1980

Personality and ability characteristics of ski instructors and their relationship to the effective teaching of skiing skills

Carol Austin Bridgwater
The University of Montana

COPYRIGHT ACT OF 1976

THIS IS AN UNPUBLISHED MANUSCRIPT IN WHICH COPYRIGHT SUB-
SISTS. ANY FURTHER REPRINTING OF ITS CONTENTS MUST BE APPROVED
BY THE AUTHOR.

PERSONALITY AND ABILITY CHARACTERISTICS
OF SKI INSTRUCTORS
AND THEIR RELATIONSHIP TO THE
EFFECTIVE TEACHING OF SKIING SKILLS

By

Carol Austin Bridgwater

B.A., Stanford University, 1976

Presented in partial fulfillment
of the requirements for the degree of

Master of Arts
University of Montana
1980

Approved by:

Xames A. Walsh
Chairman, Board of Examiners

R C Murray
Dean, Graduate School

_6-11-80_
Date

UMI Number: EP36034

UMI®

Dissertation Publishing

UMI EP36034

ProQuest®

As the product of my first independent

research endeavor, this thesis is dedicated

in memory of my father,

Dr. George A. Austin.

His contribution to the field of psychology

is not forgotten and has often served as an

inspiration in my pursuit of higher education.

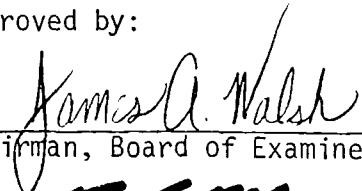## ACKNOWLEDGEMENTS

Bridgwater, Carol Austin, M.A., June 1980   Psychology

Personality and Ability Charateristics of Ski Instructors and Their Relationship to the Effective Teaching of Skiing Skills (92 pp.)

Director: James A. Walsh

The purpose of this study was threefold. The primary purpose was to develop a reliable and valid measure of teacher effectiveness in ski instructors. Secondly, normative data on the Personality Research Form was gathered for ski instructors as a specific occupational group. Third, the efficacy of PRF scales as predictors of teacher effectiveness was investigated.

A 51 item rating scale was validated on criterion groups of effective, mediocre, and ineffective ski instructors (n=72) from eight Rocky Mountain ski resorts. The criterion measure was the director's overall rating of instructors' general teacher effectiveness. Global ratings were also obtained from up to four supervisors at each resort. A correlation of .94 between supervisor and director overall ratings revealed high reliability for the criterion measure. Directors then rank ordered their instructors on each of the 51 items, with total score on the rating scale equaling the sum of the ranks.

Evidence for criterion related validity was established through group separation. An analysis of variance yielded significant differences between total scores for the three criterion groups (p < .001). Post-hoc analysis by means of the Newman-Keuls test showed significant differences at the .001 level for every pair of means. Further evidence for the scale's validity was shown by the correlation of .87 between total score and the criterion measure. Finally, an internal cross validation procedure using two equal random samples (n=36) yielded similar results. The scale's reliability was established by a coefficient alpha of .99, and a mean item correlation of .70 with total score.

The PRF was administered to 118 ski instructors. Their average scores differed significantly from PRF norms on 17 of the 21 scales, with mean differences ranging in significance from .05 to .0001. Twenty-eight of these subjects were also rated on the effectiveness rating scale. Several PRF scales correlated significantly with levels of teaching effectiveness, and two scales were significant predictors of teaching effectiveness using univariate regressions. However, a step-up multiple regression showed total score on the effectiveness scale, and level of instructor certification to be the most efficient combination of predictor variables for teaching effectiveness.

The study demonstrated that a reliable and valid measure of ski instructor teaching effectiveness could be constructed. The scale's high internal consistency revealed the homogeniety of the construct. The significant differences between ski instructors and PRF norms reiterate that college normed summary statistics must be applied with caution to noncollege populations.

iv

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The purpose of this study stems primarily from the fact that the component dimensions of effective ski instructing have yet to be empirically investigated. Consequently, there appears to be a scarcity of standardized, objective, and valid criteria by which ski school directors can evaluate teaching performance in both pre-seasonal employee selection and post-seasonal evaluations. At present, evaluation procedures appear to be mainly subjective and idiosyncratic to a particular director and resort.

Many professions require a licensing or certification process intended to function as a screening mechanism to differentiate between the qualified and unqualified in a given occupation. Theoretically, a ski director could then utilize the criterion of certification vs. non-certification as a rudimentary measure of teaching effectiveness. Unfortunately, there are severe limitations to this criterion.

The first limitation is the fact that hiring policies at ski resorts are such that employment is not contingent on certification. That is, many uncertified instructors seek employment and are hired for seasonal work. Thus, the ski director is in need of some criteria other than certification by which to evaluate this particular pool of

applicants.

Secondly, like any professional certification procedure, it is not infallible and relatively ineffective teachers are to be found among the population of certified instructors. Consequently, merely being certified is no absolute guarantee of good teacher performance.

Finally and most important, the certification process itself is a non-standardized procedure. Certification is awarded on a regional basis and the criteria and methods by which to measure them vary from region to region. Moreover, there is no empirical evidence to show that the set of criteria utilized in any particular region truely discriminates between effective and ineffective teachers. This is not meant to degrade the current methods, for on an intuitive level the domains evaluated appear to be relevant and multidimensional. Empirical validation of the procedures are lacking, however, making the criterion of certification as the sole indicator of teacher effectiveness a questionable one.

Returning to the original issue of employee evaluation by ski directors, some objective (i.e. quantitative) data may currently enter into the decision process at some resorts. Their efficacy as valid predictors of teacher effectiveness has not been demonstrated, however. For example, it is not uncommon for records to be kept concerning special requests for certain ski instructors by former students. While this may indicate the instructor is well liked by his students, no valid link can be made apriori between this and his overall teaching effectiveness of skiing skills. Certainly given the multidimensional

nature of teaching in general, a limited criterion such as this
could not begin to account for a significant amount of the variance.
In fact, it is estimated that for any <u>single</u> variable contributing
to teacher effectiveness in general, a correlation larger than .4
with the overall criterion measure cannot be expected (Gage, 1978).

In view of the deficiencies in the area of instructor evaluation,
the primary purpose of this study is to develop a reliable and valid
measure of teacher effectiveness in ski instructing in the form of
a standardized rating scale. This instrument could be utilized by
ski school directors in evaluating both their certified and uncertified
employees. Moreover, the information obtained from this study would
undoubtedly be of value in refining the current certification process
both in terms of training and evaluation.

The second purpose of this study is to gather normative data on
the Personality Research Form (PRF) (Jackson, 1967) for ski instructors
as a specific occupational group. Some research of this nature has
been conducted with special educational and psychiatric populations.
However, no studies were located which sought to obtain normative
data on the PRF for occupational groups distinctly separate from the
college norms on which the scale was standardized.

Finally, the efficacy of PRF scales as predictors of teacher
effectiveness in ski instructors will also be investigated.

Three distinct bodies of literature will be reviewed in light
of these three goals. Since the central focus of this research is
measurement, the basic principles of psychological testing will be
reviewed. The research on teacher effectiveness will also be summarized,

and an overview of objective personality tests and research on special groups will conclude the review.

CHAPTER II

REVIEW OF THE LITERATURE

Principles of Psychological Testing

Information for this review was obtained from both recent and
classic texts in the area of statistics and psychological testing
(Anastasi, 1976; Brown, 1970; Edwards, 1973; Ghiselli, 1964;
Ghiselli & Brown, 1948; Jackson & Messick, 1967; Lanyon & Goodstein,
1970; Nunnally, 1967; and Snedecor & Cochran, 1967).

Reliability

The concept of reliability refers to the stability, consistency,
and repeatability of measurement. If an assessment instrument were
valid and 100% reliable, theoretically it would be assumed that any
variations between individuals on that measure would be due to "true"
differences between them on that particular trait, characteristic, or
ability. Likewise, any changes in an individual's score over time
would be solely reflective of an actual change in the level or strength
of what was being measured. Obviously, such a psychometrically perfect
instrument is beyond the capabilities of our discipline at this time.
Measures of reliability are therefore necessary in order to estimate the
accuracy of measurement for a test.

Reliability can thus be related to the concept of measurement
error, which can either be systematic or random. Systematic error re-
fers to correctable mistakes in test construction or administration.

Reliability measures reflect random error, which is what remains
after all the systematic biases have been identified and removed as
much as possible. A reliability coefficient is then an estimate of
the correlation between scores on a test and the corresponding
theoretical "true" score of a respondent.

There is disagreement among authorities concerning the philo-
sophical assumptions underlying the concept of reliability and the
type of error it presumably reflects (Ghiselli, 1964; Nunnally, 1967).
Since an appropriate index of reliability is conceptually linked to
its definition, a variety of measurement methods have consequently
evolved.

Test-retest reliability

Test-retest reliability refers to the correlation between scores
obtained from two separate administrations of the same test. The error
variance being measured thus corresponds to the random fluctuation of
performance over time. The greater this type of reliability, the more
test results can be generalized over time and presumably the more
stable the function being measured. For example, high test-retest
reliabilities are expected for valid tests of intelligence or person-
ality traits. This type of reliability measure is problematic in some
instances, however. For example, when practice effects or memory
components are likely for a given test, this method is inappropriate
since the coefficient would be artificially inflated.

Split-half reliability

Split-half reliability involves the correlation between two
comparable halves of a test and reflects the consistency of content

sampling. Again, it is not an appropriate measure for some tests, particularly if comparable halves are difficult to obtain. The more heterogeneous the domain being measured, the more difficult this task would presumably be.

The development of alternate forms and the correlation between them is another index of reliability which is computationally similar to split-half reliability. Like test-retest, it can be problematic when practice effects are likely, or when pragmatic considerations prevent the development of a parallel form.

Kuder-Richardson reliability coefficients

Kuder-Richardson reliability formulas (Kuder & Richardson, 1937) measure error variance due to both content sampling and content heterogeneity. The method takes into account both the standard deviation of the test, and its interitem consistency as reflected in the correlation between items and the total score. The heterogeneity or homogeneity of the domain being measured will affect the K-R coefficient. Heterogeneous criterion would necessarily be measured by a less homogeneous item pool and would thus result in lower interitem consistency.

Standard error of measurement

Standard error of measurement is another way to express a test's reliability. It is used exclusively for interpreting an individual's performance and estimates by how much his obtained score deviates from his true score. In this sense it is used independently of the reliability coefficients previously mentioned, although the statistic itself is derived from the reliability coefficient of the particular

test in question.

In general, reliability coefficients will be affected by three factors. The extent to which the characteristic in question is heterogeneous in nature and its resulting influence on reliability has already been mentioned. Secondly, the constitution of the standardization sample is important. The more heterogeneous the sample with respect to the trait being measured, the more likely the reliability coefficient is to be higher, since it is affected by the range of individual differences in the normative group.

Cronbach was the first to detect the significance of the third factor affecting reliability, which is specifically, response biases (Cronbach, 1942; 1946; 1950). Response biases or distortions can take the form of either a response style or a response set. A response style is generally considered to be independent of the test's content and reflects a disproportionate tendency to respond in a certain way. For example, an acquiescent response style refers to a tendency to respond "true" on true-false inventories.

Response set refers to a distortion of responses resulting from a person's desire to present himself in a particular way. This may be either conscious or unconscious. The well-known phenomenon of a socially desirable response set was first investigated by Edwards (1953) when he detected a tendency for college students to endorse socially desirable items on a personality test, rather than responding in a way which might have been more truely representative of themselves.

Issues relating to a test's reliability should ideally be confronted during the construction phase of the instrument. To the extent

that this is true (e.g. controlling for response distortions in the test's development), the more reliable the measure will generally be.

Validity

Validity refers to the extent to which a test measures what it is intended to. Evidence for different kinds of validity are generally emphasized according to the nature and purpose of the particular test in question. The three types of validity generally recognized are content validity, criterion-related validity or predictove validity, and construct validity.

Content validity

Content validity reflects the degree to which test items are representative of the domain being measured. The more definitive the domain, the more the inclusion of representative items is facilitated in test construction and the easier it is to evaluate the adequacy of the sample. It is thus particularly relevant for achievement and aptitude tests where the behaviors of concern are more easily specified. Although it is considered in aptitude and personality tests, the issue of content validity is secondary to the other types of validity mentioned above.

Content validity is commonly determined by the judgement of experts concerning the relationship between a test and the domain it is measuring. Problems are encountered when judges disagree and/or when the domain in question is not adequately defined. The lack of a standardized quantitative index of content validity is also problematic. This can be partially circumvented by obtaining judges' ratings on each item and using the extent of agreement between judges as a measure of content validity.

