California State University, San Bernardino

CSUSB ScholarWorks

Methods to reduce leniency bias in self-assessments and self-appraisals

Michael Robert Dolen

METHODS TO REDUCE LENIENCY BIAS IN SELF-ASSESSMENTS AND SELF-APPRAISALS

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Industrial/Organizational Psychology

by
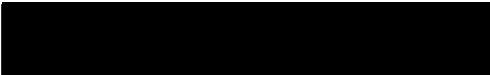
Michael Robert Dolen

June 1995

METHODS TO REDUCE LENIENCY BIAS IN SELF-ASSESSMENTS AND SELF-APPRAISALS

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

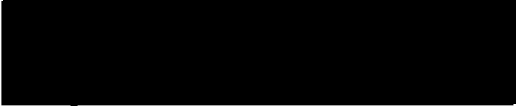Michael R. Dolen

June 1995

Approved by:

Jaet L. Kottke, Chair, Psychology   Date   5/31/95

Matt L. Riggs

Glria A. Cowan

ABSTRACT

Self-assessments are a desirable source of information due to their relatively inexpensive cost of administration and the rich information that they yield, however, they also have several deficiencies, the most significant of which has been termed leniency bias. The current research attempted to identify the situations in which self-assessments tend to be most lenient and whether it is possible to statistically correct for leniency.

Analyses were conducted using ANOVA and ANCOVA on subject's self-assessments. The independent variables were comparison condition (social comparison vs absolute judgment), expectation of validation (high vs low), and purpose of the assessment (research vs grading). The covariate in the ANCOVA was a "lie scale" that consisted of bogus items.

The results of the ANOVA demonstrated that there was significantly less leniency evidenced: when subjects rated themselves on a social comparison basis, when subjects thought that their self-assessments would be validated, and when self-assessments were collected for research purposes. Results of the ANCOVA demonstrated that it is possible to correct (statistically) for an individual's propensity to be lenient.

The results of this research demonstrate the conditions under which conditions self-assessments and self-appraisals are relatively free from leniency bias, which has been their major drawback. Further, it is possible to correct for an individual's proclivity to present themselves in a more favorable manner.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

Introduction

Self-assessments and self-appraisals (SA/SAs) have been used for a variety of purposes both in academic and applied settings. Specifically, self-assessments and self-appraisals have been used for performance appraisal, identification of training needs, selection, criterion measurement, and construct validation in research studies (Thorton, 1980). Self-assessments and self-appraisals are a valuable source of information for an organization, university, or any other institution that seeks to measure the abilities or performance of its members. The characteristic of SA/SAs that makes them so desirable is their low cost. Self-assessments and self-appraisals are relatively inexpensive compared to other forms of assessment (e.g., written tests, physical abilities tests).

There are many advantages that SA/SAs have over other forms of evaluation. In his seminal paper "The Self System in Reciprocal Determinism," Albert Bandura describes many of these advantages. First, every person has detailed information of his/her attributes under many different situations. Second, each of us experiences the consequences of our behavior in a variety of natural settings. Third, we gain information about ourselves by observing others. Fourth, we all have the opportunity to authenticate our feelings regarding our performance as well as who we are. Finally, we all receive feedback from others, whether it is intentional or unintentional, regarding ourselves and our performance which helps us to better judge our performance (Bandura, 1978). There are also other theoretical explanations as to why self-assessment and self-appraisal may be a desirable source of information. Self-perception theories such as that by Jones and

Nisbett (1971) and Bem (1967) note that an individual has access to every behavior that one has in every situation. Social comparison theory states that an individual will constantly observe others and make comparisons to oneself. This constant self-evaluation therefore leads to a better understanding of oneself, possibly a better understanding than any other person may have.

According to social comparison theory (Festinger, 1954), an individual first tries to make an evaluation of his/her ability based on physical features. If this comparison is not possible, one turns to comparing oneself to others around him/her. In the real world, physical characteristics with which to evaluate behavior are extremely limited, and even if they are available, social comparisons are still needed to give meaning to the results. An example will demonstrate this point: Greg just finished running 5 miles in 25 minutes. This physical information (time) does not give any indication as to how Greg performed. However, when Greg considers that the average time for people his age and roughly the same physical condition is 40 minutes, he has a much more detailed understanding of his behavior. In this case Greg may want to consider racing as a career.

Social Comparison Theory has two main implications for self-assessment and self-appraisal. First, instructions to rate oneself in relative terms, which shall be discussed in detail later, lead to less lenient self-appraisals than when subjects were instructed to use absolute ratings. Second, social comparison theory explains the different motivational operations that may take place in self-assessment and self-appraisal. Individuals are either driven to produce accurate self-evaluations or driven to enhance one's appearance. The

later characteristic is common when the result of the self-evaluation is linked to a desired goal. In other words, the purpose of the appraisal may determine the accuracy of the evaluation. If the self-evaluation is to decide who gets an award, for example, then there is a greater probability that the evaluator will produce an inappropriately high evaluation. This phenomenon will also be discussed in greater detail later in this paper.

There are also drawbacks to the use of SA/SAs. Bandura (1978) states that the process of self-assessment and self-appraisal is very complex and therefore makes it difficult for the assessor to accurately assess his/her behavior. There are several different processes in which an individual is engaged during self-evaluation. First, there is an observation component which is described as a process where the individual monitors his/her behavior on several different dimensions. This makes it very difficult for the individual to be accurate because he/she may have to observe himself/herself on as many as eight different dimensions. Second, there is a judgment component in which the individual has to process all of his/her observations (on as many as eight dimensions) and then evaluate all of the resulting information.

To make the evaluation process even more arduous, there are several different factors that affect self-evaluations in a number of ways. Among these factors are: personal values (whether or not the behavior is a valued one), personal standards, reference groups, and the attribution of cause (i.e., external influences or personal dispositions). All of the above reasons make self-assessment and self-appraisal difficult by reason of cognitive complexity of the task. A third disconcerting characteristic is that the

individual may have some reaction to the evaluation process itself, either positive or negative, that may abate the accuracy of the evaluation.

Based on the above, an evaluation may become deficient because of poor observation, poor judgment, or personal reactions to the process. Additionally, it should be noted that this process is made even more difficult when one considers that all of this processing is taking place when one is actually performing the behavior, and observing the reactions of others (Bandura, 1978).

Based on the previous discussion of the advantages and disadvantages of self-assessments and self-appraisals, it is arguable that self-appraisals of work performance and self-assessments of abilities are an efficacious source of evaluative information, despite their disadvantages. However, to date their use has been somewhat restricted. Why?

Perhaps the most probable explanation for this is the tendency for self-assessment and self-appraisal to be inappropriately raised, which has been termed both "inflation bias" and "leniency bias". These are two different terms with the same definition and shall be used interchangeably throughout this paper. The reason for this interchangeable usage is to present past research in its original terminology. Leniency bias refers to a heightened rating which may be different from a different rating source or from an objective measure. This elevated rating can occur in any type of rating, be it self-, peer, or supervisor (Holzbach, 1978).

There has been much research done that attempts to ascertain the different situations in which individuals are more likely to "inflate" their scores. Further, there has

also been much research done that has attempted to control or eliminate this tendency. However, before examining this research, it is important to examine some of the early research on self-assessment and self-appraisal.

Early Research

Clearly, there are advantages (e.g., low cost, rich information) to the use of self-ratings; unfortunately individuals have been found to demonstrate a leniency bias in several research findings. In early studies that attempted to determine the usefulness of SA/SAs, specifically, their usefulness in an organizational setting, concerns about the veracity of SA/SAs were raised. Parker, Taylor, Barrett, and Martens (1959) were the first to raise such questions when they demonstrated that self-appraisals showed more leniency and less variability than supervisory appraisals. In this study, self-assessors' mean appraisals were consistently higher than the appraisals of their supervisors. Further, subjects were asked to estimate what they thought their supervisory appraisals would be. Subjects consistently estimated that their supervisory appraisals would be lower than their self-appraisals, yet their estimate of supervisory ratings was not as low as the actual supervisory appraisal. In this study, these results held true for both job-related dimensions as well as for person-related dimensions (those dealing with the employee's personality).

Factors Related to Inflation of SA/SA

Since the Parker et al study, many other studies have been done to examine what types of people/employees would demonstrate this tendency to inflate their scores (i.e., demonstrate leniency bias). The tendency for people to be lenient was demonstrated by

5

individuals employed in technical positions. Inflation has also been demonstrated in self-assessments of traits, i.e., personal characteristics, and for behavioral ratings, i.e., ratings of work performance (Kirchner, 1966). The tendency toward inflation has been demonstrated to hold true for supervisors (Waldman & Thorton, 1979) and in second level supervisors (Prien & Liske, 1962). The tendency to be lenient has also been demonstrated in executives. One-hundred eighty professional and managerial employees in a manufacturing organization demonstrated greater leniency effects compared to superior ratings and peer ratings (Holzbach, 1978). Lawler (1967) demonstrated that older workers who are dissatisfied with their pay are far more likely to inflate their self-appraisals than are younger workers or people who are satisfied with their pay.

Fishbein (1965) and Keefer (1969) reported stability in self-evaluations of ability over time in college students' assessments of academic performance. However, the same finding does not hold true in an applied setting. According to Thorton (1968), lenient self-appraisals tend to increase over time. The reasons for this are not delineated in the literature, but the author would speculate that the purpose of the appraisal may moderate this relationship. Specifically, in academic settings, self-assessments are rarely gathered for any purpose other than a developmental aid, whereas, in an organizational setting, self-assessments could be obtained for a number of different reasons. In an organization, self-assessments are often gathered for a variety of reasons, ranging from identification of training needs (where less leniency is exhibited) to personnel decisions (where leniency is

more pronounced). Therefore, it is possible that an individual's self-assessment can vary from one time to another.

Whereas many of the previously mentioned studies examined the relationship between self-appraisals and appraisals made by supervisors, self-appraisals have also been compared to appraisals made by peers. The relationship between self-appraisals and peer appraisals is very similar to the self-supervisor relationship in that self-appraisals consistently demonstrate a leniency bias (Holzbach, 1978; Klimoski & London, 1974; Lawler, 1967).

