trustworthiness. A multiple regression analysis was used to test hypothesis 4. This analysis was used to determine which of the three factors (expertise, familiarity, or trustworthiness) contributed most to the explained variance of perceived overall credibility. A standard regression method was used in this process.

THESIS STUDY RESULTS

Questionnaires were distributed to all available faculty. Twelve of the questionnaires were returned due to the fact that some faculty were unable to be reached through campus mail. One hundred and forty three questionnaires were returned to the sender. Of the 143 returned, ten were unanswered. Thus leaving 133 data sets to be used in the analyses giving us a response rate of 33%. The questionnaires which asked the respondent to rate students as evaluation sources numbered 87. Forty six completed questionnaires asked respondents to rate faculty sources of evaluation.

Factor Analysis

A factor analysis was utilized in order to determine the nature of the items in the questionnaire after five items were added to the original form used in the pilot study. The five questionnaire items were added in order to ensure that the three factors identified in the pilot study had items reflecting the nature of each. The eigenvalues that resulted in the extraction analysis as shown in Table 6, allowed three factors to be retained for further analysis.

Table 6.

<p style="text-align: center;"><u>Thesis Study</u></p> <p style="text-align: center;"><u>Eigenvalues Before Rotation</u></p>			
Factor	Eigenvalue	% of Var	Cum %
1 (Ability to rate)	10.54	52.7	52.7
2 (Expertise)	1.79	8.9	61.7
3 (Familiarity)	1.22	6.1	67.8

After oblique rotation, 11 items that loaded highly on Factor 1 appeared to be measuring the ability to rate, a construct not anticipated to result from the analysis. Task expertise seemed to be the construct being defined by the three items which loaded highly on Factor 2. Factor 3 appeared to be measuring familiarity.

Item number three in the questionnaire was designated as the single item to define overall credibility as the dependent measure. Subsequently, it was not entered into the initial factor analysis or rotation.

One construct that did not factor out clearly was trustworthiness. Items supposedly tapping into the characteristics of trust cross-loaded with familiarity and

expertise items. The factor loadings for the major study questionnaire items can be found in Table 7. The results of the reliability analysis can be found in Tables 8, 9, and 10.

Table 7.

<u>Thesis Study</u>			
<u>Factor Loadings after Rotation</u>			
<u>Scale:</u>	Factor 1	Factor 2	Factor 3
Item 1	.173	.381	.496
Item 2	<u>.837</u>	-.210	.215
Item 4	.053	<u>.728</u>	-.208
Item 5	<u>.913</u>	-.307	.108
Item 6	<u>.703</u>	.070	.234
Item 7	<u>.746</u>	.091	.221
Item 8	.145	.354	.574
Item 10	.001	<u>.795</u>	.201
Item 12	<u>.792</u>	-.118	-.067
Item 13	<u>.796</u>	.102	.074
Item 14	.495	.460	.083
Item 15	.215	-.134	<u>.665</u>
Item 16	<u>.629</u>	.240	-.182
Item 17	.155	.365	<u>.523</u>

Item 18	<u>.764</u>	.113	-.114
Item 19	<u>.766</u>	.202	-.219
Item 20	<u>.720</u>	.163	-.347
Item 21	.372	.488	-.193
Item 22	<u>.595</u>	.166	-.145
Item 23	.352	<u>517</u>	-.213
Item 24	.022	<u>.792</u>	-.154

Table 8.

Reliability Analysis of Thesis Study (Factor 1)
(Ability to Rate)

<u>Items:</u>	Corrected Item- Total Correlation	Alpha if Item Deleted
2	.766	.933
5	.754	.933
6	.771	.933
7	.851	.929
9	.467	.943
12	.642	.937
14	.748	.933
16	.780	.932
18	.830	.930
19	.780	.932
20	.645	.937
22	.685	.936

alpha = .939

N = 126

Table 9.

Reliability Analysis Of Thesis Study (Factor 2)
(Task Expertise)

<u>Items:</u>	Corrected Item- Total Correlation	Alpha if Item Deleted
1	.612	.811
4	.400	.864
10	.742	.775
23	.693	.787
24	.763	.766

alpha = .836

N = 130

Table 10.