Criterion-related validity

Criterion or predictive validity is measured by a test's relationship to some criterion. Astin (1964) makes a distinction between the global concept of the criterion which he calls "conceptual criterion", and the operational device utilized to measure it. Criterion validity thus technically reflects how well test results can predict the performance on some independently determined criterion measure.

A test can be validated against a number of criterion measures usually chosen because of their relevance to the purpose for which a test is designed. The process of prediction is not restricted to criterion obtained sometime in the future, but may also involve some external criterion measured simultaneously. Thus, predictive and concurrent validity are conceptually the same.

Criterion measures should ideally be relevant, reliable, and free from bias, and should also be selected with consideration to pragmatic concerns (Brown, 1970). Although multiple criteria may seem optimal, it complicates methodology and may not always be desirable in a single validation study (Ghiselli, 1956).

As mentioned above, appropriate criterion measures differ depending on the type of test. Academic achievement is a common criterion for intelligence tests. Job performance has been used to validate aptitude and personality tests. Scores from existing tests proved to be psychometrically sound are often used as criteria for new tests measuring the same domain.

Of particular relevance to this review is the use of ratings by judges (teachers, supervisors, etc) as a criterion measure. Ratings

can be obtained for quite subjective, globally defined characteristics
as well as ones which are more behaviorally precise. Despite the pro-
blems inherent in this procedure, it is nonetheless the most appropriate
criterion to use in some cases. Since it is the criterion of choice
for this study, it warrants a more thorough examination.

Ratings are expressions of opinions and are subject to the errors
characteristic of human judgement. Nunnally (1967) prefers to distinguish
between "judgements" and "sentiments" where ratings are concerned.
"Judgements" he considers to be responses indicating the correctness or
incorrectness of some type of veridical comparison. "Sentiments" refer
to all responses involving personal reactions, preferences, attitudes,
etc. Most rating systems in criterion validation will involve the
latter category, though the terms will be used interchangeably for
simplicity's sake in this review.

In order to ensure their independence, it is necessary to obtain
ratings on the relevant criterion from several judges. Successive
ratings by the same judge at a later date would not be sufficient since
the second rating would be influenced by the first and would thus be
nonindependent. Problems with using multiple judges, however, necessarily
involve the issue of disagreement. Disagreements may arise not only
because of a true difference of opinion concerning the characteristic
in question, but also may occur because of varied exposure of the judge
to the individual being rated. Disagreements may also result from
response styles involving errors of leniency or central tendency (Ghi-
selli & Brown, 1948). Moreover, Anastasi (1976) addresses the issue
of criteria contamination. This occurs when a judge's criterion rating
is influenced by knowledge of the ratee's scores on the predictor

measure.

These problems can be minimized by 1) controlling for criteria contamination, 2) if possible, providing a precise definition of the criterion dimension in question, 3) selection of judges who are most familiar with the individual's being rated, and 4) allowing sufficient time for the actual rating procedure.

Requiring judges to rank individuals or items on the relevant criteria can provide more information than merely rating them. A paired comparisons technique can facilitate the process, but becomes cumbersome as numbers increase. In general, ranking is not recommended for over twenty to thirty individuals or items, due to the difficulty of judgement (Ghiselli & Brown, 1948). Ranking methods share the same problems as rating methods, which can be minimized by the procedures mentioned above.

A number of criterion measures have been discussed. Once the appropriate one is selected, the nature of the criterion group is the next issue of concern. As in any sampling procedure, size and representation are crucial factors. The larger the sample, the more likely significance will be obtained. Moreover, the criterion group should ideally be representative of the population for which the test is designed. Cross-validation should be pursued when possible to ensure the generalizability of the test's predictive powers.

Criterion validity can be measured or evaluated either through using a validity coefficient, or by a method of group separation. A validity coefficient represents how accurately the criterion can be predicted from a test score. The coefficient will be underestimated to the extent that 1) individual differences are restricted on either

the predictor or criterion measures, or 2) the predictor and criterion measure are non-linearly related (Brown, 1970).

The validity coefficient is advantageous because it concisely summarizes the relationship between the predictor and criterion, while also allowing for the prediction of criterion performance for an individual through the utilization of regression. It is also a common method of evaluating predictive validity and thus allows for a comparability between studies. Disadvantages include problems encountered when there is a non-linear predictor-criterion relationship.

Criterion validation through contrasted groups is also common. If predictor scores can differentiate between groups representing the extremes of the distribution of the dimension of interest, this provides evidence of criterion validity.

Construct Validity

The issue of construct validity was brought to the attention of the psychological community by the three classic articles of Cronbach and Meehl (1955), Loevinger (1957), and Campbell and Fiske (1959). Construct validity is relevant when a test attempts to measure an abstract domain for which there is no single operational definition or precise criterion. Nunnally (1967) points out that the larger the domain of observables related to a construct, the more difficult the component variables of the construct are to define. Moreover, the more abstract the concept in this sense, the more difficult the validation process becomes. Construct validation of a test thus involves the accumulation of data from a variety of validation studies, rather than utilizing a standardization method or statistical procedure. The degree of validity of a test is thus constantly being refined and

re-evaluated as discrepant or supportive evidence emerges.

The definition or meaning of a construct is nomothetic in nature and is thus derived from a theory and the laws and propositions pertaining to that theory (Cronbach & Meehl, 1955). The theory's postulates are the vehicles by which observable data can be related to and provide evidence of the construct of interest. This inferential process is then a continual one as empirical evidence of construct validity is gathered.

Loevinger (1957) expanded this concept by arguing that construct validity can only be established by converging lines of evidence relating to three mutually exclusive components. The substantive component refers to the extent to which the content of the items included in a test relate theoretically and empirically to the broadly defined domain which the test proports to measure. The structural aspect of construct validity is an analysis of the internal structure of a pool of items and incorporates concepts such as homogeneity. The external component includes predictive and concurrent validity by analyzing a test's relationship to non-test behavior. Other subdivisions of this aspect include a test's factorial pattern and the relationship of test scores to other tests. Loevinger argues that evidence for construct validity must be broken down to incorporate these three aspects, which she considers to be mandatory components of the concept.

Brown (1970) distinguishes between inter- and intra-test methods through which validity information is obtained. Content validity would be an intra-test method since it refers solely to the internal structure of an instrument. It helps to more accurately define the relevant domain of a construct, but does not provide evidence that the test actually

measures the construct. It is thus a necessary but not sufficient condition for construct validity.

Inter-test methods are numerous and variable and involve determining if two tests measure the same construct. Methods of criterion-validation are of this type and have already been discussed at length. Other inter-test methods involve convergent and discriminant validation, factor analysis, and experimental manipulation.

The multitrait-multimethod matrix introduced by Campbell and Fiske (1959) is an efficient way to demonstrate convergent and discriminant validity. By performing this analysis on a given number of tests and characteristics, one can evaluate the amount of error variance which is attributable to the trait itself, and how much is a result of the particular method used to measure it. The correlation between different methods measuring the same trait is evidence of convergent validity. The correlations between different traits measured by the same method provides evidence for discriminant validity. It is hoped that through the latter process, evidence will accrue for the independence of a trait from other traits to which it is theoretically unrelated.

Factor analysis is a statistical procedure used to condense a number of variables into a smaller number of definable and distinct categories or factors. When factor analysis is performed on several tests, the issue of concern is how many factors or constructs are needed to account for the intercorrelations among test scores. To the extent that certain scores "load" on a factor, they share a common variance and can be considered to measure the same construct. What factors or constructs determines the scores obtained on a particular test is also analyzed. The factorial validity of a test then refers to the correla-

tion of a test with each factor identified as determining its scores (Anastasi, 1976).

Finally, test-retest reliability would be an example of construct validation through experimental manipulation with time serving as the manipulated variable. To the extent that test scores are unchanged, the stablity of the construct being measured is revealed. Like content validity, this is evidence for construct validation, but is not sufficient in and of itself.

In summary, the issue of validity is complex and multidimensional. The measurement of validity is not standardized, but is specific to the individual test, given what it measures and the purpose for which it is intended. The assessment of validity is thus a subjective judgement involving the accumulation of supportive and discriminant data through which that judgement is continually refined.

Teacher Effectiveness

General problems in the literature

The problem of precisely identifying the measureable character-
istics of good teaching is a long standing one which has inspired
thousands of investigations spanning several decades. The scope of
these studies vary as much as the methodology utilized, and range
from questionably small subject pools to massive studies involving
hundreds of separate research projects and thousands of teachers
(Ryan, 1971). Despite this voluminous research, collective results
would hardly be considered definitive. In fact, much of the research
on teaching has been summarized as a fruitless search for consistent
relationships between teacher variables and effectiveness criteria
(Doyle, 1978; Shavelson & Dempsey, 1976).

A seemingly infinite list of problems can be identified as the
source of this unfortunate state of affairs, ranging from a lack of
reliable definitions of composite traits of effective teachers, to a
lack of adequate, concrete, objective, and standardized criteria for
teaching ability. Heath & Nielson (1974) identify the variability of
methodological flaws between studies as the salient culprit, and pay
specific attention to randomization problems and frequent violations
of the assumptions of various statistical techniques (e.g. normality,
homogeniety, linearity).

However, some researchers contend that there is a general lack
of relationship between the quality of evaluation studies and the
nature of the outcome, i.e., the same conclusions are generally reached
and may still be valid (Stickell, 1974; Yin, Bingham & Heald, 1976).
N.L. Gage (1978) is the major proponent of this position. He points

out several major errors committed by proponents of the view that
the yield of teaching research has been almost completely equivocal,
non-significant, and inconsistent.

First, he specifically selects Dunking and Biddle (1974) as
reviewers illustrative of committing Type II errors in their analysis
of the teaching literature. A summary of Gage's criticisms should
be prefaced by noting that although teaching is multidimensional,
many studies have investigated the relationship of only one element
of teaching to pupil achievement. Many of these studies have yielded
nonsignificant results and are used by critical reviewers as evidence
of the ambiguous and inconsistent relationship between these traits
and teacher effectiveness.

Gage argues, however, that because of the multidimensional nature
of teaching, a very low and nonsignificant correlation (e.g. .1 - .4)
of any single variable to teacher effectiveness is to be expected.
Moreover, even for research involving multiple variables, sample sizes
are typically small, which raises the critical value needed for the
correlation coefficient to reach significance. Investigations of
single dimensions of teaching behavior and multidimensional studies
on small samples would almost always be nonsignificant because of
these considerations. Consequently, reviews based solely on the
presence or absence of statistical significance will necessarily be
bleak and contradictory.

This brings up the second issue contributing to the lack of
consensus among reviewers of the teaching literature. Specifically,
there are numerous methods of synthesizing the available research
(e.g. statistical significance, sole examination of the consistency

of direction, etc.), which necessarily produce different conclusions. Gage proposes a method of evaluation involving testing of significance of combined results of many studies utilizing a method proposed by Jones and Fiske (1953). A conversion technique relates the results of single studies into chi square values, which are summed across studies to determine their joint significance. Thus, a cluster of studies examining the same process variable or teaching behavior can be tested with a technique employing greater statistical power. When applying this technique to several variables found to be non-significant by several reviewers, certain behaviors were in fact discovered to be significant contributers to teacher effectiveness.

A third and obvious issue contributing to the incohesive nature of this area is the phenomenal variety of populations on which research is conducted. Subject matters examined range from reading to science; schools from private to public and large to small; pupils range from pre-school to graduate school; and teachers from non-certified and inexperienced to tenured university professors.

Despite this massive literature, no systematic studies could be located concerning the instruction of individual (as opposed to team) sports. Suffice to say that the only direct bearing this literature has on the proposed study is 1) the establishment of teaching as a multidimensional process, the component variables of which have not been clearly defined, and 2) the widespread usage of rating scales subcategorized into various domains as the criterion measure of teaching ability.

There is one study, however, which bears particular relevance to the efficacy of personality characteristics as predictors of teacher

effectiveness (Murray, 1975). Although several similar studies have failed to demonstrate relationships between these two variables (at least in college teaching) (Bendig, 1955; Sorey, 1968), Murray successfully identified characteristics significantly correlated with ratings of teacher effectiveness. Moreover, particular attention is paid to this study since characteristics were assessed via peer ratings directly derived from the PRF.

Peer ratings were obtained by 45 full time faculty members on 20 characteristics directly paralleling the PRF scales. Mean trait ratings on 36 instructors of psychology were computed. The teacher effectiveness of these instructors was measured by a student questionnaire composed of items organized into eight areas. The categories and their split-half reliability coefficients are as follows: Communication (.96), Interest (.93), Rapport (.88), Grading (.75), Textbook (.72), Impact (.77), Difficulty (.83), and General Evaluation (.96). A rating scale of 1 to 9 was utilized and an Overall Teacher Rating (OTR) (.95) was derived from these categories. A principal-axis, varimax rotation factor analysis of mean items yielded factors which correspond closely to the logically defined questionnaire categories, except that Communication, Interest and General Evaluation items loaded on a single large factor.