Factors Affecting the Accuracy of Self-Assessments

Another common method for assessing the accuracy of self-assessments is to compare them with external criteria, such as a personality measure or objective test. One such study compared pretest ratings with scores on the WAIS-R (Wechsler Adult Intelligence Scale - Revised) (Wechsler, 1981). Self-assessed scores were collected for a verbal component, a non-verbal component, and a general intelligence component. Pretest assessments demonstrated that individuals rated their own intelligence with limited accuracy in all categories. Stated another way, subjects' scores were significantly inflated. In the same study, subjects also provided self-assessments after taking the WAIS-R. The results of the feedback about performance led to more accurate self-assessment; there was no significant inflation in self-ratings (Schmitt & Strein, 1987).

The implications of such findings could be interpreted in two ways. The first explanation is that feedback on performance can significantly increase the accuracy of

future self-assessments. The second explains the obtained results with respect to cognitive dissonance theory. Cognitive Dissonance Theory (Festinger, 1957) postulates that individuals strive to maintain a balance between their attitudes, cognitions, and behaviors. If one of these is to differ from the others, thus creating dissonance (i.e., a state of discord), something must change in order to restore a balance and reduce the uncomfortable feeling (dissonance). Most often, it is either the attitude or cognition that will change, since it is often impossible to go back in time and change a behavior that has already been committed. After giving their initial assessments, subjects obtained accurate estimations of their ability which were incompatible with their belief (assessments), thus producing dissonance. The choices that the individual had to reduce this dissonance were either to change their belief (assessments) or to ignore the new information. Since most subjects believed that the test was accurate, their only recourse was to change their beliefs. This is one possible explanation as to why the subjects in this experiment changed their assessments after getting objective feedback regarding their performance.

Recent Research

Much research conducted throughout the 1980's and into the 1990's has examined methods that, when implemented into the SA/SA setting, will either reduce or correct for the tendency for people to inflate their SA/SAs (i.e., leniency bias), thereby increasing the validity of the tests/assessments. These methods include rater training, perceived system knowledge, and self-esteem.

Rater Training

Early studies of rater training concluded that rater error training (RET), training raters about the errors typically associated with appraisals (halo, central tendency, range restriction, and leniency), generally resulted in lower mean scores (less halo) and lower scale inter-correlations (less halo) (Bernardin & Walter, 1977; Borman, 1975). However, soon thereafter it was posited that these results are usually obtained at the expense of rating accuracy. In other words, RET may have reduced halo, for example, but it probably reduced "true halo", thereby reducing the accuracy of the appraisal by creating a new response set, but not necessarily an improved response set (Bernardin & Pence, 1980). Bernardin & Buckley (1981), in an effort to overcome the previous deficiencies, developed Frame of Reference (FOR) training.

Frame of Reference training is designed to ensure that all raters have the same understanding of what comprises "good", "poor", and "excellent" performance. This is typically accomplished by giving raters examples of behaviors (critical incidents) as well as the correct ratings on behavioral ratings scales. To date, FOR training has received strong support as an alternative to RET (Athey & McIntyre, 1987; Bernardin & Buckley, 1981; Hedge & Kavanaugh, 1988; Pulakos, 1984, 1986; Stamoulis & Hauenstein, 1993) with respect to rating accuracy. In as much as self-appraisals are typically correlated with supervisory appraisals as the criterion of choice to determine their accuracy, Frame of Reference training can be useful in reducing leniency bias.

Similarly, when both supervisors and self-raters have similar "frames of reference" as to what constitutes good performance, the correlation between self-ratings and supervisory ratings is higher. In other words, to the extent that there is ambiguity regarding performance dimensions (absence of a adequate frame of reference), leniency in self-ratings is increased. This has been demonstrated for not only SA/SA, but for performance evaluations in general. When there are no well defined performance dimensions on which to base evaluations, raters base their evaluation on an overall "global" evaluation thereby decreasing the accuracy (Woehr, 1992). Further, as the number of possible criteria used in evaluating performance increases, the ambiguity of that particular performance dimension increases, and therefore, the leniency in SA/SA increase. Consequently, one method to reduce the amount of leniency in self-appraisals is to clearly define the dimensions on which behavior is to be evaluated (Dunning, Meyerowitz & Holzberg, 1989).

Perceived System Knowledge

Another line of research has indicated that the amount of knowledge that an employee has (or feels he/she has) regarding the appraisal process will moderate the relationship between self- and supervisory appraisals. Specifically, the more knowledge an employee feels that he/she has regarding the appraisal system itself, the more agreement their self-appraisals will share with their supervisor's appraisal. This is likely to result in a less threatening and more productive performance appraisal process. Additionally, this line of research demonstrates that we can make self-appraisals more accurately if we can

in some way increase the employee's knowledge about the appraisal process (Williams & Levy, 1992).

Self-Esteem

Self-Enhancement Theory (Greenwald, 1980) predicts that individuals with low self-esteem are more likely to inflate their self-assessments because people want to be perceived in a positive manner. Since people with low self-esteem feel that they are not as adequate as other individuals, they should inflate their assessments in order to look like they are just as good as other people.

By contrast, according to theories that describe maintaining consistencies between the different components of attitudes and feelings such as Cognitive Dissonance Theory (Festinger, 1957) and Congruity Theory (Osgood & Tannenbaum, 1955), people who have high self-esteem are more likely to inflate their self-assessments so they have consistency among their attitudes, cognitions, and behaviors. That is, to maintain their high self-image, people with high self-esteem would have to produce higher self-assessments in order to maintain consistency. Whether individuals simply pay more attention to behaviors that are more consistent with their self-esteem or they actually intentionally inflate their behaviors is unknown.

The available research that has been conducted to examine the relationship between self-esteem and self-assessments supports theories such as Cognitive Dissonance and Congruency. That is, the available research suggests that individuals with high self-

esteem are more likely to inflate their ratings than are those individuals with low self-esteem (Wells & Sweeney, 1986; Farh & Dobbins, 1989).

The Aim of This Research

The previously mentioned characteristics are interesting and important in developing measures of SA/SAs. However, perhaps the characteristics that hold the most promise for correcting, controlling, anticipating, or reducing leniency bias are: the comparison instructions given to self-raters, the expectations of validation, the purpose of the appraisal, and using a "lie scale" to correct for inflation. The aim of this research is to examine the effects that each of these conditions has on SA/SAs.

Social Comparison

The existing literature on SA/SAs which illustrates the effects that social comparisons have on SA/SAs demonstrates that when individuals make their self-ratings in relative terms (compared to others), they correlate much higher with other evaluations and with other criterion measures (i.e., less lenient) than when ratings are made in absolute terms (Farh & Dobbins, 1989b; Felson & Reed, 1986; Mabe & West, 1982). For example, if a pilot were to make his/her self-evaluation ratings as either poor, acceptable, or excellent (absolute ratings), it would be expected that his/her ratings would be much more inflated than they would be had he/she made his/her ratings as either below average, average, or above average (relative ratings). Further, it would be expected that the pilot's self-ratings would be less inflated if the reference group was well defined (Mabe & West, 1982; Raven & Fishbein, 1965).

The research that has been done examines the effects of absolute ratings on leniency or it examines the effects that social comparisons have on leniency. However, there is not much research that directly compares the two types of ratings and their effects on leniency. The only known existing study of this type is that done by Farh and Dobbins (1989) which directly manipulated this relationship. The results of this study supported previous research and assumptions which state that there is a higher correlation between SAs and criterion measures when individuals are instructed to make their ratings on a "compared to others" basis.

Expectation of Validation

An individual's expectation that his/her SA/SA will be compared to a criterion measure of performance, such as an objective test or supervisor's appraisal, has a sizable effect on the proclivity to inflate his/her SA/SA. When individuals believe that their SA/SAs will be compared to external criteria the tendency to be lenient is significantly decreased because of the fear of reprisal for dishonesty (Weary & Arkin, 1982; Zuckerman, 1979). In their meta-analysis, Mabe and West (1982) reported the results of 43 studies of self-appraisals. They demonstrated that when individuals feel that their SA/SA will be compared to external criteria, they tend to lower their SA/SAs so as not to appear as though they are being disingenuous.

Similar support can be inferred from the results obtained by Fox and Dinur (1988) who examined candidates for a prestigious military training program in Israel. Based on what we have already seen, one would anticipate a large amount of leniency in the SAs

13

due to the highly desirable goal. However, leniency was not found. The subjects in this study were in the final stage of a screening period before induction into the training program. All candidates had undergone extensive testing (interviews, medical exams, and paper and pencil tests) and were taking part in a 9 day "screening" period which was comprised of more extensive testing which included physical and mental tests, class exercises, and field simulations. On the last day, subjects were asked to fill out self-evaluation measures. It was communicated to the experimental group that their self-evaluations would be compared to peer evaluations, field officer evaluations, and other criterion measures: the control group did not receive these instructions.

Although the manipulation of giving instructions of expectation of validation to the "experimental group" did not produce effects that differentiated them from the control group, it is argued that the obtained results were due to expectations of validation on behalf of all subjects. That is, although the experimental group was specifically told their scores were going to be compared to the other measures, it is argued that even the members of the control group assumed that their scores would be compared to these other measures. This is evident if you consider that the entire purpose of participating in the screening period was to get a number of evaluations of their abilities.

Perhaps the strongest support for the argument that expectations of validation lead to less lenient SA/SAs comes from a study by Farh and Werbel (1986). This study demonstrated that there is a large expectation of validation effect. That is, when

instructions of expectation of validation instructions are given, there is not as large a propensity in individuals to inflate their SA/SAs.

Taken together, these studies illustrate that the expectation that SA/SAs will be validated against external criteria significantly reduces the likelihood that an individual will inflate their SA/SAs. This tendency appears to hold true even if the goal is desirable to the individual doing the SA/SA.

Purpose of the Appraisal

Simply because an appraisal system demonstrates reliable and valid results for one purpose does not necessarily mean that the same system will have equally reliable and valid results when used for another purpose (Cleveland & Murphy, 1992). Therefore, it is imperative that we examine the "quality" of appraisal systems that are used for different purposes. It is somewhat intuitive that SA/SAs that are conducted for research purposes are less likely to exhibit leniency bias than when they are being conducted for administrative purposes, such as to determine who gets a promotion.