Reliability Analysis of Thesis Study (Factor 3)
(Familiarity)

<u>Items:</u>	Corrected Item- Total Correlation
---------------	--------------------------------------

15	.394
----	------

17	.394
----	------

alpha = .564

N = 127

Student's t-Test

A t-test was used to analyze the mean differences in levels of ability to rate, familiarity, and expertise between our two groups of data. It was found that the faculty sources of evaluation were perceived as possessing higher levels of expertise, $t(129) = 9.44$, $p < .05$., one-tailed significance.

Upon analyzing the results from the factor analysis, the construct identified as ability to rate or evaluate (Factor 1), was not expected to fall out as a separate factor from expertise. In essence, expertise broke into two

factors (a) ability to rate or evaluate, and (b) task expertise. The results from the t-test on Factor 1 - ability to rate are as follows: $t(127) = 1.47$, $p > .05$, one-tailed test. These results indicate that there are no significant differences in the sample means. Neither the faculty source nor the student source of evaluation was perceived as possessing more ability to rate.

Hypothesis 2 was not supported. The student source of evaluation was not perceived as being more familiar with the ratee's performance when compared to faculty sources. Because the factor trustworthiness failed to emerge from the factor analysis, no scale was available for hypothesis three.

Multiple Regression Analysis

Hypothesis 4 was not supported by the results obtained in the regression analysis. This procedure was carried out to determine the contributions of the three measures; ability to rate, task expertise, and familiarity with task performance, to the criterion variable overall credibility. The method used was simultaneous entry of predictor variables. The resulting $R = .79$ ($R^2 = .63$, $F(4,119) = 50.20$, $p < .001$). The variance accounted for by all three predictor variables (ability to rate, familiarity, and expertise) was 63%. Unfortunately the trust scale was never

established and a source of variance was lost. Factor 1, the ability to rate had a beta weight of .76. In comparison to Factor 2 expertise, (beta = $-.018$) and Factor 3 familiarity, (beta = $-.0001$), Factor 1 is the best predictor of overall credibility.

Additional tests were run to explore the possibility of tenured and non-tenured faculty responding differently to the scales. An analysis of variance with a two-by-two factorial design was used. No significant differences were found. The two main effects that were being examined were, tenured faculty versus non-tenured faculty and subjects who rated students as evaluation sources versus subjects who rated other faculty. The interaction effects (tenured status by experimental manipulation) of these variables were not significant, therefore revealing that tenure did not make a difference in the way the subjects responded to the questionnaire. (Interaction effects: Scale 1: ability to rate $F(1A) = .026$. Significance of $F = .873$; Scale 2: Expertise $F(1) = .060$. Significance of $F = .808$; Scale 3: Familiarity. $F(1) = .046$. Significance of $F = .831$).

A correlation analysis was used to test the relationship between number of years teaching and the scales generated from the questionnaire. Those subjects who rated students ($N=87$) versus those who rated other faculty ($N=46$) were compared on their responses to Scales 1, 2, and 3. For

Scale 1: Ability to rate; the results showed that instructors who had more years in teaching, perceived that students are less able to rate their performance ($r = -.145$) and that faculty are perceived as being more capable to carry out this task ($r = .069$). For Scale 2: Task expertise; the results indicated that as instructors get older they perceived students to be less expert ($r = -.094$) than faculty ($r = -.091$). For Scale 3: Familiarity; it was found that as instructors get older, they perceived students as being less familiar ($r = -.142$) with their performance than other faculty members who evaluate them ($r = .057$).

DISCUSSION

The present findings lend support to one of three hypotheses tested by the Student's t-Distribution. Faculty sources of evaluation were perceived as being more expert than student sources. This result is consistent with Heppner and Handley's (1982) position on supervisory behavior. They suggested that trainees who perceived their supervisor as engaging in evaluative supervisory behavior were more apt to also perceive their supervisor as more expert. The items that tapped into the expertise dimension stated that the rater understands the evaluation process and that the rater's experience assists him/her in the evaluation process. It is evident through the responses on these items that university instructors consider their fellow faculty members as being generally more credible than students.