Eleven of the twenty personality traits correlated significantly with Overall Teacher Rating. A stepwise multiple regression analysis yielded personality traits as predictor variables with OTR serving as the criterion. A linear combination of four personality traits (Leadership, Objectivity, Extroversion, and Anxiety) accounted for

67% of the variance on OTR, R (5, 31) = 82, p. < .01).

The study must be evaluated in light of the following issues. First, there is support of the validity of using student ratings as evidence of teacher effectiveness (Costin, 1978; Costin, Greenough, & Menges, 1971). There have been a few studies yielding high negative correlations between these variables (Bendig, 1953; Rodin & Rodin, 1972), but these have been largely criticized for methodological problems (Frey, 1973; Kulik & McKeachie, 1975).

Secondly, significant correlations have been found between PRF scores and peer ratings using definitions of the traits which PRF scales intend to measure (Jackson, 1967; Jackson & Guthrie, 1967; Kusyszn, 1968). Moreover, a multimethod factor analysis (Jackson, 1966) shows that peer ratings loaded on the appropriate factors yielded by the analysis (Jackson & Guthrie, 1967). It could be safely hypothesized, then, that similar results would have been obtained, had the PRF been used as the method of personality assessment instead of peer ratings.

The data imply that personality traits are useful predictors of teacher effectiveness in college faculty. The usefulness of these findings for ski instructing of course depends on their generalizability across teacher populations. Although this is an empirical question, it can be partially examined here.

Given the definitions provided by Murray, four of the eight categories on his rating scale would appear, apriori, to be generalizable to the teaching of any subject or sport. Specifically, Communication, Interest, General Evaluation (which all load on the same factor), and Rapport. Personality traits significantly correlating with all four of

these categories are: extroversion, liberalism, leadership, exhibition,
anxiety (negatively), personal warmth, lightheartedness, objectivity,
and defensiveness (negatively). The likelihood of finding significant
correlations between certain PRF scales and teacher effectiveness in
ski instructing can then be hypothesized on 1) the basis of these
findings given the general methodological soundness of Murray's study,
and 2) the significant correlation between peer ratings and PRF scores.

Personality Tests

The term "personality test" is generally applicable to assessment instruments intended to measure characteristics such as emotional adjustment, interpersonal relations, intrapersonal dynamics, motivations, attitudes, and interests. They can be designed to reveal pathological tendencies in psychiatrically disturbed or deviant populations. Conversely, many are intended for use in normal populations and tap traits and characteristics commonly associated with "normal" functioning. Assessment devices may be performance or situational in nature, or they may utilize non-projective or projective techniques. This review will be concerned with non-projective instruments intended to measure normal personality functioning.

Objective personality tests employ norms which represent the test performance of the subjects constituting the standardization sample. Ideally the sample obtained should be a representative cross section of the population for which the test is designed. Unfortunately, there are obvious pragmatic obstacles which render this task nearly impossible except for very restrictive tests designed for use in highly specific populations. Considerations of economy and efficiency have thus made high school and college students the most commonly employed normative populations. The crucial issue is: How well do non-college samples conform to college normed summary statistics? To the extend that they do not, the instrument's scope and applicability are considerably restricted unless efforts are made to gather normative data on other specific groups of interest.

Suffice to say that personality tests are subject to the same

psychometric issues of reliability and validity previously discussed.

While these properties will be briefly addressed for several tests,

the central concern of this review is with deviations of special groups

from the normative sample on which the test was standardized. This

discussion will thus be further limited to several of the most commonly

employed inventories on which research has also been conducted concern-

ing characteristics of special occupational groups within a normal pop-

upation.

California Personality Inventory

Reviews of the California Personality Inventory (CPI) range from

commendation to condemnation and reveal the controversial nature of the

test. While favorably evaluated by Kelly (1965), its utility is deemed

questionable by such respected psychometric authorities as Lee Cronbach

(1959). Moreover, a negative judgement by Thorndike (1959) is accented

by what can best be described as a scathing review by Walsh (1972).

Major shortcomings include the extreme group criterion-oriented approach

to scale construction and the excessive number of scales which are re-

dundant and highly intercorrelated. Although it seemingly borders on

ill-repute, research utilizing the CPI has been voluminous. Additionally,

its sizable and varied norm groups are illustrative of the need for norm-

ative data on special groups and render it worthy of mention in this re-

view.

The CPI is one of the few inventories standardized on a large (6,200

males, 7,150 females) non-college population (Gough, 1975). It is not

claimed that this is a true random sample of the general population, how-

ever. Unfortunately, the manual fails to provide specific demographic

information other than it included a "wide range of ages, socio-economic groups and geographic areas".

A decided virtue of the CPI is that norms are provided for thirty additional educational, occupational, and "miscellaneous" samples. Of these, eleven of the male samples and seven of the female samples would be considered special occupational groups such as physicians, scientists, machine operators, writers, office workers, etc. Examination of these norms reveals a substantial number of significant mean differences between various groups and the population on which the test was standardized. It should be noted that the utility of these special norms may be limited due to marginal test-retest reliabilities ranging from .55 - .75, for a one year period. The point, however, is the significant differences between these occupational groups and the standardization sample which is at least somewhat representative of a general population. Considering that many inventories are standardized on restricted college and high school populations, the need for normative data across tests for special groups becomes increasingly evident.

Edwards Personal Preference Schedule

The Edwards Personal Preference Schedule (EPPS) is also relevant to this review since it was the instrument utilized in the only study located on personality characteristics of ski instructors (Agocs & Suvak, 1977). The findings of this study must be interpreted in light of the limitations of the inventory. The psychometric properties of the EPPS thus warrant brief examination.

Unlike the CPI, the development of the EPPS was guided by the theoretical foundations of a personality theory. Specifically, it was

one of the first inventories designed to assess the strength of the manifest need system proposed by Murray (1938). It has been widely used in research for decades and has served as a catalyst for research and psychometric issues. Unfortunately, it has also been widely misused for reasons which will soon become clear.

The unique feature of the EPPS is its paired-choice format which yields ipsative scores. The strength of each need is not measured in absolute terms, but is evaluated relative to the strength of the individual's other needs. Two individuals with identical scores on the EPPS may differ greatly in terms of the absolute strength of their needs. The appropriate reference in ipsative scoring is thus the individual, not the normative sample. The lack of absolute measures obviously makes group comparisons using the EPPS problematic and largely inappropriate.

Construct validity depends to a large extent on the procedures followed in the development and selection of items for each of the scales. Validity issues were not substantially addressed in the stages of scale construction, however (McKee, 1972), thus increasing the importance of evidence of validation accumulated after the instrument's publication. Unfortunately, such studies are few in number and are confounded by the fact that many researchers did not account for the ipsative nature of the scoring system (McKee, 1972; Anastasi, 1976). This oversight is an extremely common error and has contributed to the misuse and misinterpretation of the EPPS in other research besides validity studies.

Not only is the external evidence of the inventory's construct validity meager, McKee points out other weaknesses, one of which is the non-independence of the scales since each choice affects two scales.

Secondly, despite the fact that items were paired for equal social desirability, the EPPS still does not effectively control for a socially desirable response set which again leaves open the question of validity (Heilbrun & Goodstein, 1961; McKee, 1972; Rorer, 1965).

The scales do represent an important cross section of normal personality dynamics, however. Other positive features include acceptable scale reliabilities, a stable norm group, and relatively low interscale correlations (Heilbrun, 1972). Note, however, that ipsative scores cannot be properly analyzed by the usual correlational procedures since the mean intercorrelation of individual scales using this system tend to be negative (Hicks, 1970).

In summary, most reviewers have not found the EPPS to be a particularly useful research instrument because of the weaknesses cited (McKee, 1972; Radcliffe, 1965; Stricker, 1965).

With these issues in mind, findings from a study of ski instructors scores on the EPPS can be discussed more appropriately and interpreted with caution. The sample included 53 ski instructors from the northern rocky mountain area (15 females and 38 males). They represented all levels of certification and ranged from one to fifteen years of experience in ski instructing.

Within group comparisons revealed no significant differences between experienced vs. inexperienced instructors, and non-certified vs. associate vs. certified instructors. Instructors were compared with both college norms and norms for the general population. All differences were reported at the .05 level, though it should be noted that the exact method of statistical analysis is unclear.

The _relative_ _strength_ (my emphasis) of the need for "change" in male ski instructors was significantly higher than college males, while "dominance" was lower. Compared to the general male population, "intraception", "change", "heterosexuality" were stronger in ski instructors, while "deference", "order", "abasement", and "endurance" were lower.

Female instructors scored relatively higher than college women on "autonomy" and "endurance", but lower in "affiliation". Compared to the general female population, women instructors scored relatively higher on "autonomy", "dominance", "change", "endurance", and "heterosexuality". Scores were relatively lower in "deference", "order","affiliation", "abasement", and "nurturance".

The shortcomings of this study are ample. First, the inappropriateness of making group comparisons with the EPPS have been discussed at length, and for this reason alone the conclusions may be considered questionable because none of the significant differences can be interpreted in absolute terms due to the ipsative nature of the scoring system. Secondly, the sample is small and too heterogeneous to warrant generalization to the general population of ski instructors. For the purposes of this review, the sample cannot be treated as a specific occupational group since there is no mention that subjects viewed ski instructing as their primary occupation. In fact, experience suggests that it is more likely that subjects represented persons in a variety of occupations who also happen to be part time ski instructors. This would most certainly be true of the uncertified instructors included in the sample. Finally, the total sample size is too small to make reasonable within group comparisons. Likewise, a sample size of thirteen is insufficient for between group

comparisons for female instructors.

Personality Research Form

The Personality Research Form (PRF) is the assessment instrument selected for use in this study. It has generally been reviewed as being highly psychometrically sound (Anastasi, 1972, 1976; Kelly, 1972; Wiggins, 1972). A review of its development and psychometric properties is relevant, as is a summary of the research on special groups utilizing the PRF.

The development of the PRF relied heavily on theoretically-oriented definitions of personality characteristics basically paralleling Murray's manifest need system. The behaviorally-oriented mutually exclusive definitions of 20 traits were also intended to represent bipolar personality dimensions. These definitions served as guidelines to the selection of item pools on the basis of proposed conceptual links to the constructs being measured.

The item pools were administered to college students and 20 items were selected for each scale on the basis of high biserial correlations with the total scale score, and low correlations with scores on the Desirability Scale and other remaining scales. In keeping with their bipolar nature, half the items on each scale were written in terms of one end of the pole, and the other half in terms of the opposite end of the dimension being measured.

The PRF has been shown to adequately control for acquiescence and socially desirable response biases (Jackson & Lay, 1967; Trott & Jackson, 1967). Two validity scales (Infrequency and Desirability) also provide indexes of nonpurposeful responding and an unusually conforming

response set.

Reliability

Reliability considerations were built into the PRF at the con-
struction stage. Controlling for response biases aided in producing
reliable scores. Odd-even reliability coefficients for the 20 scales
range from .48 - .90. K-R coefficients range from .54 - 86 (Jackson,
1967). Test-retest reliability using form AA for a one week period
ranged from .69 - .90 for college students (Anastasi, 1972).

Validity

Convergent and discriminant validity of the PRF scales have been
investigated using the multitrait-multimethod framework proposed by
Campbell and Fiske. PRF scores were correlated with pooled peer ratings
and self ratings, yielding a median correlation of .52 and .56 respective-
ly (Jackson & Guthrie, 1967). A multimethod factor analysis of the 20
traits and three methods yielded factors corresponding to the trait
scales, with appropriate loadings across all three methods. Jackson
cites this as evidence for both convergent and discriminant validity.

This anlaysis did not conform to the conventional Campbell and Fiske
method, however, as noted by Wesler and Loevinger (1972). By not computing
the intercorrelations of self ratings among themselves and peer ratings
among themselves, the major purpose of the multitrait-multimethod matrix
was circumvented, i.e., the relative contribution of trait variance and
method varience was not analyzed.

PRF and special groups

Although Jackson warns test users to exercise caution in applying
PRF norms to non-college populations, research intended to privide alter-

native norms has been surprisingly scarce.

Significant mean differences for certain PRF scales have been reported for several special groups, however. Within educational groups, differences were found between PRF norms and student nurses (Hoffman, 1970a), and highly rejected and highly accepted college students (Adinolfi, 1970). Differences from normative data have also been reported for several groups within a psychiatric population, specifically male tranvestites (Bentler & Prince), alcoholics (Hoffman, 1970b), and physically disabled persons (Winegardner, 1978).