Early research on this property of SA/SAs failed to directly examine the differences between the two conditions. Studies that examined leniency in research conditions (Holzbach, 1978; Klimoski & London, 1974) typically reported less leniency than did studies that examined leniency bias in SA/SAs conducted for administrative purposes (Parker, Taylor, Barrett, & Martens, 1959; Prien & Liske 1962; Lawler, 1968; Thorton, 1968; Anderson, Warner, & Spencer, 1984).

More recently, researchers have begun to examine the purpose of the appraisal with greater scrutiny. In a study that directly compared differing purposes of appraisal to determine when leniency is most prevalent supported previous claims that when appraisals are used for administrative decisions, there is a much higher tendency for ratings to demonstrate leniency than when appraisals are used for research purposes (Harris, Smith, & Champagne, 1995). Similarly, in a study that directly assessed differing purposes of self-appraisal, it was demonstrated that individuals were much more likely to inflate their self-ratings when the ratings were being used for administrative purposes (Farh & Werbel, 1986). However, the number of studies that directly compares different purposes of appraisal is still limited.

Lie Scale

One effort aimed at correcting for inflation bias is to create a lie scale which will assess the amount of inflation prevalent in each person's SAs, and "correct" for this inflation. Anderson, Warner, and Spencer (1984) developed one such scale for applicants for clerical positions. The application process required applicants to rate their experience for several tasks (identified through a task analysis) required to adequately perform the job. Anderson et al also included bogus items - tasks that did not actually exist - on which the applicants were to rate their experience (e.g., preparing routines renselaer reports, locating sums through use of decreased equivalents, and filing rhetaguards). The extent to which the applicants reported that they had experience on the bogus items was assumed to be the same as for the real tasks. This rating was called their exam score, which Anderson

et al could "correct" for the amount of inflation in the real tasks. They proposed two methods to do this correction, a regression method and an inflation proportion method. The regression method used the following equation: $Xc = X - r(Sx/Sy)Y$, where Xc = corrected exam score, X = individual applicant's examination score, r = correlation between exam score and inflation scale scores, Sx = standard deviation of exam scores, Sy = standard deviation of inflation scale scores, and Y = inflation scale score.

It was assumed that this correction would have low applicant acceptance (face validity), so they also created the inflation proportion (IP) method. The IP method assumes that the extent of inflation prevalent on the bogus items was the same as on the real tasks. Therefore, by taking the proportion of inflation on bogus items out of the real task ratings, one would correct for inflation.

The results of this research demonstrated that the amount of inflation was high, but it was possible to successfully correct for this inflation. Anderson et al determined that they successfully corrected for inflation in this situation by using an external criterion measure, a typing test, as a comparison. Specifically, the amount of inflation indicated by the lie scale was very comparable to what was evident in self-assessments of typing ability which was determined by the difference between subjects' assessments of the number of words they could type accurately per minute and the actual number that they achieved on a test of typing ability.

In summary, previous research regarding SA/SAs has examined many different conditions under which leniency bias is expected to be minimized. However, to date there

is no comprehensive analysis that has attempted to identify if it is possible to simultaneously use these different conditions in an attempt to reduce, or eliminate, any effect of leniency in self-assessments. Further, there is a strong need for experimental research that systematically manipulates the circumstances under which self-assessments are conducted to determine the best possible context for self-assessments and self-appraisal. Consequently, the present study will attempt to examine these factors by testing four hypotheses:

Hypothesis 1: Self-assessments are less likely to demonstrate leniency when subjects are asked to make ratings while comparing themselves to others (social comparison) than when they are asked to make global ratings of knowledge (absolute judgments).

Hypothesis 2: Self-assessments are less likely to demonstrate leniency when subjects are told that their assessments will be validated against an external criterion measure than when they are instructed that their self-assessments will not be compared to an external criterion.

Hypothesis 3: Self-assessments are less likely to demonstrate leniency when the assessments are being conducted for research purposes than when they are conducted for grading purposes.

Hypothesis 4: Leniency Bias can be corrected for by using an inflation scale as a

covariate.

Method

Subjects

The subjects for this experiment were 302 undergraduate students enrolled in three

undergraduate psychology courses at a state university in the southwest United States.

Both courses were taught using the same form of instruction (lecture) and both utilized

the same grading procedures (multiple choice tests). Class number one was an

introductory class which covered basic concepts in psychology (n=130). Class number

two was a mid-level developmental psychology class which covered middle childhood

(n=83). Class number three was an introductory developmental psychology course which

addressed basic concepts in developmental psychology (n=89). Courses consisted of 89

males and 201 females (12 missing).

Procedure

The self-assessment questionnaire was randomly distributed to all subjects in each

class along with their midterm examinations. There was no mention of different

questionnaires, so all subjects were led to believe that they received the same

questionnaire as everyone else. The questionnaires were deliberately distributed with the

test to minimize the chance that subjects would confer with each other and discover that

there was more than one questionnaire. Subjects were instructed to complete the

questionnaire prior to taking the exam and to turn it in along with the exam.

Experimental Manipulations

The study design was a 2 x 2 x 2 model with a covariate. The IVs in this study were expectation of validation (high vs low), purpose of the appraisal (research vs grading), and comparison (social vs absolute) with the lie scale score used as a covariate. The experimental manipulations for the three different conditions were achieved by altering the instructions given to subjects in each of the eight conditions. These manipulations are described next.

Expectation of Validation

Subjects who were in the "high expectation of validation" groups received, as part of their overall instructions, the following instructional set: To ensure accuracy, the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class.

Subjects who were in the "low expectation of validation" groups received, as part of their overall instructions, the following instructional set: Self-assessments are a valuable source of information and can often give accurate assessments of one's abilities. Self-assessments can often provide information that tests cannot due to the fact that tests do not perfectly measure a person's knowledge. Therefore, the results of this questionnaire will not be compared to test scores. Further, the results of this survey will be reported only as class data, with no individual responses will be examined.

Purpose of the Appraisal

Subjects who were in the "research condition" were instructed that the self-assessment questionnaire was being administered to gather information pertaining to students' knowledge of psychology. They received, as part of their overall instructions, the following instructional set: This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way affect anyone's grade in this class.

Subjects who were in the "grading condition" were instructed that borderline grades would be determined by their self-assessments. Subjects in this condition received the following instructions: The instructor of this class has decided that borderline grades will be determined by your self-assessments. In other words, if you are on the border of an A and A-, A- and B+, B+ and B, etc., your grade will be determined by your self-assessment. Complete examples of the instructions given to each of the eight groups are provided in Appendix A.

Comparison Condition

The two different conditions in the comparison manipulation were achieved by instructing the subjects to make their assessments either on an absolute basis or by comparing themselves to other people who have taken the class (social comparison). The subjects in the absolute rating condition were instructed to make their ratings on a 7 point Likert scale. The Likert scale had anchors of 1 = No Knowledge, 2 = Very Little

Knowledge, 3 = Little Knowledge, 4 = Some Basic Knowledge, 5 = Basic Knowledge, 6 = Detailed Knowledge, and 7 = Comprehensive Knowledge. The subjects in the social comparison condition were also instructed to compare the amount of knowledge they had compared to others in the class. Their Likert scale had anchors of 1 = Less than Anybody, 2 = Less than Most, 3 = Slightly Less than Others, 4 = About as much as Others, 5 = Slightly More than Others, 6 = More than Most, and 7 = More than Anybody.

Further, to assess the adequacy/effectiveness of the manipulations, a manipulation check was employed. The manipulation check for expectation of validation read "Is your self-assessment going to be compared to your test score?". The manipulation check for the purpose of the appraisal manipulation read "Will your self-assessment affect your grade in this class?".

Dependent Variable

The dependent variable (DV) in this study was the subjects' ratings of knowledge of psychology. In the study of leniency bias, this is often the construct of choice for the DV. Further, the situations which subjects were presented with in the present study is very similar to situations that they are faced with in the "real world" when self-assessments are collected and leniency may be expected. Differences between the experimental conditions could then be examined by comparing subject's self-assessments in the different conditions.

Development of Self-Assessment Questionnaires

Self-assessment questionnaire items (Appendix B, C, D) were developed so that they would directly represent the material that was covered in the classes. The dimensions that were covered in the questionnaire items were taken directly from the syllabus and/or study guides for each of the classes. Each questionnaire contained 34 items on which subjects rated their knowledge. Twenty-six of the thirty-four items were assessments of course content. The questionnaires were pilot tested to assure that instructions were clear and understandable and that the dimensions were relatively easy to rate.

Development of Lie Scales

The lie scale items (Appendix E) were developed by creating "fake" dimensions that appeared plausible for the material of the class. Eight of the 34 questions were bogus and constituted the lie scale. To ensure that the erroneous dimensions appeared plausible, they were appraised by 3 subject matter experts. These bogus items were dispersed among the real items and subjects rated them in the same manner as the real items. Further, subjects were not told of the possibility of the bogus items.

Results

Scale Development

Self-assessment questionnaires for all three classes were analyzed using factor analyses to determine the number of underlying factors. Reliability analysis were conducted to determine the internal consistency of each of the questionnaires.

Factor analyses with principle components extraction indicated the presence of one underlying factor for the self-assessment questionnaires in all three classes. The presence of one factor was determined by an examination of scree plots (Please see Table 1). The factor that emerged for class 1 accounted for 35.4% of the variance, the factor that emerged for class two accounted for 42.1% of the variance, and the factor that emerged for class 3 accounted for 38.9% of the variance.