Hypothesis 2 was not supported in the direction expected. This goes against what was established in the literature review. Students who were designated as subordinates to instructors should have been perceived as being more familiar with performance than faculty. Cusella (1982) indicated that supervisors who consequently do not interact on a regular schedule with the individuals they evaluate were less familiar, with the relevant job or work unit. Evidently, instructors may in fact feel that

their peers and department chairpersons are more familiar with their performance in class, as well as, familiar with their task responsibilities since most have been instructors at some time during their academic careers.

Although this was not the expected outcome, Goffman (1959) offers a dramaturgical model for organizations that would support the idea that faculty and peers are more familiar with an instructor's performance in class. In this model, a stage production is used as a comparison to the organization. The players in the production, be they actors onstage, members of the crew backstage, or even the audience are said to have important roles in making the show a success. Those acting and those observing the actors before the show is presented to an audience, are expected to be most familiar with the production since they interact most with the actors. Whereas, the audience does not possess the knowledge that the crew and actors themselves possess. The members of the audience are there usually to be entertained as are students in a classroom. The interaction present between actor and audience is brief and restricted. This interaction can be paralleled to that between a professor and his/her students. More information on the dramaturgical model can be found in Goffmann's book The Presentation of Self in Everyday Life.

Trustworthiness was not a resulting factor in this

experiment. Because the items cross-loaded badly with the expertise and familiarity dimension, it was difficult to use these as measures of perceived level of trust. It was obvious that usage of this factor would not improve the following analyses.

The ability to rate was the unexpected factor that fell out of the expertise variable. It appears from this evidence, that ability is of significant importance to performance evaluations. Although the t-test for this variable was not significant, and being that it was unexpected, future research in performance evaluations might further explore testing this variable.

The regression analysis indicated that ability to rate or evaluate would be the best predictor of perceived overall credibility. Although this finding was not hypothesized, future research should focus on testing this factor which affected perceived credibility more than the other two predictor variables.

A problem with scale construction made it somewhat difficult to obtain clean concise results to support our hypotheses. Some scale items caused confusion allowing subjects to "best guess" what the researcher was asking for in terms of a response. It is noted that the items were written as absolutes, where the beginning statement "rater is" was used primarily, therefore not addressing the

possibility that faculty and students as raters will undoubtedly vary in their ability to evaluate an instructor's performance. The purpose of the questionnaire was to assess the differences in an instructor's perception of believability when comparing students and faculty as raters. Although it is beyond the scope of this study, a comparison of students to students by grade level, with regard to perceived credibility, may prove conducive to better understanding the role of students as raters.

Due to the recurring difficulties of the measurement, it is obvious that any generalization of results be made with much caution and remain within the boundaries of the university setting. It should be noted that instructions for filling out the questionnaire should be more concise and unambiguous. Future research should pay closer attention to detail in the development of questionnaire items dealing with the issue of performance evaluation since this activity, as exercised by different organizations, affect promotion, retention, and tenure. Undoubtedly, the performance evaluation issue is not to be made less important through these results nor is the student source of evaluation to be discredited.

Appendix A

Circle the following items which apply:

Male/Female

Rank: Professor Associate Assistant Lecturer
Professor Professor

Tenure: Tenured Probationary Non-Tenure Track
Full-time/Part-time

How many years have you been teaching?_____

How many classes (approximately) have you taught per year?_____

INSTRUCTIONS: Respond to the following items keeping in mind that students are the sources of your performance evaluation. The raters in this case are students. Place the number on the space provided which appropriately matches your reaction.

STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE
5	4	3	2	1

1.				
2.				
3.				
4.				
5.				
6.				
7.				