Thus far there does not appear to be research available on personality characteristics of specific occuational groups using the PRF. However, Seiss and Jackson (1970) have correlated PRF scores with vocational interests as measured by the Strong Vocational Interest Blank (SVIB). A multimethod factor analysis procedure yielded seven interpretable factors determined by measures from both the SVIB and PRF. Moreover, for each factor, loadings emerged for more than one scale in each inventory. Factors were associated with an orientation towards achievement related goals, human relations management, impulse expression, practical goals, managerial control, aesthetic-intellectual goals, and social contact.

Unfortunately subjects were limited to college males, however some basis for the interpretation of vocational interests within the framework of personality theory could be speculated for the general population. To the extent that vocational interests are predictive of entry into a chosen occupation, particular occupational groups could be theoretically hypothesized to differ from PRF norms on the scale which loaded on the

same factor as the interest scale for that occupation.

For purposes of the proposed study, however, the Seiss and Jackson research has no overwhelming significance. Differences between occupational groups and available norms on other personality inventories have been cited and similar findings utilizing the PRF would be expected.

The importance of examining personality characteristics in ski instructors has thus far been discussed with respect to this group as 1) teachers, and 2) a specific occupational group. They could also be considered an athletic group (specifically, a subgroup of skiers), which would warrant a brief review of sports psychology and personality characterisitcs of athletes in general. However, the purpose of this study does not bear any particular significance to the issue of ski instructors as athletes. Moreover, the population under investigation in sports psychology is typically composed of competitive athletes, a definition not applicable to most ski instructors. Thus, in light of the lack of convergence between the thrust of this tudy and the nature of the field of sports psychology, that literature will not be reviewed.

CHAPTER III

PURPOSE OF RESEARCH AND HYPOTHESES

Purpose and Significance of Study

The purpose of this study was threefold. The primary goal was
the development of a reliable and valid assessment instrument of teacher
effectiveness in ski instructors. The instrument would most appropriately
be used by a person in a supervisory capacity who has had an opportunity
to observe the individual in question during his teaching performance.
It should have utility for both the pre-seasonal employee selection pro-
cess as well as for mid- or post-seasonal evaluation. The latter use
would facilitate employer/employee communication concerning job performance
in identifying both strong points, as well as problem areas which deserve
more attention in a particular instructor. Finally, information obtained
in this study may be useful in refining the current certification proce-
dures.

Secondly, the possibility of using certain PRF scales as predictors
of teacher effectiveness was investigated. Studies previously cited not
only imply that scales can be treated as distinct, but they are also
highly correlated with others' ratings of behavioral descriptions of
traits measured by the PRF. Thus, if specific predictors of teacher ef-
fectiveness were identified, this information could also serve as guide-
lines in the employee selection process without having to administer the
relevant PRF scales themselves.

Hypotheses

Given the purpose of this study, the following research hypotheses were generated:

1) Reliable scales or subscales would be obtained for the proposed assessment instrument based on the items in the refined item pool.

2) From the data gathered at the resorts, it would be possible to produce a scale composed of relatively distinct domains or clusters of items related to instructor characteristics and abilities which are significantly related to the effective teaching of skiing.

3) Evidence of validity would be established through an internal cross validation between two randomly divided halves of the resorts sampled.

4) Evidence of criterion-related validity through group separation would be established.

5) There would be significant correlations between certain PRF scales and the criterion measure for teacher effectiveness.

6) Significant differences on some PRF scales would result between the PRF normative sample and the sample of ski instructors.

CHAPTER IV

METHODS

The methods and procedures of this study involved 1) the develop-
ment of the assessment instrument and the selection of materials and
subjects utilized in this procedure, 2) the selection of subjects and
procedures related to the assessment of personality characteristics of
ski instructors.

Development of Teacher Effectiveness Scale

Item Pool

A preliminary item pool for the teacher effectiveness scale was
generated from the experimenter's personal interviews with five examiners
certified with the Northern Rocky Mountain region of the Professional Ski
Instructors of America (PSIA). PSIA examiners are certified ski instruc-
tors who have received additional training in evaluating ski instructors
for purposes of official certification.

During the interview the experimenter asked the examiner to verbally
list the characteristics and abilities he considered necessary or desirable
for ski instructors to be effective teachers. The experimenter then asked
questions intended to produce more behaviorally specific descriptions of
the traits and abilities originally mentioned. If paraphrasing in the
recording process was necessary, care was taken to read the paraphrased
item back to the examiner to ensure that the content was still accurate.

After the interviewing process, efforts were made to eliminate duplicate items contributed by different examiners and a final list of 50 items was compiled. This pool was then divided into several logically derived subdivisions including Attitude, Personality and Interpersonal Skills, Communication Skills, Error Correction, Class Handling, Skiing Skill and Knowledge, and Personal Appeal (see Appendix A).

The lists were sent to all ski directors of PSIA member ski schools in the Pacific Northwest, Rocky Mountain, Northern Rocky Mountain, and Intermountain regions of PSIA (n = 79). Directors were asked to rate each item in terms of its importance in contributing to the effective teaching of skiing. The rating system utilized a Likert scale ranging from one to seven, with the extremes and midpoint respectively labeled "not at all important", "extremely important", and "moderately important". Directors were also asked to list and similarly rate any characteristics or abilities not included which they also considered to be important. A self addressed business reply envelope was enclosed to facilitate a large response.

Forty-eight questionnaires were returned and mean ratings for each item were caluculated to establish the degree of agreement with regard to which items were viewed as more or less important across the ski directors sampled. All 50 items received a mean rating greater than 4.3, indicating that the directors considered all of them to be at least moderately important in contributing to teacher effectiveness. Several respondants also suggested that "dependability" be added to the list of items. The item pool utilized in the validation procedure thus consisted of a total of 51 items.

Subjects

The criterion groups used to validate the teacher effectiveness scale consisted of instructors considered to be representative of "effective", mediocre", and "ineffective" teachers at each of eight PSIA member ski schools. The selection procedure entailed a supervisor at each ski school identifying three instructors whom he considered to be his best teachers, three average teachers, and three of the most ineffective teachers at his resort. Both certified and uncertified instructors were eligible for selection, though supervisors were asked to choose individuals who had been employed by the resort for at least a year since familiarity with the relevant characteristics included on the scale was important. The selection procedure yielded a total sample of 72 instructors, with 24 in each group. The mean number of years that these instructors had been employed at their respective resorts was 5.5 with s = 72, indicating that persons in supervisory capacities were probably quite familiar with their teaching skills.

The eight resorts selected were located in Montana, Colorado, Utah, and Idaho. One-half of the resorts were large and nationally known "destination resorts" including Sun Valley, Aspen, Vail, and Steamboat Springs, and one-half were small resorts serving more local clientele (Solitude, Parkwest, Copper Mountain, and Alta).

Validation Procedure

After the supervisor selected the nine individuals from his resort, he was asked to rank order them from one to nine on the basis of their overall teaching effectiveness. If more than one supervisor was available for consultation, they were given a randomized list of the nine instructors identified by the first supervisor and also asked to rank order them on

their global teacher effectiveness. The number of supervisors involved in this procedure ranged from one to four across the eight resorts.

A randomized list of the nine instructors was then presented to the ski school director who was asked to rank order them on each of the 51 items from the item pool. The items were administered verbally by the experimenter one at a time. The director then wrote down the names of the instructors in the appropriate order, or communicated this verbally to the experimenter, depending on which method he found to be easier. To avoid arbitrary rankings on items for which he considered instructors to be indistinguishable, he was instructed to give them equal rankings. Upon completion of this process the director was asked to rank order the nine instructors on the basis of their overall teaching ability. This was done to determine the degree of agreement between persons in supervisory capacities, and the magnitude of possible criterion contamination.

Instructor Hiring Scale

In the course of this investigation it became evident that a similar procedure to that described above could be conducted using individuals who had recently been hired as instructors. Specifically, Sun Valley and Parkwest had just completed their preseason instructor clinics, a process whereby job applicants participate in a four to five day on the snow clinic and are selected for employment on the basis of their performance.

The experimenter selected 33 items from the original item pool which pertained to abilities that could be reasonably assessed by clinic leaders after a few days exposure to the applicant's skiing and teaching ability (see Appendix B). Three clinic leaders from each resort were asked to select six individuals from the respective groups which they led during the clinic. Three of these individuals were applicants which were hired

upon completing the clinic, and three were applicants which were not hired. This selection procedure yielded a total sample of 36 instructors, 18 of whom had been hired following the preseason clinic.

The clinic leaders were asked to rank order the six individuals they had selected on each of the 33 items from the original item pool. Upon completion of this process they assigned global ranks from one to six on the basis of the individual's overall teaching effectiveness. A second opinion on the overall rankings was not obtained since there was no one else who would have had sufficient exposure to the individuals in each group to reliably make such a judgement.

Personality Characteristics of Ski Instructors

Subjects

One hundred and twenty-three instructors were solicited on a volunteer basis from the eight ski resorts previously mentioned, plus one additional resort (Brighton, Utah). The Infrequency Scale score on the Personality Research Form exceeded two standard deviations above the mean for five subjects, and their protocols were thus excluded from the analysis. Of the 118 remaining instructors, 28 had also been rated on the teacher effectiveness items by their ski school director.

Testing Materials

Form AA of the Personality Research Form was used as the personality assessment instrument. Scale descriptions are included in Appendix C. An information sheet (Appendix D) was also provided for each subject and obtained the following: name, age, years as instructor, years employed at this resort, current level of certification, part time or full time employment, primary on-season occupation, primary off-season occupation, and educational level. The subject's name was necessary in order to

identify those persons who were included in the validation procedure for the teacher effectiveness scale.

Procedure

A group meeting was arranged at each resort and the investigation was presented as a research project intended to study personality characteristics of ski instructors. Standardized PRF instructions were given, test booklets and answer sheets with the information sheet attached were distributed and collected at the end of the testing session.

CHAPTER V

RESULTS

Teacher Effectiveness Scale

The teacher effectiveness scale was constructed in the usual
manner for homogeneous tests (Nunnally, 1967). Before this procedure
is described, however, it is necessary to note that the criterion against
which the scale was validated was shown to be a strikingly reliable one.
Pearson product moment correlations were computed between the super-
visor's overall rankings (ranging from one to nine), and the director's
rankings of the same nine individuals. In instances where rankings were
obtained from more than one supervisor, the mean supervisor rank was
computed for each subject and correlated with the director's rank. The
final analysis yielded a correlation of .94 between supervisor and direc-
tor rankings. Thus, there was a high level of agreement between persons
in supervisory capacities on the criterion measure of overall teacher
effectiveness.

Total scores on the teacher effectiveness scale were computed for
each subject by adding the ranks assigned to them by their directors
across all 51 items. Overall director rank correlated .87 with total
score, while the mean supervisor rankings correlated .76 with total
score. The high correlation between the supervisors' ranks and total
score is perhaps the most convincing indicator of the validity of the
criterion and relative freedom from criterion contamination. These two

variables were independent of each other since the director ranked the subjects on each of the 51 items without knowledge of the supervisors' overall rankings of them. The intercorrelations among the variables just discussed and related variables are shown in the matrix on Table 1.

Evidence for criterion related validity through group separation was demonstrated by an analysis of variance between the director's classifications of teacher effectiveness (effective, medium, and ineffective), with toal score on the teacher effectiveness scale serving as the dependent variable. The analysis yielded a significant difference between groups (F (2, 213) = 87.5, $p < .001$) as shown on the summary table provided on Table 2. Post-hoc analysis by means of the Newman-Keuls test showed significant differences at the .001 level for every pair of means ($\overline{X}_1$ = 375, $\overline{X}_2$ = 593, $\overline{X}_3$ = 807).

An analysis including all 51 items from the original item pool yielded a coefficient alpha of .99. The average item correlation with total score was .70, with correlations ranging from .19 - .86. The distribution was highly skewed, with only two items showing correlations less than .40.

The 15 items showing the highest correlations with total score were then selected for further anlaysis. A scale composed of these 15 items yielded a coefficient alpha of .97 and a mean item correlation of .82 with total score. Correlation coefficients ranged from .80 to .86. An internal cross validation of these results was conducted by dividing the 72 subjects into two equal random samples (n = 36). The above procedure was duplicated for each sample and yielded results similar to those described above. For purposes of clarity the results are summarized on Table 3.