Table 1

Factor Scree Plots

Self-Assessment Questionnaire (all classes)

Scree Plot -- Class 1

Factor Number

Scree Plot -- Class 2

Factor Number

26
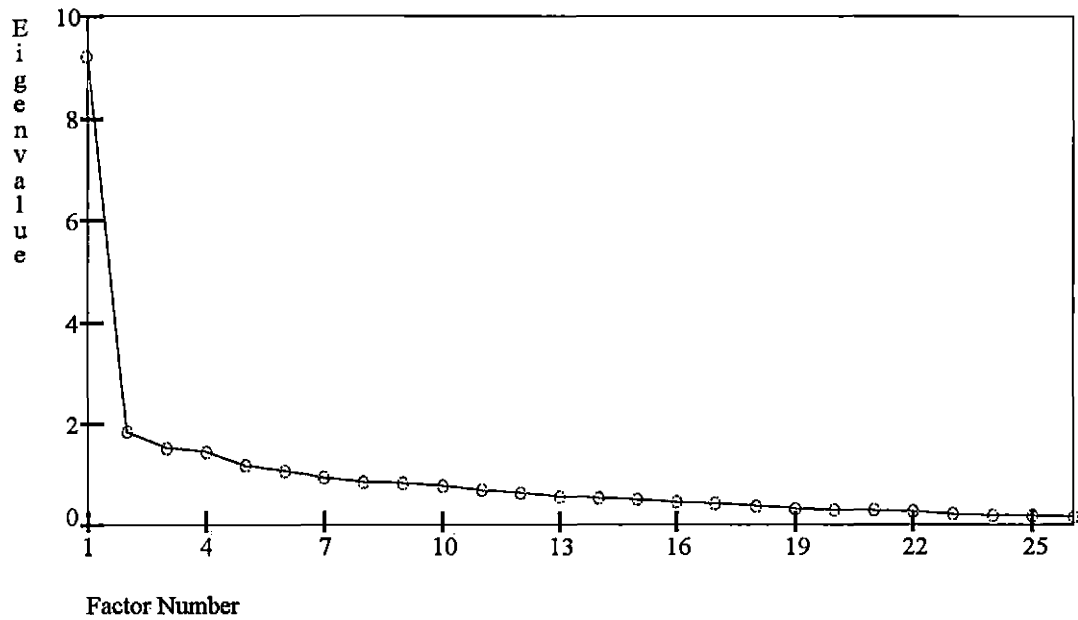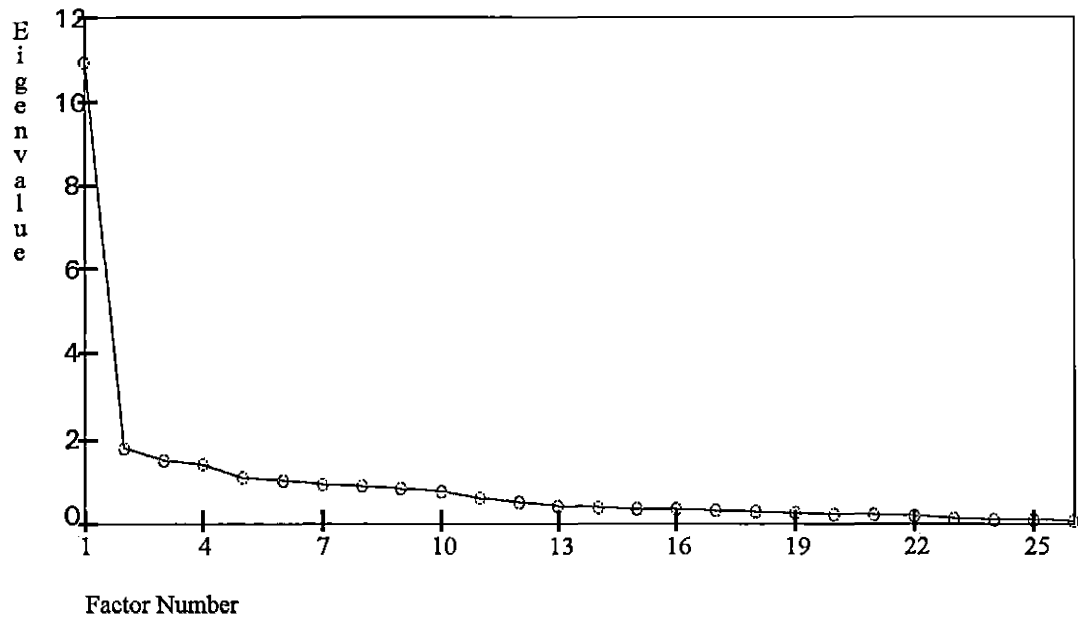
Scree Plot -- Class 3
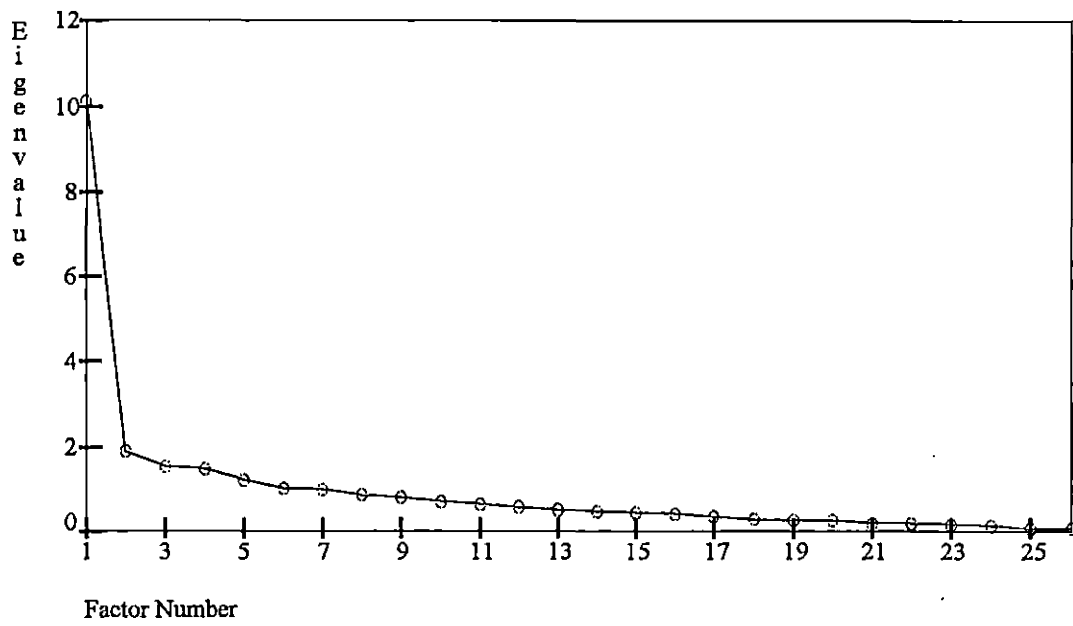
As can be seen from Tables 2, 3, and 4, factor loadings are quite high for all three classes, ranging from .35 to .73 (median of .62). The relatively low loadings for items 2 and 9 (Class 2 -- Table 3) may be due to the inclusion of ambiguous terms that were generally less understood by the class. Factor loadings for class 3 (Table 4) ranged from .26 to .76 (median of .62).

Table 2

Principle Components Analysis

Self-Assessment Questionnaire -- Class 1

Item	Factor Loading
Item 1	.493
Item 2	.429
Item 3	.639
Item 4	.535
Item 5	.730
Item 6	.572
Item 7	.675
Item 8	.650
Item 9	.461
Item 10	.646
Item 11	.358
Item 12	.714
Item 13	.630
Item 14	.666
Item 15	.666
Item 16	.460
Item 17	.621
Item 18	.594
Item 19	.695
Item 20	.527
Item 21	.610
Item 22	.641
Item 23	.677
Item 24	.685
Item 25	.443
Item 26	.421

Self-Assessment Questionnaire (Class 1): Bartlett Test of Sphericity = 1571.21 ($p < .001$); Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .877; N = 130; Variance Accounted for by Factor 1 = 35.4%.

Table 3

Principle Components Analysis

Self-Assessment Questionnaire -- Class 2

Item	Factor Loading
Item 1	.676
Item 2	.295
Item 3	.393
Item 4	.633
Item 5	.752
Item 6	.417
Item 7	.701
Item 8	.772
Item 9	.218
Item 10	.689
Item 11	.748
Item 12	.397
Item 13	.798
Item 14	.610
Item 15	.685
Item 16	.759
Item 17	.779
Item 18	.514
Item 19	.710
Item 20	.691
Item 21	.701
Item 22	.410
Item 23	.759
Item 24	.815
Item 25	.753
Item 26	.634

Self-Assessment Questionnaire (Class 2): Bartlett Test of Sphericity = 1291.27 ($p < .001$); Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .859; N = 79; Variance Accounted for by Factor 1 = 42.1%.

Table 4

Principle Components Analysis

Self-Assessment Questionnaire -- Class 3

Item	Factor Loading
Item 1	.669
Item 2	.490
Item 3	.578
Item 4	.703
Item 5	.588
Item 6	.721
Item 7	.608
Item 8	.635
Item 9	.496
Item 10	.564
Item 11	.581
Item 12	.761
Item 13	.632
Item 14	.697
Item 15	.267
Item 16	.560
Item 17	.759
Item 18	.740
Item 19	.544
Item 20	.556
Item 21	.621
Item 22	.656
Item 23	.725
Item 24	.658
Item 25	.597
Item 26	.594

Self-Assessment Questionnaire (Class 3): Bartlett Test of Sphericity = 1292.98 ($p < .001$); Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .819; N = 85; Variance Accounted for by Factor 1 = 38.9%.

Reliability Analysis

The reliability analysis for all three classes indicated that the three scales had high internal consistency (Please see Tables 5, 6, 7). Coefficient alpha for class 1 was .92, whereas class 2 had a coefficient alpha of .93. Item total correlations and squared multiple correlations are shown in Tables 5, 6, and 7, for classes 1, 2, and 3 respectively. For class 1, item-total correlations ranged from .34 to .67 (median of .58) and squared multiple correlations ranged from .32 to .72 (median of .54). For class 2, item-total correlations ranged from .16 to .77 (median of .61) and squared multiple correlations ranged from .31 to .80 (median of .67). Class 3 had item-total correlations that ranged from .24 to .73 (median of .56) and squared multiple correlations that ranged from .40 to .83 (median of .66).

Table 5

Reliability Analysis

Self-Assessment Questionnaire -- Class 1

Items	Corrected Item Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Item 1	.457	.548	.921
Item 2	.404	.422	.922
Item 3	.594	.590	.919
Item 4	.487	.370	.921
Item 5	.677	.666	.918
Item 6	.524	.504	.920
Item 7	.633	.564	.918
Item 8	.617	.540	.918
Item 9	.395	.486	.922
Item 10	.593	.551	.919
Item 11	.339	.375	.923
Item 12	.655	.722	.918
Item 13	.592	.594	.919
Item 14	.626	.639	.918
Item 15	.622	.585	.918
Item 16	.431	.439	.921
Item 17	.568	.526	.919
Item 18	.534	.629	.920
Item 19	.652	.657	.918
Item 20	.498	.486	.920
Item 21	.562	.511	.919
Item 22	.591	.518	.919
Item 23	.638	.579	.918
Item 24	.629	.603	.918
Item 25	.382	.414	.922
Item 26	.389	.325	.923

Alpha = .923; Standardized Item Alpha = .924; N = 130
Self-Assessment Questionnaire (Class 1): Mean = 111.32; Variance = 424.30;
Standard Deviation = 20.60.