- ___ 8. The rater is familiar with the activities I carry out on a daily basis.
- ___ 9. The feedback I receive from the rater is favorable.
- ___ 10. Rater understands completely the importance of the evaluation process.
- ___ 11. Rater's experience and background assists him/her in the evaluation of my performance.
- ___ 12. The feedback I receive from the rater is taken as constructive criticism.
- ___ 13. Rater is a competent individual capable of evaluating my performance.
- ___ 14. Rater can be trusted to deliver feedback that is unbiased and impartial.
- ___ 15. Rater is usually present during my working hours.
- ___ 16. I trust the rater to do a consistent evaluation on my performance.
- ___ 17. Rater is familiar with the amount of time I spend on carrying out my daily activities.
- ___ 18. Rater can be relied upon for good performance feedback
- ___ 19. Rater has evaluated my performance fairly and accurately.
- ___ 20. I agree with the rater on the adjustments suggested to improve my performance based on the evaluation.
- ___ 21. Rater has the training and experience needed to understand my position as an instructor.
- ___ 22. I perceive the rater as being the most qualified to evaluate my performance.

- ____ 23. Rater is aware of the training needed to carry out the responsibilities I have as a university instructor.
- ____ 24. Rater is familiar with the performance evaluation process.

- ____23. Rater is aware of the training needed to carry out the responsibilities I have as a university instructor.
- ____24. Rater is familiar with the performance evaluation process.

Appendix B

Thesis Study Demographic Information

	<u>Frequency</u>	<u>Percent</u>
<u>GENDER:</u>		
Male	53	39.8
Female	41	30.8
Missing	39	27.8
	<hr/>	<hr/>
	N = 133	

<u>RANK:</u>		
Professor	30	22.6
Associate Professor	36	27.1
Assistant Professor	31	23.3
Lecturer	30	22.6
Missing	6	3.8
	<hr/>	<hr/>
	N = 133	

<u>TENURE:</u>		
Tenured	40	30.1
Probationary	53	39.8
Non-Tenured Track	26	19.5

Missing	14	9.8
---------	----	-----

N = 133

STATUS:

Full-time	95	71.4
Part-time	19	14.3
Missing	19	13.5

N = 133

Frequency

Percent

YEARS TAUGHT:

0 (Missing)	3	2.3
1	11	8.3
2	8	6.0
3	8	6.0
4	5	3.8
5	7	5.3
6	8	6.0
7	8	6.0
8	6	4.5

9	3	2.3
10	6	4.5
11	4	3.0
12	4	3.0
13	9	6.8
15	1	.8
16	4	3.0
17	4	3.0
18	5	3.8
19	2	1.5
20	7	5.3
21	2	1.5
22	3	2.3
24	3	2.3
25	1	.8
26	3	2.3
27	2	1.5
30	3	2.3
35	1	.8
38	1	.8
39	1	.8

N = 133

CLASSES TAUGHT PER YEAR:

0 (Missing)	2	1.5
1	1	.8
2	3	2.3
3	7	5.3
4	7	5.3
5	4	3.0
6	12	9.0
7	10	7.5
8	27	20.3
9	35	26.3
10	5	3.8
11	2	1.5
12	9	6.8
13	2	1.5
14	1	.8
15	4	3.0
17	1	.8
20	1	.8

N = 133

REFERENCES

- Baird, L. S., (1977). Self and supervisor ratings of performance: As related to self-esteem and satisfaction with supervision. Academy of Management Journal, 20, 291-300.
- Cohen, P. A., (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 51, 281-309.
- Cusella, L. P., (1982). The effect of source expertise and feedback valence on intrinsic motivation. Human Communication Research, 9, 17-32.
- Dowell, D. A. & Neal, J. A., (1982). A selective review of the validity of student ratings of teaching. Journal of Higher Education, 53, 51-62.
- Doyle, K. O., Jr., & Whitely, S. E., (1974). Student ratings as criteria for effective teaching. American Educational Research Journal, 11, 259-274.
- Elliot, D. H., (1950). Characteristics and relationships of various criteria of college and university teaching. Purdue Studies in Higher Education, 70, 5-61.
- Frey, P. W. (1973). Student ratings of teaching: Validity of several rating factors. Science, 182, 83-85.
- Gessner, P. K. (1973). Evaluation of instruction. Science, 180, 566-569.
- Giffin, K., (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. Psychological Bulletin, 68, 104-120.
- Goffman, E. (1959). The Presentation of Self in Everyday Life. Doubleday & Company, Inc. New York.
- Harris, M. M. & Schaubrook, J. (1988). A meta-analysis of self-supervisor, self-peer, peer-supervisor ratings. Personnel Psychology, 41, 4-62.
- Heppner, P. P. & Handley, P., (1982). The relationship between supervisor behavior and perceived supervisor expertness, attractiveness, or trustworthiness. Academy

of Management Journal, 22, 37-46.