TABLE 1

CORRELATIONS WITH TEACHER EFFECTIVENESS SCORE
AND DIRECTOR RANKS WITH OTHER VARIABLES

	Total Score	Overall Director Rank	Mean Supervisor Rank	Years at Resort	Level of Certification
Total Score	1.00	.87 ****	.76 ****	.08	.60 ****
Overall Director Rank		1.00	.94 ****	.10	.55 ****
Mean Supervisor Rank			1.00	.11	.63 ****
Years at Resort				1.00	.38 **
Level of Certification					1.00

 ** $p < .01$

 *** $p < .001$

**** $p < .0001$

TABLE 2

SUMMARY OF ANOVA OF TEACHER EFFECTIVENESS

Source of Variation	SS	df	MS	F
Between	39,205	2	19,602	87.5 ****
Within	47,653	213	224	
Total	86,858	215		

TABLE 3

COEFFICIENT ALPHA, MEAN AND RANGE OF ITEM CORRELATION WITH
TOTAL SCORE FOR TEACHER EFFECTIVENESS SCALES (k = 51, k = 15)
DERIVED FROM THE ORIGINAL SAMPLE AND TWO RANDOM HALVES OF
THE ORIGINAL SAMPLE

	Total Sample (n = 72)	Random Sample 1 (n = 36)	Random Sample 2 (n = 36)
51 Item Scale			
$\alpha_{k\,=51}$	.99	.99	.98
$\bar{r}_{it}$	.70	.68	.73
	$(.19 < r_{it} < .86)$	$(.21 < r_{it} < .84)$	$(.12 < r_{it} < .91)$
15 Item Scale			
$\alpha_{k\,=51}$	.97	.96	.97
$\bar{r}_{it}$	.82	.79	.87
	$(.80 < r_{it} < .86)$	$(.76 < r_{it} < .84)$	$(.85 < r_{it} < .91)$

Table 4 lists the 15 best items yielded by the original analysis using the total sample of 72 subjects. Also shown are the 15 best items for the two random samples, indicated by s_1 or s_2 following each item. For example, Item 1 was among the top 15 from the original sample, and was also among the top 15 in Random Sample 1. As indicated on the table, 11 items were among the top 15 in all three analyses. There were also two items which appeared only in the top 15 for Random Sample 1, and two which appeared only in the analysis for Random Sample 2.

The coefficient alpha for the 15 items scale using the total sample was sufficiently large to warrant a scale composed of only these items. However, also included in the final scale (see Appendix E) are the additional four items which were among the top 15 in Random Samples 1 and 2. The inclusion of the remaining 32 items in the final form of the scale would be largely redundant and would not add to the scale's reliability.

Instructor Hiring Scale

The construction of the hiring scale was based on the procedure for homogeneous tests and was validated on 36 subjects. Coefficient alpha for the 33 item hiring scale was .97. The range of item correlation with total score was .54 - .87 with a mean of .70. A scale composed of the 15 items showing the highest correlation with total score had a coefficient alpha of .96. The range of item correlations on this scale with total score was .74 - .87, with a mean of .79. Again, coefficient alpha was sufficiently high to justify a hiring scale composed of only those 15 items which are listed on Table 5. Group separation was established by testing the significance of the difference between the mean total scores of the two criterion groups ($\underline{t}$ = 2.94, $\underline{p} < .01$, $\underline{df}$ = 17).

TABLE 4

TEACHER EFFECTIVENESS SCALE

Fifteen Best Items from Total Sample

1. Seems comfortable in teaching students of all ability levels (s_1) (s_2)

2. Has ability to grasp and hold class's attention when teaching (s_1) (s_2)

3. Knows when class is not understanding him and takes corrective steps (e.g. changes approach or explains more carefully) (s_1) (s_2)

4. Effectively communicates own ideas about philosophy of skiing and teaching (s_1) (s_2)

5. Is effective in pointing out errors without criticizing student (s_1) (s_2)

6. Has ability to effectively correct error (i.e. employs the best corrective exercise or teaching method for a given individual or group (s_1) (s_2)

7. Can choose most appropriate maneuver for a given terrain and snow condition (s_1) (s_2)

8. Has ability to demonstrate maneuvers properly (s_1) (s_2)

9. Understands teaching progressions and can implement them effectively (s_1) (s_2)

10. Has ability to accurately and quickly assess group's skiing abilities and deficiencies (s_1) (s_2)

11. Has ability to evaluate student's overall skiing ability and develop an individual program with flexibility to alter according to terrain and snow conditions (s_1) (s_2)

12. Handles class diplomatically (s_1)

13. Has ability to speak clearly and distinctly in front of class (s_1)

14. Is able to balance encouragement and support with criticism (i.e. is not overly supportive or overly critical of student) (s_2)

15. Gives adequate individual attention in group lessons when possible (s_2)

---

s_1 = Items among top 15 from Random Sample 1

s_2 = Items among top 15 from Random Sample 2

(continued)

Additional Items from Random Sample 1

1. Seeks out ways to improve teaching skills and abilities (e.g. is receptive to new teaching methods and techniques)

2. Shows interest in recent developments in teaching techniques

Additional Items from Random Sample 2

1. Places emphasis on keeping class moving (i.e. provides optimal balance between talking and skiing)

2. Has ability to appropriately use technical knowledge when teaching given a particular individual or group

TABLE 5

INSTRUCTOR HIRING SCALE

1. Seems to enjoy teaching

2. Displays interest in teaching skiing (i.e. primary motivation is to teach skiing and help others learn as opposed to being an instructor for pure ego fulfillment)

3. Seeks out ways to improve teaching skills and abilities (e.g. is receptive to new teaching methods and techniques)

4. Seems dedicated to teaching skiing

5. Shows interest in recent developments in teaching techniques

6. Is willing to give up own desires without begrudging when teaching

7. Has friendly attitude towards students

8. Relates to students as individuals (e.g. conveys a personal interest in them)

9. Is personable

10. Imparts enthusiasm

11. Has ability to grasp and hold class's attention when teaching

12. Is effective in pointing out errors without criticizing student

13. Has ability to effectively correct error (i.e. employs the best corrective exercise or teaching method for an individual or group)

14. Has ability to explain what has just been demonstrated

15. Has an attractive appearance

A correlation matrix for the variables obtained during the con-
struction of the instructor hiring scale appears on Table 6. The
variables include:total score on the hiring scale, whether or not
the applicant was hired, clinic leader's overall ranking, and whether
the individual was certified.

Prediction of Teacher Effectiveness

Several analyses were performed using data from the sample of 28
instructors who were ranked on the teacher effectiveness scale and who
also took the PRF. This sample was composed of ten "effective" teachers,
ten "medium" teachers, and eight "ineffective" teachers. Several signi-
ficant correlations were found between levels of effectiveness and cer-
tain PRF scales, as well as other relevant variables. Specifically,
increasing levels of effectiveness were negatively correlated with three
PRF scales, including Abasement ($r = -.44$, $p < .05$), Succorance ($r = -.38$,
$p < .05$), and Sentience ($r = -.38$, $p < .05$). Effectiveness was positively
correlated with certification level ($r = .86$, $p < .001$), years of experience
($r = .72$, $p < .001$), years at resort ($r = .43$, $p < .01$), and age ($r < .51$,
$p < .01$).

A series of univariate regressions (Nunnally, 1967) were performed
with teacher effectiveness as the dependent variable as measured by the
director's overall classifications of ineffective, medium and effective
teachers. The independent variables were: total score from the teacher
effectiveness scale, age, sex, years of experience, certification level,
education level, and five PRF scale scores (Abasement, Impulsivity,
Nurturance, Sentience, and Succorance). Impulsivity and Nurturance

TABLE 6

CORRELATIONS WITH TOTAL SCORE ON THE HIRING SCALE
WITH THE CRITERION MEASURE
AND OTHER RELEVANT VARIABLES

	Total Score	Hired vs. Not Hired	Clinic Leader Rank	Certification
Total Score	1.00	.88****	.86****	.40*
Hired vs. Not Hired		1.00	.86****	.43**
Clinic Leader Rank			1.00	.37*
Certification				1.00

Note: n = 36

* $p < .05$

** $p < .01$

*** $p < .001$

**** $p < .0001$

were included because the previous analysis showed trends towards significant correlations between these scales and teacher effectiveness. Table 7 summarizes the univariate regression F values for these independent variables.

A stepup multiple regression procedure (Nunnally, 1967) was used to determine the combination of variables which would best predict teacher effectiveness. Since total score on the teacher effectiveness scale accounted for the greatest amount of variance in the dependent measure (F = 114.09, p $<$.0001), it was combined with all the other independent variables individually. The two variable combination which accounted for the greatest amount of variance was retained. The process of adding the best variable to the combination of variables was continued until no variable added to the equation could significantly increase the amount of variance accounted for in the dependent measure. The following regression equation was thus derived:

$$Y = 2.56 - (.005)X_1 + (.03)X_2$$

where: Y = level of teacher effectiveness

X_1 = total score from teacher effectiveness scale

X_2 = level of certification

PRF Scores of Ski Instructors

Subjects who took the PRF consisted of 118 ski instructors (34% female, 66% male) from nine Rocky Mountain ski resorts. The average age of those sampled was 28.4 years old. The vast majority were officially certified instructors who considered ski instructing to be their primary occupation. On the average, subjects had been teaching for over five years and most had been employed at the resort at which they were

TABLE 7

UNIVARIATE REGRESSION F (1, 27) VALUES
FOR EACH OF THE ELEVEN INDEPENDENT VARIABLES
ON THE DEPENDENT VARIABLE OF TEACHER EFFECTIVENESS

Independent Variable	F Value (1, 27)
Total Score on Instructor Hiring Scale	114.09 ****
Age	7.46 *
Sex	.64
Years Experience	23.94 ***
Certification	76.03 ****
Education Level	.16
Abasement	8.94 **
Impulsivity	1.66
Nurturance	2.30
Sentience	4.77 *
Succurance	.17

* $p < .05$

** $p < .01$

*** $p < .001$

**** $p < .0001$

tested for four years or more. The group was generally well educated, with 93% having attended some college and 56% holding a bachelors or advanced degree. In general, then, subjects in this sample were established as professionals in their field, were employed in steady jobs, and were non-transient since most had been at their respective resorts for over four years.

To provide a comparison between the present sample and the normative sample (mean = 50, standard deviation = 10), means and standard deviations were computed and t-statistics calculated between the two groups (see Table 8). Seventeen of the 21 PRF scales (including Social Desirability) differentiated at a significant level between the two groups, with nine scales differing beyond the .001 level.

Significant correlations were found between certain PRF scales and other demographic and relevant variables. These relationships are summarized on Table 9.

TABLE 8

MEANS, STANDARD DEVIATIONS, AND t-VALUES
REFLECTING DEGREE OF DEPARTURE FROM PRF NORMATIVE SAMPLE

Scales on which instructors scored higher than PRF norm group

Scale	Mean	Standard Deviation	t
Abasement	51.95	8.43	2.54 *
Achievement	56.51	8.94	7.48 ****
Autonomy	57.89	9.44	9.07 ****
Cognitive Structure	52.57	9.09	3.07 **
Dominance	54.05	8.91	4.94 ***
Endurance	57.67	9.47	8.81 ****
Exhibition	51.83	8.98	2.20 *
Nurturance	51.92	9.03	2.39 *
Order	52.03	8.83	2.50 *
Sentience	60.46	9.69	11.75 ****
Understanding	52.84	9.89	3.12 **
Desirability	57.92	9.88	8.70 ****

Scales on which instructors scored lower than PRF norm group

Scale	Mean	Standard Deviation	t
Aggression	45.63	7.89	- 6.07 ****
Defedence	47.65	10.73	- 2.37 *
Harmavoidance	46.92	7.42	- 4.54 ***
Social Recognition	44.78	7.46	- 7.56 ****
Succorance	48.11	8.06	- 2.55 *

* $p < .05$
** $p < .01$
*** $p < .001$
**** $p < .0001$

TABLE 9

SIGNIFICANT CORRELATIONS BETWEEN PRF SCALES AND
OTHER DEMOGRAPHIC AND RELEVANT VARIABLES

Variable	PRF scale	r	
Age	Aggression	.27	**
	Endurance	.28	**
	Harmavoidance	.20	*
	Defedence	- .22	*
	Sentience	- .33	***
Education	Cognitive Structure	.20	*
	Desirability	.24	**
	Endurance	.21	*
	Understanding	.26	**
	Aggression	- .18	*
	Play	- .22	*
Female	Aggression	.27	**
	Succorance	- .18	*
Experience of Instructor	Endurance	.22	*
	Play	- .18	*
Years at Resort	Achievement	.18	*
	Endurance	.19	*
	Play	- .21	*
Certified Instructors	Abasement	- .22	*
Instructing is Primary Occupation	Abasement	- .20	*

* $p < .05$
** $p < .01$
*** $p < .001$

CHAPTER VI

DISCUSSION

The results of this study will be discussed by first presenting
data in support of the six research hypotheses postulated at the pro-
ject's inception. More elaborative discussions of the study's various
subsections will follow in this order: teacher effectiveness and hiring
scales, prediction of teacher effectiveness, PRF scores of ski instruc-
tors, and evidence contributing to the construct validity of the PRF.