Table 6

Reliability Analysis

Self-Assessment Questionnaire -- Class 2

Items	Corrected Item Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Item 1	.604	.628	.931
Item 2	.234	.536	.935
Item 3	.344	.459	.934
Item 4	.584	.588	.931
Item 5	.704	.733	.929
Item 6	.398	.644	.934
Item 7	.653	.763	.930
Item 8	.695	.748	.929
Item 9	.164	.313	.938
Item 10	.638	.652	.930
Item 11	.700	.750	.929
Item 12	.601	.554	.930
Item 13	.753	.787	.928
Item 14	.575	.667	.931
Item 15	.612	.636	.930
Item 16	.680	.717	.929
Item 17	.709	.800	.929
Item 18	.533	.544	.932
Item 19	.618	.768	.930
Item 20	.662	.738	.930
Item 21	.679	.674	.930
Item 22	.328	.457	.934
Item 23	.717	.698	.929
Item 24	.769	.756	.928
Item 25	.674	.683	.924
Item 26	.567	.597	.931

Alpha = .933; Standardized Item Alpha = .937; N = 79
Self-Assessment Questionnaire (Class 2): Mean = 127.58; Variance = 427.25;
Standard Deviation = 20.67.

Table 7

Reliability Analysis

Self-Assessment Questionnaire -- Class 3

Items	Corrected Item Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Item 1	.615	.690	.928
Item 2	.472	.503	.930
Item 3	.513	.669	.930
Item 4	.647	.758	.928
Item 5	.545	.488	.929
Item 6	.693	.694	.927
Item 7	.569	.630	.929
Item 8	.593	.755	.928
Item 9	.457	.599	.930
Item 10	.544	.522	.929
Item 11	.538	.662	.929
Item 12	.727	.699	.927
Item 13	.579	.677	.929
Item 14	.645	.700	.928
Item 15	.242	.409	.933
Item 16	.510	.493	.928
Item 17	.718	.791	.927
Item 18	.699	.823	.927
Item 19	.515	.619	.930
Item 20	.524	.676	.930
Item 21	.575	.586	.929
Item 22	.601	.690	.928
Item 23	.714	.829	.926
Item 24	.601	.652	.928
Item 25	.558	.643	.929
Item 26	.547	.615	.929

Alpha = .931; Standardized Item Alpha = .935; N = 85
Self-Assessment Questionnaire (Class 3): Mean = 125.36; Variance = 435.04;
Standard Deviation = 20.86.

Lie Scale Development

A factor analysis with principle components extraction was performed on the lie scale items to determine the number of underlying factors and a reliability analysis was performed to assess the internal consistency of the scale. Results of the principle components analysis identified one underlying factor that accounted for 57.2% of the variance. The presence of one factor was determined by an examination of the scree plot (Please see Table 8). Factor loadings for the eight items ranged from .52 to .84, with a median of .81 (Please see Table 9).

Table 8

Factor Scree Plot

Lie Scale

Scree Plot -- Lie Scale

Factor Number
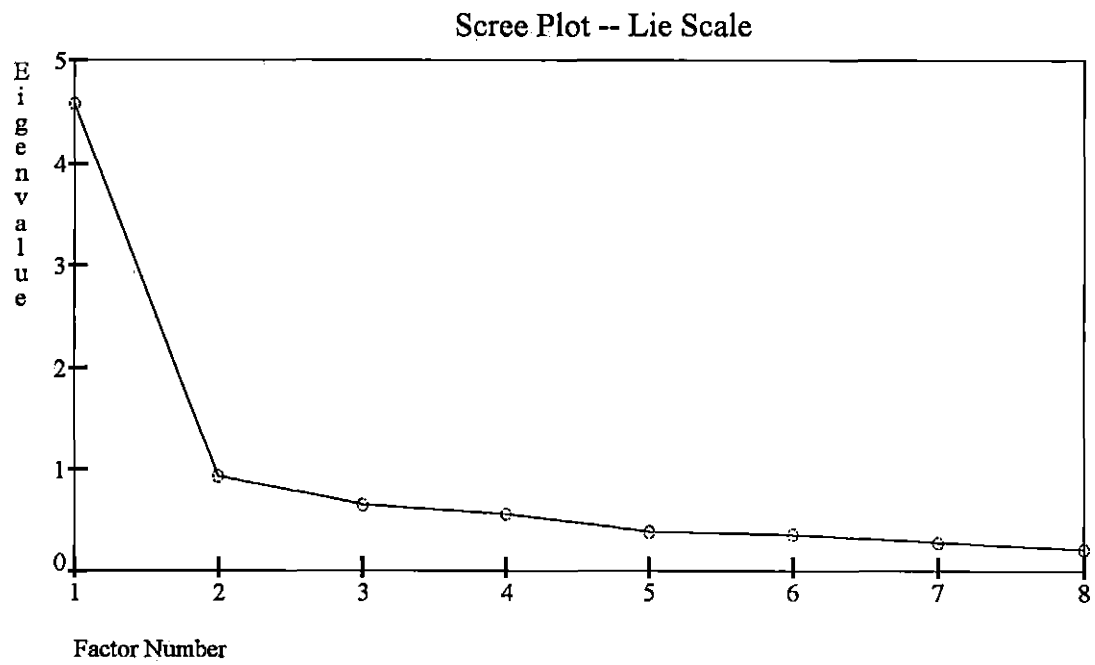
Table 9

Principle Components Analysis

Lie Scale

Lie Scale Item	Factor Loading
Item 1	.525
Item 2	.820
Item 3	:844
Item 4	.826
Item 5	.844
Item 6	.813
Item 7	.753
Item 8	.541

Lie Scale (all 3 classes): Bartlett Test of Sphericity = 1252.18 (p < .001); Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .895; N = 294; Variance Accounted for by Factor 1 = 57.2%.

Reliability Analysis

Reliability analysis performed on the lie scale items indicated a high level of internal consistency. Coefficient alpha for the eight items was .88. Squared multiple correlations, which ranged from .25 to .66 (median of .60), and item-total correlations, which ranged from .43 to .76 (median of .73), can be seen in Table 10.

Table 10

Reliability Analysis

Lie Scale

Items	Corrected Item Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Item 1	.439	.256	.892
Item 2	.748	.641	.861
Item 3	.764	.661	.860
Item 4	.750	.594	.859
Item 5	.747	.628	.860
Item 6	.724	.606	.863
Item 7	.654	.497	.869
Item 8	.474	.249	.889

Alpha = .884; Standardized Item Alpha = .890; N = 298
Lie Scale (all 3 classes): Mean = 20.06; Variance = 66.32;
Standard Deviation = 8.14.

Test of Hypotheses

To test hypotheses 1, 2, and 3, a 2 (social comparison versus absolute judgment) x 2 (high versus low expectation of validation) x 2 (research purposes versus grading purposes) Analysis of Variance (ANOVA) was performed on subject's self-assessment scores. Analysis was performed using SPSS for Windows (v.5.0.2).

Due to differences in means between classes (111, 127, and 125, for class one, two, and three respectively), a transformation to standardized scores was performed. The standard deviations and variances for all classes were very similar (Please see Table 11), By standardizing scores, it was possible to equalize the classes. Scores were transformed by creating z-scores (for each class separately), then multiplying the z-scores by 10, and adding a constant of 50 to create T-scores. T-scores were chosen because they are a widely used standardized score that allow for easy comparisons. After transformation, there were no significant differences between classes.

Table 11

Means, Standard Deviations, and Variances

for All Three Classes

Self-Assessment Questionnaire (without Lie Scale)

	Mean	Standard Deviation	Variance
Class 1 (n=130)	111.32	20.60	424.30
Class 2 (n=79)	127.58	20.67	427.25
Class 3 (n=85)	125.36	20.86	435.04

Results of evaluation of the assumptions of normality of sampling distributions, linearity, homogeneity of variance, homogeneity of regression, and outliers all proved satisfactory (Please see Table 12). Six cases were rejected due to missing data, therefore, 296 total cases were utilized in the analysis. Unequal cell sizes were adjusted by weighting means by cell.

.

Table 12

Results for the Assumptions of ANOVAs

Dependent variable: TSUM
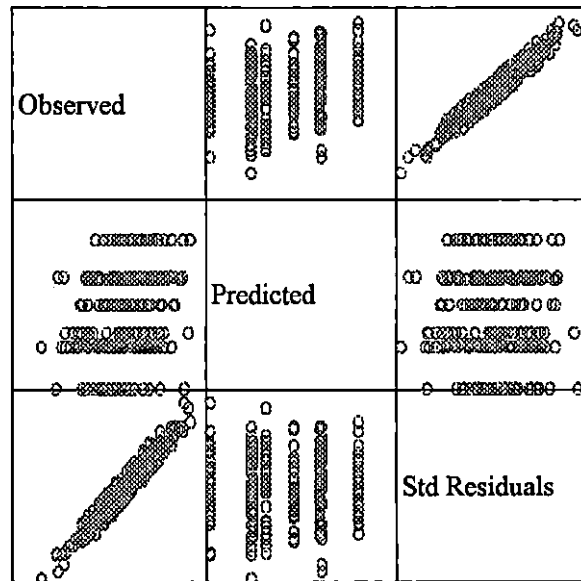
Hypothesis 1 predicted that subjects would be more likely to demonstrate leniency when they made comparisons on an absolute basis, compared to making a social comparison. The results supported this hypothesis, with $F (1, 295) = 7.02$, $p = .008$ (Please see Table 13 for means.) $\eta^2 = .024$.

Hypothesis 2 predicted that subjects would be less likely to demonstrate leniency when they are instructed that their self-assessments will be compared with an external criterion than when they are instructed that their self-assessments will not be compared with an external criteria. The results support this hypothesis, with $F (1, 295) = 4.19$, $p < .042$ (Please see Table 13 for means.) However, the strength of this relationship was weak, with $\eta^2 = .014$.