Herold, D. M., Liden, R. C., & Leatherwood, M. L. (1987). Using multiple attributes to assess sources of performance feedback. Academy of Management Journal, 30, 826-835.

Holzbach, R. L. (1978). Rater bias in performance ratings: Supervisor, Self, and peer. Journal of Applied Psychology, 63, 579-588.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S., (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349-371.

Jewell, L. N. (1985). Contemporary industrial/organizational psychology. Minnesota: West Publishing Co.

Keeley, M., (1977). Subjective performance evaluation and person-role conflict under conditions of uncertainty. Academy of Management Journal, 20, 301-314.

Klimoski, R. J., & London, M., (1974). Role of the rater in performance appraisal. Journal of Applied Psychology, 59, 445-451.

Knight, W. H. & Holen, M. C., (1985). Leadership and the perceived effectiveness of department chairpersons. Journal of Higher Education, 56, 677-689.

Korman, A. K., (1968). The prediction of managerial performance: A review. Personnel Psychology, 21, 295-322.

Kurz, R. S., Mueller, J. J., Gibbons, J. L., & DiCataldo, F., (1989). Faculty performance: Suggestions for the refinement of the concept and its measurement. Journal of Higher Education, 60, 49-55.

Larson, J. R. Jr., (1986). Supervisors' performance feedback to subordinates: The impact of subordinate performance valence and outcome dependence. Organizational Behavior and Human Decision Processes, 37, 391-408.

Larson, J. R. Jr., (1989). The dynamic interplay between employees' feedback-seeking strategies and supervisors' delivery of performance feedback. Academy of Management Review, 14, 408-422.

- Latham, G. P. & Wexley, K. N. (1982). Increasing productivity through performance appraisal. Phillipines: Addison-Wesley Publishing Co.
- Morsch, J.E., Burgess, G. G., & Smith, P. N. (1956). Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 47, 79-88.
- O'Reilly, C. A. & Anderson, J. C., (1980). Trust and the communication of performance appraisal information: The effect of feedback on performance and job satisfaction. Journal of Applied Psychology
- Overall, J. U., & Marsh, H. W., (1982). Student evaluations of teaching: An update. Research Currents/American Association for Higher Education Bulletin, December, 9-12.
- Remmers, H. H., Martin, E. D., & Elliot, D. N., (1949). "Are students' ratings of instructors related to their grades?". Purdue University Studies in Higher Education, 66, 17-26.
- Roadman, H. E., (1964). An industrial use of peer ratings. Journal of Applied Psychology, 48, 211-214,
- Rodin, M., & Rodin, B., (1972). Student evaluations of teachers. Science, 177, 1164-1166.
- Schermerhorn, J. R. Jr., Hunt, J. G., & Osborn, R. N. (1988). Managing organizational behavior (3rd ed.) New York: John Wiley & Sons.
- Shrauger, J. S. & Lund, A. K., (1975). Self-evaluation and reaction to evaluations from others. Journal of Applied Psychology, 43, 94-108.
- Sullivan, A. M. & Skanes, G. R., (1974). Validity of student evaluations of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 66, 584-590.
- Ulanski, S. L., (1987). A quantitative assessment of student evaluations of an introductory physical geology course. Journal of Geological Education, 35, 25-27.
- Whitely, S. E. & Doyle, K. O., Jr., (1979). Validity and generalizability of student ratings from

between-classes and within-class data. Journal of Educational Psychology, 71, 117-124.

Zammuto, R., London, M., & Rowland, K. M. (1982). Organization and rater differences in performance appraisals. Personnel Psychology, 35, 643-658.