Support of Research Hypotheses

An internal consistency approach was used to demonstrate the relia-
bility of the teacher effectiveness scale. An estimation of the amount
of systematic variance in a scale was on the basis of the average inter-
item correlation through coefficient alpha. Hypothesis One was thus
supported since the analysis yielded a coefficient alpha of .97.

The second hypothesis that the scale would be composed of several
factors relating to overall teacher effectiveness was not borne out.
Teacher effectiveness in ski instructors appears to be unidimensional,
as indicated by the high degree of homogeneity of the rating scale.

Evidence contributing to the validity of the teacher effectiveness
scale was demonstrated through the support of Hypotheses Three and Four,
as well as by the analysis indicating that the criterion measure was
a highly reliable one. With regard to the latter point, the correlation

of .94 between the director's and supervisor's overall rankings of the same nine individuals shows a striking degree of agreement on the criterion measure, making the possibility of criterion contamination a remote one. Thus, the correlation of .87 between the criterion and total score provides strong evidence for the validity of the teacher effectiveness scale. Additionally. Hypothesis Three was supported by the internal cross validational procedure which largely duplicated the results yielded by the original analysis using the total sample. This suggests that sampling fluctuations which would lead to different results would be unlikely. Finally, criterion-related validity was established through group separation using total score on the teacher effectiveness scale as the dependent measure. The analysis of variance yielded a highly significant difference between groups of effective, medium, and ineffective instructors.

Hypothesis Five was supported by the significant correlations between three PRF scales and level of teacher effectiveness. Effective teachers scored lower on the Abasement (r = -.44, $p < .05$), Succorance (r = -.32, $p < .05$), and Sentience scales (r = -.39, $p < .05$) than did ineffective teachers. Interpretations of these relationships will be discussed later in this section.

Finally, Hypothesis Six was supported by the significant differences between ski instructors and the PRF normative sample on 17 of the 21 scales. A summary of these results and possible interpretations will also be discussed later in this section.

Teacher Effectiveness and Hiring Scales

The results of this study demonstrate that a reliable and valid measure of teacher effectiveness in ski instructing can be constructed. Moreover, it is clear that teacher effectiveness is a domain composed of a tightly knit group of observable behaviors which co-occur in a predictable fashion. The domain is not multi-dimensional in nature as originally hypothesized. The homogeneity of the construct is most convincingly illustrated by the high degree of internal consistency of the rating scale produced in this study.

In addition to identifying the observables which relate to teacher effectiveness in ski instructing, the teacher effectiveness scale constructed in this study has practical utility in several areas. First, it can serve as a mid- or post-seasonal evaluative tool by objectively identifying both strengths and weaknesses in the teaching skills of an existing ski school staff. A standardized evaluation procedure could help facilitate employer/employee communication about an instructor's teaching performance while also motivating him to improve certain areas which have been identified as weaknesses.

If the scale was used on a regular basis at a number of different resorts, scores could also become part of the employee's permanent record. In the event that an instructor applies for a job at a different ski area, the hiring director would have a standardized, reliable and valid indication of that instructor's teaching effectiveness over time and could use this information in his personnel decisions.

Finally, the information yielded by this research could aid in refining the current PSIA certification process. At present this is

a non-standardized procedure under regional jurisdiction, with variable criteria and selection processes. Further knowledge of the skills proved empirically to be related to effective ski instructing may eventually contribute to a more unified, objective, and standardized certification procedure.

A second major finding of this study was the successful construction of an instructor hiring scale. While data collection in this area was not originally anticipated, it became clear during the course of the investigation that the data were available at two of the resorts visited. It is recognized that a cross validation of this scale is necessary since it was validated on a fairly small smaple (6 clinic leaders and 36 applicants). However, the results from the present sample strongly suggest that cross validation would also produce a reliable and valid scale.

The high reliability of the 15 item hiring scale constructed in this study was indicated by a coefficient alpha of .96. The criterion against which the scale was validated appeared to be a solid one since overall rankings assigned by clinic leaders correlated .88 with whether or not the applicant was hired. Evidence contributing to the validity of this scale was demonstrated by a correlation of .86 between total score and the criterion measure. A correlation of .86 between total score and clinic leader's overall rankings also provides convergent evidence for the scale's validity. Finally, criterion-related validity was demonstrated by significant differences on total score between the two criterion groups.

The present hiring scale is one composed of a variety of items which

can be reliably assessed following a relatively brief observational period of approximately four to five days. Assuming that these results are supported through cross validation, this scale can aid in objectifying the pre-seasonal instructor selection process.

Typically, the pool of applicants at a resort is divided into groups of approximately ten, led by a clinic leader who stays with them throughout the duration of the clinic. Sometimes leaders are rotated in order to increase the exposure to more applicants. Selection criteria range from subjective global opinions of clinic leaders, to numerical scores on a variety of areas such as free skiing, ability to demonstrate maneuvers, etc. Global rankings by clinic leaders were indeed found to correlate .88 with whether or not the applicant hired for the sample used in this study. If the selection criterion is limited only to the leader's subjective opinion, however, there is a problem in making objective comparisons among applicants from different clinic groups at a given resort. The use of a hiring scale would circumvent this problem by providing a reliable index which would allow for comparisons between groups. Moreover, items on the scale have been proved empirically to relate to teacher effectiveness whereas the other criteria sometimes used at various resorts have undoubtedly not been put to this test. For example, certification is often influential in deciding whether or not the applicant should be hired. It is interesting to note, however, that for the present sample, presence or absence of certification was only moderately correlated with clinic leader rankings, total score on the hiring scale, and whether or not the applicant was hired ($.37 < r < .43$).

While all items from the original item pool of 51 were shown to tap

the common domain of overall teacher effectiveness, it is interesting

to note that the final forms of each scale share little item overlap.

It should be recalled, however, that only 33 of the 51 original items

were considered suitable for ranking applicants from preseason clinics.

Items excluded pertained to skills and abilities which logically could

not be reliably assessed after a four to five day observational period

(e.g. items dealing with how well an instructor handled a class of

students). Of the 19 items on the final form of the teacher effective-

ness scale, only eight were included in the item pool from which the

hiring scale was constructed. Of those eight items, five were among

the top 15 in the hiring scale analysis and were thus included in the

final form of that scale. Thus, the relatively small degree of item

overlap between the two final forms of the scales does not indicate a

lack of agreement between the two studies concerning which items are

most highly related to teacher effectiveness. Rather, differing content

of the two scales is primarily due to the time variable which places a

restriction on the appropriateness of items which could be included on

the hiring scale.

A brief content analysis of the two scales further explains the

differences between them. The majority of the items on the hiring

scale seem to tap attitudinal variables such as dedication to teaching,

interest, motivation, enthusiasm, friendliness, etc. The items on the

teacher effectiveness scale tend to tap more specific teaching skills

such as error correction and class handling. The differences between the

two scales seems logical given the following argument.

From the experimenter's experiential knowledge of instructor clinics

in general, it is evident that the proficiency of skiing skill among applicants for ski instructing jobs is at the expert level. Therefore, what distinguishes among those who are hired and those who are not hired does not pertain primarily to skiing skill since the group is so homogeneous with respect to this variable. It has already been mentioned that distinctions cannot be made on the basis of specific teaching skills (e.g. class handling), since clinic leaders have not had the opportunity to observe applicants in actual teaching capacities. There-fore, the variables which best distinguish between those applicants who are hired and those who are not are basically attitudinal in nature as indicated by the items which emerged in the top 15 of the instructor hiring scale.

Following a longer time period (e.g. a ski season) during which instructors can be observed in teaching capacities, judgements regarding teacher effectiveness can be further refined to include the finer com-ponents of this domain. The items on the teacher effectiveness scale show which areas are most related to the assessment of teaching effective-ness under these conditions. Indeed, most of the items which differenti-ated best between effective and ineffective teachers could be classified under the broad categories of communication skills, error correction and class handling. Thus, as pointed out earlier, the differing content of the two scales can best be explained by a time variable which determines the amount of exposure the rater has had to the instructor's actual teach-ing performance.

Prediction of Teacher Effectiveness

One of the three major purposes of the present study was to investi-

gate the possibility of using certain PRF scales as predictors of
teacher effectiveness. The hypothesis was supported that certain
PRF scales were significantly correlated with teacher effectiveness.
Specifically, ineffective teachers scored significantly higher on the
Abasement, Succorance, and Sentience scales. That ineffective teachers
should be more self abasing is an expected relationship which follows
logically given their lower proficiency level. Likewise, they are
probably less confident in themselves and their teaching ability, and
would be more likely to depend on the support and advice of others as
suggested by the higher Succorance scale. The higher score on the Sen-
tience scale is somewhat puzzling, except when explained within the
context of age. Younger subjects from the total sample (n = 118) tended
to score higher on Sentience than did their older counterparts, a rela-
tionship which makes intuitive sense. Since age was positively correlated
with teacher effectiveness (r = .52, $p < .01$), it is possible that the more
ineffective teachers scored higher on Sentience largely as a function of
age rather than their lack of proficiency in teaching skiing.

The series of univariate regressions of the five PRF scales on
teacher effectiveness did in fact show that the Abasement and Sentience
scales accounted for a significant amount of the variance in the dependent
measure when used separately as the sole predictors. However, the step-
up multiple regression procedure indicated that the optimal combination
of predictor variables for teacher effectiveness was restricted to the
total score on the rating scale and certification level. It is evident,
then, that these PRF scales do not contribute anything over and above
the information already contained in these two major variables with
regard to teacher effectiveness. Therefore, they cannot be considered

to be efficient predictors if information regarding the other two variables is available.

Finally, it is interesting to note that total score on the teacher effectiveness scale was weighted more heavily in the regression equation than was certification level. This argues that for this sample, mere certification in and of itself does not necessarily guarantee that an instructor will be a good teacher. There is additional information to be gained concerning an instructor's teaching effectiveness through the use of more objective and explicit criteria as measured by the scale developed in this study.

PRF Scores of Ski Instructors

The third purpose of this study was to gather normative data on the PRF for ski instructors as a specific occupational group. Instructors differed significantly from PRF norms on 17 of the 21 scales, including the Social Desirability scale. Mean differences ranged in significance from the .05 to the .0001 level. Those scales on which instructors scored significantly higher at the .05 level were Abasement, Exhibition, Nurturance, and Order. At the same significance level, they scored lower than the PRF normative group on the Defedence and Succorance scales. Instructors scored higher at the .01 level on the Cognitive Structure and Understanding scales.

For purposes of clarity, this discussion will be primarily confined to the 11 scales which were significantly different beyond the .001 level, since these serve to highlight the dominant characteristics of this sample. The description of the present sample in relation to the PRF norms is summarized as follows:

High: Achievement   Low: Aggression
        Autonomy   Harmavoidance
        Dominance   Social Recognition
        Endurance
        Sentience
        Desirability

The first pattern which emerges in these results is the significantly higher scores on the Achievement and Endurance scales. Subjects in this sample described themselves as being highly achievement oriented, diligent in the pursuit of personal goals, and willing to work long hours. Indeed, casual observation during data collection suggests that many instructors put in 50 - 60 hours per week during the high season, often working six days a week.

The high Dominance scale suggests that instructors exhibit leadership qualities by holding and adhering to strong opinions and being influential and persuasive with others. The high Autonomy scale and low score on Social Recognition suggests that they are also highly independent people who are generally self-determined and individualistic. They do not appear to be overly concerned about what others think of them, nor do they work for the approval or recognition of others. In short, they seem to be a self confident group of people who adhere to the conventional values of goal orientation and hard work, but who prefer to do so in a way which is least inhibiting to their personal freedom and flexibility.

While instructors may have strong personalities in the manner described above, they are not likely to be abrasive in manifesting these characteristics, as indicated by a very low score on the Aggression scale and a moderately high score on Nurturance. This pattern suggests that they are a well socialized group and have found more socially acceptable ways to express their dominating, leadership qualities. Another indication of a

high level of socialization is the elevated Desirability scale, indicating that instructors describe themselves in a way which is favorable and traditionally socially desirable.

The low score on the Harmavoidance scale was an expected finding and is also the scale which yielded the least variability between subjects. Instructors consistently described themselves as enjoying exciting and dangerous activities, showing little concern for physical harm. Thus, they are not a cautious group physically which appears to be functional and adaptive given the nature of their job.