Hypothesis 3 predicted that individuals would be more likely to inflate their self-assessments if the purpose of the appraisal were involved in determining their grades in the class, rather than simply for research purposes. The results of the ANOVA support this hypothesis, with $F (1, 295) = 5.58$, $p = .019$ (Please see Table 13 for means). Again, the strength of this relationship was weak, with $\eta^2 = .019$.

Table 13

Means for 3 Conditions

Self-Assessment Questionnaire

		Mean
Comparison Condition	Social Comparison	48.5
	Absolute Rating	51.5
Expectation of Validation	High Expectation	48.8
	Low Expectation	51.2
Purpose of the Assessment	Research Purpose	48.7
	Grading Purpose	51.4

No statistically significant interaction effects were observed, with $\underline{F}$s ranging from .03 to .29 (ps ranged from .59 to .87). Variance accounted for (η^2) by the combined interactions did not total one percent.

Hypothesis 5 predicted that it would be possible to correct for leniency bias by using a lie scale as a covariate. To test this hypothesis, a 2 x 2 x 2 Analysis of Covariance (ANCOVA) was performed on standardized self-assessment questionnaire scores, with the lie scale score serving as the covariate. The independent variables were comparison condition, expectation of validation, and purpose of the appraisal. The correlation between the lie scale and the standardized self-assessment questionnaire scores was .31 (p < .001), indicating that the covariate was significantly related to the self-assessment scores.

After adding an individual's tendency to inflate self-ratings, the variance accounted for by the analysis went from 4 percent to 15 percent (R^2 went from .03 to .15). After correction, the social comparison condition remained significant, as did the purpose of the appraisal condition. However, interestingly, the $\underline{F}$ for the expectation of validation condition no longer reached significance, $\underline{p}$ = .189 (Please see Table 14 for observed and adjusted means of the eight cells).

Table 14

Observed and Adjusted Means

		Observed Mean	Adjusted Mean
Comparison Condition	Social Comparison	48.5	47.6
	Absolute Rating	51.5	52.4
Expectation of Validation	High Expecation	48.8	49.3
	Low Expectation	51.2	50.7
Purpose of the Assessment	Research Purpose	48.7	48.8
	Grading Purpose	51.4	51.2
Cell 1	Social Comparison High Expectation Research Purpose	46.1	45.9
Cell 2	Social Comparison High Expectation Grading Purpose	49.0	48.0
Cell 3	Social Comparison Low Expectation Research Purpose	48.3	49.6
Cell 4	Social Comparison Low Expectation Grading Purpose	52.0	53.8
Cell 5	Absolute Rating High Expectation Research Purpose	48.4	47.6
Cell 6	Absolute Rating High Expectation Grading Purpose	50.5	49.3
Cell 7	Absolute Rating Low Expectation Research Purpose	51.8	52.4
Cell 8	Absolute Rating Low Expectation Grading Purpose	54.0	53.6

Discussion

As hypothesized, and consistent with past research (Farh & Dobbins, 1989b; Felson & Reed, 1986; Mabe & West, 1982), this study found less lenient appraisals when individuals made their self-assessments on a social comparison basis than when they provided an absolute judgment. The most probable explanation for this finding is that social comparison gives an individual a context in which to judge his or her knowledge or performance. Students in classes talk to each other, study together, and gain information from each other in a variety of ways. Therefore, rather than ask someone to judge his or her knowledge on a scale that is not well defined, people rank themselves amongst peers with whom they know. These results are even more promising for the prognosis of self-assessments and self-appraisals if you consider that in an applied setting, where people have detailed knowledge of peers' knowledge and performance capabilities, an individual would be able to rate his or her own knowledge or performance even more accurately.

In concordance with previous research (Farh & Werbel, 1986; Fox & Dinur, 1988; Mabe & West, 1982; Weary & Arkin, 1982), when an individual believes that his or her self-assessment will be examined and compared to external criterion (test scores in this case) there tends to be less leniency in their self-ratings. The most intuitive reason for this occurrence is that individuals do not want to appear dishonest to others and risk public scrutiny or ridicule. Conversely, when an individual does not believe that his or her self-assessment will be compared to external criteria, or there is no risk in inflating their scores. There are many reasons that an individual may want to present themselves in a favorable

49

manner: to receive a promotion, to get a better grade in a class, to get the more preferable

job assignment, or simply to boost one's ego. No matter what the motivation to appear

desirable, the result is leniency bias, which is much more pervasive under conditions of

low expectations of validation.

The present study affirmed previous research (Farh & Werbel, 1986; Harris, Smith,

& Champagne, 1995; Holzbach, 1978; Klimoski & London, 1974; Lawler, 1968; Parker,

Taylor, Barrett, & Martens, 1959; Prien & Liske, 1962; Thorton, 1968) regarding the role

of the purpose of the appraisal on leniency in self-assessments. Self-assessments were

found to be less lenient when they were conducted for research purposes than when they

were being conducted for an administrative (grading in this case) purpose. The logic

behind this finding is obvious when you consider the rater's goals when completing his or

her self-assessment. Although this finding does not lend itself to changes or modifications

in how self-assessments should be administered, it does allow academicians and

practitioners with the knowledge that they must "guard" against leniency if they are using

self-assessments or self-appraisals for a purpose that can be associated with a desirable

goal for the rater.

The findings of the current research are analogous to the findings of Anderson,

Warner, and Spencer (1984), who demonstrated that it was possible to correct for inflation

by incorporating a lie scale into the self-assessment questionnaire. The results of the

present study demonstrate that it is possible to correct for leniency by utilizing a lie scale

as a covariate. After correcting for an individual's tendency to inflate his or her self-

50

assessments, the variance accounted for by the analysis jumped from five percent to fifteen percent.

There are two additional findings that arise from the use of a lie scale that are interesting. First, the means of the two groups in the social comparison condition became more disparate. Specifically, the social comparison condition mean was adjusted downward (48.5 to 47.6) and the absolute judgment condition mean was adjusted upward (51.5 to 52.3). Secondly, the groups in the expectation of validation condition were shifted so that they were closer together. The most plausible explanation for this is that when individuals are making social comparison, they often believe that they know more than their peers, but they would not be willing to announce this publicly. However, on a questionnaire, they would not hesitate to indicate this feeling. This seems to contradict the finding that assessments based on social comparisons demonstrate less leniency, but in fact it does not. The reason that social comparison self-assessments demonstrate less leniency is that the gain from having a comparison point outweighs the tendency for people to overestimate their abilities.

The implications of this finding are advantageous to academicians and practitioners when developing a self-assessment questionnaire, in that it may no longer necessary to foster an expectation in people that their self-assessments will be validated. Rather than hope that people will be honest and give accurate ratings, it is possible to statistically correct for an individual's penchant to be lenient. What this means to anyone developing a self-assessment measure is that it may not be necessary to develop instructional sets that

attempt to manipulate the beliefs of everyone who complete the instrument, rather it is only necessary to create some bogus items that can be used to correct for an individual's tendency to inflate their ratings.

One suggestion resulting from the present study would be to examine whether these findings are replicable in an organizational setting. In reference to the purpose of the appraisal, even though it is something that would not be easy to manipulate in an organization, it is argued that this effect would be more pronounced since an individual would be more inclined to associate the outcome of the self-assessment with a favorable situation (promotion, raise, etc.). This inference magnifies the possible usefulness of lie scale, in that, if people are more inclined to be lenient, then any correctional method becomes more valuable. In light of this, it appears that the best use for self-assessments, from an organizational perspective, remains for identification of training needs. However, this research does provide promise for self-assessments with respect to selection.

Recommendations for Future Research

Future research should examine the effect of such manipulation in applied organizational settings to determine whether the same effects can be realized. There may be differing motivational processes taking place when an individual is assessing him or herself and the results could have a direct impact on retaining a job. Speculation would lead one to believe that the effects observed in the present study would be magnified, but applied research is needed to elucidate all effects.

Another possibility for future research would be to examine whether the effects demonstrated in the comparison condition were the result of experimental manipulations (i.e., the instructions subjects received) or the result of a measurement phenomenon. One way to examine this would be to employ a self-assessment questionnaire that utilized an absolute rating and a lie scale that utilized a social comparison rating scale (and vice versa). If similar results are obtained for the different types of scales in both comparison conditions, there would be more evidence that suggested that the results truely were a consequence of the instructional manipulation.

Additionally, future research should examine whether conditions in which people make social comparisons lead to less lenient self-ratings when compared to absolute judgments when there are well defined performance dimensions (e.g., when people have an understanding of what constitutes good, average, and poor performance). Further, the present study simply examined self-assessments of knowledge of a particular topic. Future research should examine the generalizability of the current results. In other words, future research should attempt to identify the different types of situations that these findings apply to. Moreover, future research should examine the effect of such manipulations on the correlations with other sources of ratings (e.g., supervisor ratings).

Frame of reference training, or more specifically, clarified dimensional definitions, should be examined in conjunction with the different experimental manipulations to examine if the effects still hold true when individuals have a good understanding of what ambiguous absolute anchors mean.

Perhaps most important to organizations, future research should address whether such manipulations can result in self-assessments and self-appraisals adding to a selection process. In other words, is it possible to improve self-assessments and self-appraisals consistently so as to result in a selection tool that adds incremental validity to other selection measures such as test batteries or assessment centers?

Future research related to self-assessments should focus on the different types of rater training and see which is the best at reducing leniency. Based on past research, one would expect that training individuals on performance dimensions (what is good, average, and poor) would have the most benefit. Further, it would be interesting to see what effect rater training would have on the variable manipulated in the present study.

One last recommendation for future research would be to examine what is the effect on an individual's tendency to be lenient when you merely inform them that there may be a lie scale in the self-assessment. It would be interesting to see whether or not this would reduce the inflation associated with self-assessments.