Finally, the scale with the highest elevation was the Sentience scale. This group described themselves as being extremely attuned to physical sensations. They seem sensitive to many forms of experiencing and are perceptive and responding to aesthetic stimuli. They are generally tuned into their environment and consider this to be an important part of life. Given their strong orientation towards work and the diligent pursuit of personal goals, the extreme elevation on the Sentience scale seems to round out the unique profile of the ski instructor as suggested by this sample.

To generalize these findings as descriptive of instructors as a whole, we must assume that this group was reasonably representative of instructors nationwide. This assumption is tenable since efforts were taken to sample from both small resorts, as well as large destination resorts from four states in the Rocky Mountain region. The ages of the group sampled ranged from 16 to 62, and there were varying levels of certification and experience among them.

Assuming that these findings can be reasonably generalized to the population of ski instructors, the importance of this aspect of the study

is twofold. First, these findings will aid in professionalizing ski instructing and may help to defray some of the myths which seem prevalent in the minds of many recreational skiers. The highly achievement and work oriented profile of this group does not fit the image of the "laid back", carefree, glorified ski bum. This longstanding stereotype may simply be the product of what recreational skiers expect instructors to be like and may be perpetuated because skiers do not have the opportunity to prove or disprove this image through their own personal experience.

Secondly, the highly significant differences between an occupational group from a normal population and PRF norms decisively reiterates the fact that college normed summary statistics in general should be applied with caution to non-college populations. These findings highlight the need for more normative data on specific occupational groups so that the PRF can be validly used in applied settings such as businesses and outpatient clinics. At present, it is unfortunate that such a psychometrically sound instrument is confined primarily to research settings.

Evidence Contributing to the Construct Validity of the PRF

The validation of an objective personality test is an empirical procedure involving the accumulation of data supporting the theoretical postulates that relate the constructs of interest to observable behavior. Significant relationships between certain PRF scales and other variables such as age and educational level were yielded by the analysis conducted in the present study. These relationships fit a pattern one would predict on an apriori basis from the theoretical networks defining the constructs which are measured by the PRF. It is recognized that the sample of 118 ski instructors is a highly circumscribed group both occupationally and

and geographically. Certain relationships reported here are thus only tentatively generalizable to the population as a whole. The data seem consistent enough, however, to warrant further discussion.

The first major pattern emerged in the form of significant scale correlations with the subject's age, which ranged from 16 to 62 with a mean of 28.4. Scores on the Harmavoidance and Endurance scales increased with age, while Aggression, Defedence, and Sentience decreased. Theses are all relationships one would expect with increasing age. The older people get, the more cautious they become in their physical activities and the more risk avoidant they are in general. Conversely, younger people tend to participate more in vigorous, potentially dangerous sports and are generally less concerned with physical safety. Age relationships with the Endurance scale appear to reflect general maturity, with older subjects describing themselves as more patient, willing work work long hours, and perservering in the face of difficulty. Likewise, maturity is suggested in the lower Aggression scale, reflecting even temperedness and diplomacy in the face of disagreement rather than being blunt, pushy, and argumentative. Related to this is a lower Defedence score for older subjects, meaning they are less sensitive to criticism and are generally more accepting of themselves. Finally, older subjects seem less attuned to bodily and environmental sensations are are not as open to sensual experiences as their younger counterparts.

A second major variable which yielded significant correlations with PRF scales was education, which ranged from the high school level to the master's degree for this particular sample. Educational level was positively correlated with Understanding, Cognitive Structure, Endurance, and Desirability, and negatively correalted with Aggression and Play. Again,

these relationships seem to make intuitive sense. First, more educated
people tend to be more probing, analytical and inquisitive, with energies
often directed towards satisfying intellectual curiosity. They tend to
have well developed systems by which to synthesize and organize information,
which is relfected in the elevated Cognitive Structure scale. The high
Endurance score suggests they are more perservering and hard working, a
characteristics which is certainly facilitative in attaining a college
degree. The lower Play scale converges well with this, by suggesting they
are a more serious group and generally less carefree and amusement oriented
than their less educated cohorts. A lower Aggression score and high
Desirability score is most reflective of the higher level of socialization
one would expect to find in a more educated group.

Female instructors were more aggressive and less succorant than
their male counterparts, which is exactly opposite from the PRF normative
group. This seems to be understandable given the sex role stereotyping
which is prevalent among skiers. Women are viewed as less able skiers
and are more likely to have to prove themselves as athletically competent.
That women would have to be more aggressive and self sufficient in order
to obtain a job as a ski instructor is not surprising. Passive, helpless,
and defenseless women would probably not gain entry into this occupation
which has tradiationally been male oriented.

The following variables were also found to be significantly correlated
with certain PRF scales: experience of instructor, years at resort,
certification, full time vs. part time, and whether or not teaching was
the subject's primary occupation. Experienced instructors scored signifi-
cantly higher on Endurance and lower on Play. Again, this is not a surpris-
ing relationship. As the years go by, instructing undoubtedly becomes more

of a job and less of a pleasurable thing to do with one's winters.
The same relationships were found between these scales and how long
an instructor had been at a particular resort. This was a logical find-
ing since there is a correlation of .71 between years of experience and
how many years an instructor had been employed at a resort. The less
transient instructors also showed a higher score on Achievement, which
suggests they are more stable and career oriented than their more mobile
counterparts.

Certified instructors and those who considered ski instructing to
be their primary occupation scored lower on Abasement. Being well esta-
blished and accomplished in one's occupation would logically tend to fos-
ter a less self-critical, humble and apologizing attitude.

In summary, a number of relationships were found between PRF scales
and other non-test variables which we might expect on the basis of our
theoretical and experiential knowledge of these variables and constructs.
The data discussed here thus contribute to the evidence in support of
construct validity of the Personality Research Form.

CHAPTER VII

SUMMARY

The primary goal of this study was the development of a reliable and valid assessment instrument for teacher effectiveness in ski instructors. A 51 item rating scale was validated on 72 instructors from eight Rocky Mountain ski resorts. Nine instructors were chosen by a supervisor at each resort as being representative of "effective", "average", and "ineffective" teachers. The total sample was thus composed of three criterion groups of 24 representing these levels of teacher effectiveness.

The validation procedure involved the ski school director ranking the nine instructors from his resort on each of the 51 items. He then assigned them ranks from one to nine on the basis of their overall teacher effectiveness. The global director rank served as the criterion measure against which the scale was validated. Global ranks were also obtained from one to four additional supervisors at each resort. A correlation of .94 between supervisor and director overall ranks revealed the high reliability of the criterion measure and the remote possiblity of criterion contamination.

Evidence for criterion related validity was established through group separation. An analysis of variance yielded significant differences between the three groups of instructors using total score on the rating scale as the dependent measure. Post-hoc analysis by means of the Newman-Keuls test showed significant differences at the .001 level for every pair of means.

An analysis including all 51 items yielded a coefficient alpha of .99

and a mean item correlation with total score of .70. A second scale composed of the 15 items showing the highest correlations with total score had a coefficient alpha of .97. An internal cross validation procedure using two equal random samples (n = 36) yielded similar results for both the 51 and 15 item scales.

The reliability of the teacher effectiveness scale was demonstrated by the high value of coefficient alpha. Evidence contributing to the scale's validity was established by the reliability of the criterion measure, the correlation of .87 between total score and the criterion measure, the successful internal cross validation of the scale's construction, and the significant separation of criterion groups using total score as the dependent measure.

An instructor hiring scale was constructed using a similar procedure to that described above. The criterion group used to validate the scale was composed of 18 instructors who had recently been hired by a resort and 18 unsuccessful applicants for the same job. Six clinic leaders ranked groups of 6 instructors (three from each criterion group) on each of 33 items from the original pool of 51.

The scale's reliability was established through a coefficient alpha of .97 for the 33 item scale. A scale composed of the 15 items showing the highest correlations with total score yielded a coefficient alpha of .96. Evidence for the scale's validity was demonstrated by a correlation of .86 between total score and the criterion measure. Group separation was also established by a significant difference between the mean total scores of the two criterion groups.

The second purpose of this study was to investigate the possibility of using certain PRF scales as predictors of teacher effectiveness.

Several analyses were performed using PRF scores from 28 instructors who were also ranked on the teacher effectiveness scale. Increasing levels of effectiveness were found to be negatively correlated with Abasement, Succorance, and Sentience.

A series of univariate regressions were performed to predict teacher effectiveness using these scales and other variables as independent measures. Abasement and Sentience were found to account for a significant amount of the variance when used separately as the sole predictors. However, a stepup multiple regression procedure revealed that the optimal combination of predictor variables was composed only of total score on the teacher effectiveness scale, and the level of instructor certification. Thus, the PRF scales were shown not to be efficient predictors of teacher effectiveness compared to these two major variables.

The third purpose of this study was to gather normative data on the PRF for ski instructors as a specific occupational group. One hundred and eighteen instructors from nine Rocky Mountain ski resorts differed significantly from PRF norms on 17 of the 21 scales. Mean differences ranged in significance from the .05 to the .0001 level. In increasing order of magnitude, instructors scored higher on the following scales: Exhibition, Nurturance, Order, Abasement, Cognitive Structure, Understanding, Dominance, Achievement, Desirability, Endurance, Autonomy, and Sentience. They scored lower on Defedence, Succorance, Harmavoidance, Aggression, and Social Recognition. Speculative descriptions of the average ski instructor were offered using the data presented above.

Significant correlations were also found between certain PRF scales and variables such as age, education, sex, and instructor experience.

These relationships were ones which might be expected on the basis of theoretical and experiential knowledge of these variables and the constructs measured by the PRF scales. The data were thus presented as contributing to the evidence supporting the construct validity of the Personality Research Form.

The results of this study demonstrated that a reliable and valid measure of teacher effectiveness in ski instructing could be constructed. The homogeneity of this construct was illustrated by the high degree of internal consistency of the rating scale produced in this study. The successful construction of an instructor hiring scale showed that reliable ratings measuring this domain could also be obtained following a brief observational period of four to five days. While certain PRF scales were shown to predict teacher effectiveness, a multiple regression procedure yielded total score on the rating scale and level of instructor certification as the most efficient predictors of teacher effectiveness. Finally, significant differences between ski instructors and PRF norms on 17 scales reiterated the fact that college normed summary statistics must be applied with caution to non-college populations.

REFERENCES

Adinolfi, A.A. The characteristics of highly accepted, highly re-
jected, and relatively unknown university freshmen. Unpublished
Doctoral Dissertation, University of Rochester, 1970. (DAI 31:
2271B).

Agocs, H., & Suvak, A. Ski instructors are different - in some ways.
Professional Ski Instructors of America newsletter, 1977.

Anastasi, A. Personality Research Form. In O.K. Buros (Ed.), The
Seventh Mental Measurements Yearbook, Highland Park, New Jersey:
Gryphon Press, 1972.

Anastasi, A. Psychological Testing. New York: MacMillan Publishing
Co., Inc., 1976.

Astin, A.W. Criterion-centered research. Educational and Psychologi-
cal Measurement, 1964, 24, 807 - 822.

Beecher, D.E. The Evaluation of Teaching, Backgrounds and Concepts.
New York: Syracuse University Press, 1949.

Bendig, A.W. The relation of level of course achievement to students'
instructor and course ratings in introductory psychology.
Educational and Psychological Measurement, 1953, 13, 437 - 448.

Bendig, A.W. Ability and personality characteristics of introductory
psychology instructors rated competent and empathetic by the
students. Journal of Educational Research, 1955, 48, 705 - 709.

Bentler, P.M., & Prince, C. Personality characteristics of male
transvestites: III. Journal of Abnormal Psychology, 1969, 74,
140 - 143.

Buros, O.K.,(Ed.). The Fifth Mental Measurements Yearbook, Highland
Park: Gryphon Press, 1959.

Buros, O.K.,(Ed.). The Sixth Mental Measurements Yearbook. Highland
Park: Gryphon Press, 1965.

Buros, O.K.,(Ed.). The Seventh Mental Measurements Yearbook. Highland
Park: Gryphon Press, 1972.

Brown, F.G. Principles of Educational and Psychological Testing.
New York: Holt, Rinehart & Winston, 1970.

Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation
by the multitrait-multimethod matrix. Psychological Bulletin,
1959, 56, 81 - 105.

Costin, F. Do student ratings of college teachers predict student achievement? Teaching of Psychology, 1978, 5, 86 - 88.

Costin, F., Greenough, W.T., & Menges, R.J. Student ratings of college teaching: reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511 - 535.

Cronbach, L.J. Studies of acquiescence as a factor in the true-false test. Journal of Educational Psychology, 1942, 33, 401 - 415.

Cronbach, L.J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475 - 494.