Summary and Conclusions

As a whole, this research suggests that self-assessments are best used for research purposes, on a social comparison basis, when the self-rater has expectations that the ratings will be checked against other criteria. In situations where there is a desirable outcome for the self-rater (selection), an individual often tries to paint a better picture of him or herself. However, the results of this study suggest that it is possible to use self-assessments across a variety of situations (selection) as long as one is careful to construct

the self-assessment questionnaire in such a way as to minimize the chances that someone will inflate their ratings (e.g., imply that their self-assessments will be checked against an external criterion). Additionally, this study demonstrated that it is possible to correct for leniency by using a lie scale as a covariate, that is, make statistical corrections that result in less lenient ratings.

Although this research demonstrated that self-assessments can be collected that are not excessively lenient, it is recommended that self-assessments be used in conjunction with other forms of assessment (e.g., interviews, supervisor appraisals), rather than in isolation. It is hoped that this research demonstrated that self-assessments and self-appraisals can be used as additional sources of information when gathering information about an individual since they are a valuable source of information, they are relatively inexpensive, and it is possible to reduce the amount of leniency, further, that they not be abandoned due to faulty beliefs about the lack of their veracity.

APPENDIX A

Example of Each Instructional Set

Low Expectation of Validation - Research Purpose - Social Comparison

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way effect anyone's grade in this class. Self-assessments are a valuable source of information and can often give accurate assessments of one's abilities. Self-assessments can often provide information that tests cannot because tests do not perfectly measure a person's knowledge. Therefore, the results of this questionnaire will at no time be compared to test scores. Further, the results of this survey will be reported only as class data, no individual responses will be examined. Please use the following scale to rate your knowledge compared to the knowledge of other people in the class.

Low Expectation of Validation - Research Purpose - Absolute Rating

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way effect anyone's grade in this class. Self-assessments are a valuable source of information and can often give accurate assessments of one's abilities. Self-assessments can often provide information that tests cannot due to the fact that tests do not perfectly measure a persons knowledge. Therefore, the results of this questionnaire will at no time be compared to test scores. Further, the results of this survey will be reported only as class data, no individual responses will be examined. Please use the following scale to rate your knowledge on each of the questions on the questionnaire. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

High Expectation of Validation - Research Purpose - Social Comparison

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way effect anyone's grade in this class. To ensure accuracy, the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class. Please use the following scale to rate your knowledge compared to the knowledge of other people in the class.

Low Expectation of Validation - Grading Purpose - Social Comparison

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The instructor of this class has decided that <u>borderline grades will be determined by your self-assessment</u>. In other words, if you are on the border of an A and A-, A- and B+, B+ and B, etc. Self-assessments are a valuable source of information and can often provide an accurate assessment of ones abilities. Self-assessments can often provide information that tests cannot due to the fact that tests do not perfectly measure a persons knowledge. Therefore, <u>the results of this questionnaire will at no time be compared to test scores</u>. Please use the following scale to rate your knowledge compared to the knowledge of other people in the class.

Low Expectation of Validation - Grading Purpose - Absolute Rating

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The instructor of this class has decided that <u>borderline grades will be determined by your self-assessment</u>. In other words, if you are on the border of an A and A-, A- and B+, B+ and B, etc. Self-assessments are a valuable source of information and can often give accurate assessments of ones abilities. Self-assessments can often provide information that tests cannot due to the fact that tests do not perfectly measure a persons knowledge. Therefore, <u>the results of this questionnaire will at no time be compared to test scores</u>. Please use the following scale to rate your knowledge on each of the questions on the survey. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

High Expectation of Validation - Grading Purpose - Social Comparison

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The instructor of this class has decided that <u>borderline grades will be determined by your self-assessment</u>. In other words, if you are on the border of an A and A-, A- and B+, B+ and B, etc., your grade will be determined by your self-assessment. To ensure accuracy, <u>the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class</u> and then determine your final grade. Please use the following scale to rate your knowledge compared to the knowledge of other people in the class.

High Expectation of Validation - Grading Purpose - Absolute Rating

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The instructor of this class has decided that borderline grades will be determined by your self-assessment. In other words, if you are on the border of an A and A-, A- and B+, B+ and B, etc., your grade will be determined by your self-assessment. To ensure accuracy, the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class and then determine your final grade. Please use the following scale to rate your knowledge on each of the questions on the survey. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

High Expectation of Validation - Research Purpose - Absolute Rating

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way effect anyone's grade in this class. To ensure accuracy, the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class. Please use the following scale to rate your knowledge on each of the questions on the survey. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

APPENDIX B

Example of Self-Assessment Questionnaire

Class 1

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way effect anyone's grade in this class. To ensure accuracy, the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class. Please use the following scale to rate your knowledge on each of the questions on the survey. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Endocrine System (including the thyroid, adrenal, and pituitary glands)

1 2 3 4 5 6 7   Early schools of Psychology (i.e., Structuralism, Functionalism, Behaviorism, Psychoanalysis, Gestalt, etc.)

1 2 3 4 5 6 7   Reversible validity

1 2 3 4 5 6 7   Different functions of the brain (including thought, language, and left/right brain functions)

1 2 3 4 5 6 7   Role of heredity in Psychology (genes, chromosomes, twin studies)

1 2 3 4 5 6 7   Structural Thantosis Syndrome

1 2 3 4 5 6 7   Different methods of observation (including survey research, case studies, naturalistic observation, etc.)

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Experimental Psychology (IVs, DVs, counterbalancing, etc.)

1 2 3 4 5 6 7   Use of subjects in research and other ethical considerations

1 2 3 4 5 6 7   Current disciplines of Psychology (i.e., Biopsychology, Cognitive Psychology, Psychodynamic, Learning Theories, Humanistic-Existential Psychology, etc.)

1 2 3 4 5 6 7   Advantages of ventralmedial color deafness

1 2 3 4 5 6 7   Different careers available to Psychologists (e.g., Clinical-Counseling, Industrial/Organizational, Experimental, etc.)

1 2 3 4 5 6 7   Basic concepts of Sensation and Perception (including absolute thresholds, difference thresholds, etc.)

1 2 3 4 5 6 7   Different visual perceptual concepts (such as the "autokinetic effect" and "phi-phenomenon")

1 2 3 4 5 6 7   Difference between a sample and a population

1 2 3 4 5 6 7   Different measures of central tendency

1 2 3 4 5 6 7   Test-retest reliability

1 2 3 4 5 6 7   Schizotypal method of personality prevention

1 2 3 4 5 6 7   Difference between reliability and validity

1 2 3 4 5 6 7   Pioneers of Psychology (e.g., Cattell, Freud, James, Thorndike, Watson, Wundt, etc.)

1 2 3 4 5 6 7   Principles of Operant Conditioning (e.g., positive reinforcement, negative reinforcement, primary reinforcers)

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Principles of Classical Conditioning (e.g., US, CS, UCS, UCR, etc.)

1 2 3 4 5 6 7   Principles of Modeling and Shaping

1 2 3 4 5 6 7   Symptoms of Ansoorian's Gender Conversion Syndrome

1 2 3 4 5 6 7   Different kinds of memory (i.e., episodic, semantic, procedural, etc.)

1 2 3 4 5 6 7   Processes involved in memory

1 2 3 4 5 6 7   Sachau's law of triangular determinism

1 2 3 4 5 6 7   Effects of drugs on consciousness

1 2 3 4 5 6 7   A negative correlation

1 2 3 4 5 6 7   Purpose of having a control group

1 2 3 4 5 6 7   Effects of undergoing Psychogenesis prior to puberty

1 2 3 4 5 6 7   Tip-of-the-tongue phenomenon

1 2 3 4 5 6 7   Symptoms of Catatonic Schizophrenia

1 2 3 4 5 6 7   Generalized specificity

                                                                    Circle One

Is your self-assessment going to be compared to your test score?   YES   NO

Will your self-assessment affect your grade in this class?   YES   NO

APPENDIX C

Example of Self-Assessment Questionnaire

Class 2

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The instructor of this class has decided that borderline grades will be determined by your self-assessment. In other words, if you are on the border of an A and A-, A- and B+, B+ and B, etc. Self-assessments are a valuable source of information and can often give accurate assessments of ones abilities. Self-assessments can often provide information that tests cannot due to the fact that tests do not perfectly measure a persons knowledge. Therefore, the results of this questionnaire will at no time be compared to test scores. Please use the following scale to rate your knowledge on each of the questions on the survey. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comprehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Factors influencing attachment

1 2 3 4 5 6 7   Children's perceptual development, including perceptual efficiency

1 2 3 4 5 6 7   Reversible validity

1 2 3 4 5 6 7   Intelligence as defined by Galton

1 2 3 4 5 6 7   Self-concept as defined by James

1 2 3 4 5 6 7   Structural Thantosis Syndrome

1 2 3 4 5 6 7   Temperament

1 2 3 4 5 6 7   Differences between morphemes and phonemes

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Language acquisition

1 2 3 4 5 6 7   Mohr's developmental model of self-concept and identity

1 2 3 4 5 6 7   Advantages of ventralmedial color deafness

1 2 3 4 5 6 7   Usefulness of recasting

1 2 3 4 5 6 7   Culture-fair test

1 2 3 4 5 6 7   Effect of parenting style on intelligence

1 2 3 4 5 6 7   Gender bias

1 2 3 4 5 6 7   Dimensions of temperament

1 2 3 4 5 6 7   Children's exploration phenomena

1 2 3 4 5 6 7   Schizotypal method of personality prevention

1 2 3 4 5 6 7   Different types of reliability

1 2 3 4 5 6 7   Linguistic interactions

1 2 3 4 5 6 7   Controversies in intelligence

1 2 3 4 5 6 7   Sexual identity

1 2 3 4 5 6 7   Effects the environment has on a child's reading ability

1 2 3 4 5 6 7   Symptoms of Ansoorian's Gender Conversion Syndrome

1 2 3 4 5 6 7   Cultural influences on identity

1 2 3 4 5 6 7   Sociocultural influences affecting language acquisition

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Sachau's law of triangular determinism

1 2 3 4 5 6 7   Disorganized attachment

1 2 3 4 5 6 7   Role syntax plays in language

1 2 3 4 5 6 7   Defining features of avoidant attachment

1 2 3 4 5 6 7   Effects of undergoing Psychogenesis prior to puberty

1 2 3 4 5 6 7   Biological foundations of reading

1 2 3 4 5 6 7   Different types of validity

1 2 3 4 5 6 7   Generalized specificity

                          Circle One

Is your self-assessment going to be compared to your test score? YES NO

Will your self-assessment affect your grade in this class?   YES NO

Example of Self-Assessment Questionnaire

Class 3

This self-assessment questionnaire is designed to assess students' knowledge of Psychology. The results of this questionnaire will be used to improve the instruction of this course in the future and will in no way effect anyone's grade in this class. To ensure accuracy, the instructor of this class will compare your self-assessments with the grades you receive on the tests in this class. Please use the following scale to rate your knowledge on each of the questions on the survey. The scale is designed to represent a continuum of knowledge with 1 equaling the least amount of knowledge and 7 representing the most knowledge.

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Adolescent egocentrism

1 2 3 4 5 6 7   Centration

1 2 3 4 5 6 7   Reversible validity

1 2 3 4 5 6 7   Different memory aids, including mnemonics and chunking

1 2 3 4 5 6 7   Influence peer groups have on behavior

1 2 3 4 5 6 7   Structural Thantosis Syndrome

1 2 3 4 5 6 7   Stages of occupational development

1 2 3 4 5 6 7   Concrete reasoning

1 2 3 4 5 6 7   Kohlberg's levels of moral development

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comprehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Features of identity, including diffusion, closure, achievement, and foreclosure

1 2 3 4 5 6 7   Advantages of ventralmedial color deafness

1 2 3 4 5 6 7   Adolescent sleep patterns

1 2 3 4 5 6 7   Different control processes

1 2 3 4 5 6 7   Fluid intelligence

1 2 3 4 5 6 7   Identity formation

1 2 3 4 5 6 7   Characteristics of adolescent drug abusers

1 2 3 4 5 6 7   Effect of group pressures on conformity

1 2 3 4 5 6 7   Schizotypal method of personality prevention

1 2 3 4 5 6 7   Stages of marital choice

1 2 3 4 5 6 7   Authoritarian versus autocratic parenting styles

1 2 3 4 5 6 7   Dialectical thinking

1 2 3 4 5 6 7   Crystallized intelligence

1 2 3 4 5 6 7   Behavioral contagion

1 2 3 4 5 6 7   Symptoms of Ansoorian's Gender Conversion Syndrome

1 2 3 4 5 6 7   Dodge's model of group entry

1 2 3 4 5 6 7   Effect that family size has on intelligence

1	2	3	4	5	6	7
No Knowledge	Very Little Knowledge	Little Knowledge	Some Basic Knowledge	Basic Knowledge	Detailed Knowledge	Comp-rehensive Knowledge

My understanding of (the):

1 2 3 4 5 6 7   Sachau's law of triangular determinism

1 2 3 4 5 6 7   Information processing

1 2 3 4 5 6 7   Hoffman's stages of empathy development

1 2 3 4 5 6 7   Characteristics of popular children

1 2 3 4 5 6 7   Effects of undergoing Psychogenesis prior to puberty

1 2 3 4 5 6 7   Formal operations

1 2 3 4 5 6 7   Moratorium

1 2 3 4 5 6 7   Generalized specificity

                                                     Circle One

Is your self-assessment going to be compared to your test score?   YES   NO

Will your self-assessment affect your grade in this class?   YES   NO

APPENDIX E

Lie Scale Items -- All Classes

Reversible validity

Structural thantosis syndrome

Advantages of ventralmedial color deafness

Schizotypal method of personality prevention

Symptoms of Ansoorian's gender conversion syndrome

Sachau's law of triangular determinism

Effects of undergoing psychogenesis prior to puberty

Generalized specificity

APPENDIX F

Debriefing Statement

DEBRIEFING

The study you have just participated in is designed to determine whether or not self-assessments demonstrate a leniency bias, and if so, whether or not this bias be reduced. Leniency bias exists when a rating by one source is consistently higher than that by another source (e.g., self-ratings are consistently higher than supervisor ratings).

The instructions you received prior to completing the self-assessment questionnaire were manipulated so as to determine the situations in which self-assessments are most lenient. Your responses to the self-assessment items, contrary to what may have been written in the instructions, will at no time be used to determine your grade in this class. Additionally, an individual's responses to the self-assessment questionnaire will at no time be examined in direct comparison with his or her grade on the exam. Furthermore, the instructor of this class will at no time see your responses to the self-assessment items. The only person who will have access to the self-assessments is a Michael R. Dolen, a graduate student at CSUSB, and at no time will he make this data available to other sources.

In order to ensure the confidentiality of your responses, your name will at no time be connected with the individual responses. The data gathered will only be analyzed as group data.

If you have any questions that cannot be answered by the individual administering this survey, or if you would like to obtain the general results of the study, please contact Dr. Janet L. Kottke at California State University, San Bernardino. Dr. Kottke can be reached in her office (PS-116), through voice mail at (909) 880-5585, or at her e-mail address (jkottke@wiley.csusb.edu).

Thank you for your cooperation in completing this survey.

APPENDIX G

Oral Informed Consent Form

ORAL INFORMED CONSENT FORM.

Part of the materials you complete today will provide data for a study being conducted by a graduate student at CSUSB. If you have any reservation about participating in this study, please see me immediately and you may withdraw, at anytime, from completing those portions of the materials.

After you have completed and turned in all materials, you will receive a written description of the project and can ask any questions that you may have. If you have a question that cannot be answered by the persons here today, feel free to contact Dr. Janet L. Kottke. I have information on how to contact Dr. Kottke, should you feel the need to do so.

Participation in this study is indicated by completion of the materials that are handed out to you, and I thank you for your cooperation.

References

Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology, 69* (4), 574-580.

Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Level-of-processing theory and social facilitation theory perspectives. *Journal of Applied Psychology, 72,* 239-244.

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist, 33,* 344-358.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74,* 183-200.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review, 6,* 205-212.

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology, 65,* 60-66.

Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology, 62,* 64-69.

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliabilities and validities of performance evaluation ratings. Journal of Applied Psychology, 60 (5), 1011-1021.

Cleveland, J., & Murphy, K. R. (1992). Analyzing performance appraisal as goal-directed behavior. In Ferris G, Rowland K. R. (Eds.), *Research in personnel and human resources management* (pp. 121-185). Greenwich, CT: JAI Press.

Dunning, D., Griffin, D. W., Milojkovic, J., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58,* 568-581.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57* (6), 1082-1090.

Farh, JL., & Dobbins, G. H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology*, 42, 835-850.

Farh, JL., & Dobbins, G. H. (1989b). Effects of comparative performance information on the accuracy of self-ratings and agreement between self- and supervisor ratings. *Journal of Applied Psychology*, 74 (4), 606-610.

Farh, JL., & Werbel, J. D. (1986). Effects of purpose of the appraisal and expectation of validation on self-appraisal leniency. *Journal of Applied Psychology*, 71 (3), 527-529.

Felson, R. B. (1981). Ambiguity and bias in the self-concept. *Social Psychology Quarterly*, 44, 64-69.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford.

Fishbein, M. (1965). *Current Studies in Social Psychology*. New York: Holt, Rinehart, and Winston.

Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology*, 41, 581-592.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603-618.

Harris, M. M., Smith, D. E., & Champagne, D. (1995). A field study of performance appraisal purpose: Research- versus administrative-based ratings. *Personnel Psychology*, 48, 151-160.

Hedge, J. W., & Kavanaugh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, 73, 68-73.

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

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3, 1-2.

Jones, E. E., & Nisbett, R. E. (1971). *The actor and the observer: Divergent perceptions of the cause of behavior.* Morristown, NJ: General Learning Press.

Keefer, K. E. (1969). Self-prediction of academic achievement by college students. *Journal of Educational Research, 63,* 53-56.

Kirchner, W. K. (1966). Relationship between supervisory and subordinate ratings for technical personnel. *Journal of Industrial Psychology, 3,* 57-60.

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

Lawler, E. E. III. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of applied Psychology, 51,* 369-380.

Mabe, P. A. III., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67* (3), 280-296.

Nisbett, R. E., Caputo, C., Legant, P., & Maracek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology, 27,* 154-164.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62,* 42-55.

Parker, J. W., Taylor, E. K., Barrett, R. S., & Martens, L. (1959). Rating scale Content: III. Relationship between supervisory- and self-ratings. *Personnel Psychology, 12,* 49-63.

Prien, E. P. & Liske, R. E. (1962). Assessments of higher-level personnel: III. Rating Criteria. A comparative analysis of supervisory ratings and incumbent self-ratings of job performance. *Personnel Psychology, 15,* 187-194.

Pulankos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69,* 581-588.

Pulankos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior and Human Decision Processes, 38,* 76-91.

Raven, B. H., & Fishbein, M. (1965). Social referents and self-evaluation in examinations. *Journal of Social Psychology, 65,* 89-99.

Schmitt, P. V. & Strein, W. (1987). Self-ratings versus actual WAIS-R performance. *Psychological Reports, 61,* 905-906.

Stamoulis, D. T., & Hauenstein, N. M. A. (1993). Rater training and rating accuracy: Training for dimensional accuracy versus training for ratee differentiation. *Journal of Applied Psychology, 78,* 994-1003.

Thorton, G. C., III. (1968). The relationship between supervisor- and self- appraisals of executive performance. *Personnel Psychology, 21,* 441-455.

Thorton, G. C., III. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 263-271.

Waldman, D. A. & Thorton, G. C., III. (1979). A comparison of supervisors' self-appraisals and their administrators' appraisals. *Medical Group Management, 26,* 22, 24, 58.

Weary, G., & Arkin, R. M. (1982). Attributional self-presentation. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New Directions in attribution research (Vol. 3).* Hillsdale, N.J.: Erlbaum.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale - Revised,* manual. New York: Psychological Corporation.

Wells, L. E., & Sweeney, P. D. (1986). A test of three models of bias in self-assessment. *Social Psychology Quarterly, 49,* 1-10.

Williams, J. R., & Levy, P. E. (1992). The effects of perceived system knowledge on the agreement between self-raings and supervisor ratings. *Personnel Psychology, 45,* 835-847.

Woehr, D. J. (1992). Performance dimension accessibility: Implications for rating accuracy. *Journal of Organizational Behavior, 13,* 357-367.

Zuckerman, M. (1979). Attribution of success and failure revisited: Or the motivational bias is alive and well in attribution theory. *Journal of Personality, 47,* 245-287.