Cronbach, L.J. Further evidence on response sets and test design. Educational and Psychological Measurement, 1950, 10, 3 - 31.

Cronbach, L.J. California Personality Inventory. In O.K. Buros (Ed.), The Fifth Mental Measurements Yearbook, Highland Park: Gryphon Press, 1959.

Cronbach, L.J., & Gleser, G.C. Psychological Tests and Personnel Decisions. Urbana: University of Illinois Press, 1957.

Cronbach, L.J., & Meehl, P.E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281 - 302.

Doyle, W. Paradigms for research on teacher effectiveness. In L.S. Shulman (Ed.), Review of Research in Education, Itasca; F.E. Peacock, 1978.

Dunkin, M.J., & Biddle, B.J. The Study of Teaching. New York: Holt, Rinehart & Winston, 1974.

Edwards, A.L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. Journal of Applied Psychology, 1953, 37, 90 - 93.

Edwards, A.L. Statistical Methods. New York: Holt, Rinehart, & Winston, 1973.

Frey, P.W. Student ratings of teaching: validity of several rating factors. Science, 1973, 182, 83 - 85.

Gage, N.L. The yield of research on teaching. Phi Delta Kappan, 1978, 229 - 235.

Ghiselli, E.E., & Brown, C.W. Personnel and Industrial Psychology. New York: McGraw-Hill Book Co., 1948.

Ghiselli, E.E. Dimensional problems of criteria. _Journal of_ _Applied Psychology_, 1956, 40, 1 - 4.

Ghiselli, E.E. _Theory of Psychological Measurement_. New York: McGraw-Hill Book Co., 1964.

Gough, H.G. _California Psychological Inventory Manual_. Palo Alto: Consulting Psychologists Press, Inc., 1975.

Heath, R.W., & Nielson, M.A. The research basis for performance based teacher education. _Review of Educational Research_, 1974, 44, 463 - 483.

Heilbrun, A.B. Edwards Personal Preference Schedule. In O.K. Buros (Ed.), _The Seventh Mental Measurements Yearbook_. Highland Park: Gryphon Press, 1972.

Heilbrun, A.B., & Goodstein, L.D. Social desirability response set: error or predictor variable. _Journal of Psychology_. 1961, 51 321 - 329 (a).

Heilbrun, A.B., & Goodstein, L.D. The relationships between individually defined and group defined social desirability and performance on the Edwards Personal Preference Schedule. _Journal of Consulting Psychology_, 1961, 25, 200- 204 (b).

Hicks, L.E. Some properties of ipsative, normative and forced normative measures. _Psychological Bulletin_, 1970, 74, 167 - 184.

Hoffman, H. Note on the personality traits of student nurses. _Psychological Reports_, 1970, 27, 1004 (a).

Hoffman, H. Personality characteristics of alcoholics in relation to age. _Psychological Reports_, 1970, 27, 176 - 171 (b).

Jackson, D.N. Multimethod factor analysis in the analysis of convergent and discriminant validity. Paper presented at the Society for Multivariate Experimental Psychology, Atlanta, Georgia, 1966. Cited in D.N. Jackson, _Personality Research Form Manual_. New York: Research Psychologists Press, 1967.

Jackson, D.N. _Personality Research Form Manual_. New York: Research Psychologists Press, 1967.

Jackson, D.N., & Guthrie, G.M. The convergent and discrimination validation of the Personality Research Form scores. Cited in D.N. Jackson, _Personality Research Form Manual_. New York: Research Psychologists Press, 1967.

Jackson, D.N., & Messick, S. (Eds.), _Problems in Human Assessment_. McGraw-Hill Book Co., New York, 1967.

Jackson, D.N., & Lay, C. Homogeneous dimensions of personality scale content. London, Canada: University of Western Ontario Research Bulletin, 1967. Cited in D.N. Jackson, Personality Research Form Manual. New York: Research Psychologists Press, 1967.

Jones, L.V., & Fiske, D.W. Models for testing the significance of combined results. Psychological Bulletin, 1953, 50, 375 - 381.

Kelly, E. California Personality Inventory. In O.K. Buros (Ed.), The Sixth Mental Measurements Yearbook. Highland Park: Gryphon Press, 1965........

Kelly, E. Personality Research Form. In O.K. Buros (Ed.), The Seventh Mental Measurements Yearbook. Highland Park: Gryphon Press, 1972.

Kuder, G.F., & Richardson, M.W. The theory of the estimation of test reliability. Psychometrika, 1937, 2, 151 - 160.

Kulik, J.A., & McKeachie, W.J. The evaluation of teachers in higher education. In F.N. Kerlinger (Ed.), Review of Research in Education. Itaska: F.E. Peacock, 1975.

Kusyszyn, I.A. A comparison of judgemental methods with endorsement in the assessment of personality traits. Journal of Applied Psychology, 1968.

Lanyon, R.I., & Goodstein, L.D. Personality Assessment. New York: John Wiley and Sons, Inc., 1971.

Loevinger, J. Objective tests as instruments of psychological theory. In Jackson & Messick, (Eds.), Problems in Human Assessment. McGraw-Hill Book Co., New York, 1967.

McKee, M.G. Edwards Personal Preference Schedule. In O.K. Buros (Ed.), The Seventh Mental Measurements Yearbook. Highland Park: Gryphon Press, 1972.

Murray, H.A. Explorations in Personality. New York: Oxford University Press, 1938.

Murray, H.G. Predicting student ratings of college teaching from peer ratings of personality types. Teaching of Psychology, 1975, 2, 66 - 69.

Nunnally, J.C. Psychometric Theory. New York: McGraw-Hill Book Co., 1967.

Radcliffe, J.A. Review of the Edwards Personality Preference Schedule. In O.K. Buros (Ed.), The Sixth Mental Measurements Yearbook. Highland Park: Gryphon Press, 1965.

Rodin, M., & Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1164 - 1166.

Rorer, L.G. The great response-style myth. Psychological Bulletin, 1965, 63, 129 - 156.

Ryans, D.G. Characteristics of teachers: their description, comparison and appraisal: a research study. Cited in L.C. Deighton (Ed.), Encylopedia of Education, The Macmillan Co. and the Free Press, 1971.

Siess, T.F., & Jackson, D.N. Vocational interests and personality: an empirical integration. Journal of Counseling Psychology, 1970, 17, 27 - 35.

Shavelson, R., & Dempsey, N. Generalizability of measures of teacher effectiveness. Review of Educational Research, 1976, 46, 553 - 560.

Snedecor, G.W. & Cochran, W.G. Statistical Methods, Aimes: The Iowa State University Press, 1967.

Sorey, L.E. A study of the distinguishing personality characteristics of college faculty who are superior in regard to the teaching function. Unpublished doctoral dissertation, 1968. (DAI 28: (12-A) 4916).

Stickell, D.W. A critical review of the methodology and results of research comparing televised and face to face instruction. Cited in T.C. Taveggia, Resolving research controversy through empirical cumulation: toward reliable sociological knowledge. Sociological Methods and Research, 1974, 2, 403.

Stricker, L.J. Edwards Personal Preference Schedule. In O.K. Buros (Ed.), The Sixth Mental Measurements Yearbook. Highland Park: Gryphon Press, 1965.

Thorndike, R.L. California Personality Inventory. In O.K. Buros (Ed.), The Fifth Mental Measurements Yearbook. Highland Park: Gryphon Press, 1959.

Trott, D.M., & Jackson, D.N. An experimental analysis off acquiescence. Journal of Experimental Research in Personality, 1967.

Walsh, J.A. California Personality Inventory. In O.K. Buros (Ed.), The Seventh Mental Measurements Yearbook. Highland Park: Gryphon Press, 1972.

Wessler, Ruth and Loevinger, J. Personality Research Form. In O.K. Buros (Ed.), The Seventh Mental Measurements Yearbook. Highlaud Park: Gryphon Press, 1972.

Wiggins, J.S. Personality Research Form. In O.K. Buros (Ed.), The Seventh Mental Measurements Yearbook . Highland Park: Gryphon Press, 1972.

Winegardner, J. Prediction of vocational outcome using the Person-
    ality Research Form. Unpublished Masters Thesis, University
    of Montana, 1978.

Yin, R.K., Bingham, E., & Heald, K.A. The difference that quality
    makes: the case of research reviews. Sociological Methods
    and Research, 1976, 5, 153 - 154.

APPENDIX A

Below is a list of qualities, characteristics and abilities which may contribute to being an effective ski instructor. You will also find a rating scale of numbers and corresponding descriptions. Please use this scale to rate the following items in terms of how important you think they are in contributing to the effective teaching of skiing.

For example, if you feel item #1 is very important in order to teach skiing effectively, put a 7 in the blank beside the item. If you think item #2 is not at all important in being an effective teacher, give that item a rating of 1. Now suppose that you consider item #3 to be somewhere between moderately important and very important. You would rate this item with a 5 or 6, depending on which rating best reflects your opinion.

Important: Please try and rate the items according to what you have found to be generally true in your experience with teaching skiing. For example, you may feel that in general, item #4 is very important. However you may know of one instructor who you consider to be an excellent teacher, yet he/she does not exhibit this quality at all. You would still rate item #4 as very important since you have found it to be true in most cases.
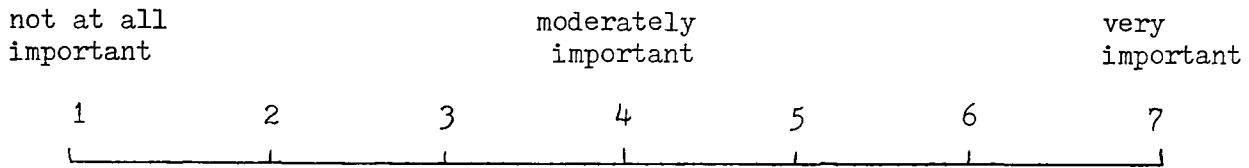
not at all important			moderately important			very important
1	2	3	4	5	6	7

Attitude

_____ 1. Seems to enjoy teaching

_____ 2. Invites comments and criticism of teaching ability

_____ 3. Displays interest in teaching skiing (i.e. primary motivation is to teach skiing and help others learn as opposed to being an instructor for pure ego-fulfillment).

_____ 4. Seeks out ways to improve teaching skills and abilities (e.g. is receptive to new teaching methods and techniques)

_____ 5. Seems dedicated to teaching skiing

_____ 6. Is willing to go beyond minimum requirements of the job (e.g. puts in extra hours if necessary, occasionally gives class longer lessons,)

_____ 7. Shows interest in recent developments in teaching techniques

_____ 8. Is willing to give up own desires without begrudging (e.g. having to teach a beginning class on a day when there's two feet of fresh powder)

Personality and Interpersonal Skills

_____ 9. Has friendly attitude towards students

_____ 10. Relates to students as individuals (i.e. conveys a personal interest in students)

(OVER)

_____ 11. Recognizes and greets students out of class

_____ 12. Has sense of humor

_____ 13. Is patient with slow learners and students far below own ability level

_____ 14. Is personable

_____ 15. Handles class diplomatically

_____ 16. Imparts enthusiasm

_____ 17. Treats students on equal level as self (i.e. does not exude an air of superiority)

_____ 18. Possesses adequate self-confidence

_____ 19. Is capable of creating a relaxed atmosphere and minimizes tension in students

_____ 20. Seems comfortable in dealing with people on both a group and individual level

_____ 21. Seems comfortable in dealing with both children and adults

_____ 22. Seems comfortable in teaching students of all ability levels

Communication Skills

_____ 23. Has ability to speak clearly and distinctly in front of class

_____ 24. Has ability to grasp and hold class's attention when teaching

_____ 25. Knows when class is not understanding him and takes corrective steps (e.g. changes approach or explains more carefully)

_____ 26. Encourages and is receptive to questions

_____ 27. Effectively communicates own ideas about philosophy of skiing and teaching

Error Correction

_____ 28. Has ability to identify errors

_____ 29. Has ability to analyze errors

_____ 30. Is effective in pointing out errors without criticizing student

_____ 31. Has ability to effectively correct error (i.e. employs the best corrective exercise or teaching method for an individual or group)

_____ 32. Can choose most appropriate maneuver for a given terrain and snow condition

_____ 33. Has ability to demonstrate maneuvers properly

_____ 34. Has ability to explain what has just been demonstrated

_____ 35. When necessary, places emphasis on corrective exercises and repeats them often during lesson

_____ 36. Is able to balance encouragement and support with criticism (i.e. is not overly supportive or overly critical)

_____ 37. Understands teaching progressions and can implement them effectively

Class Handling

_____ 38. Has ability to accurately and quickly assess group's skiing abilities and deficiencies

_____ 39. Choose appropriate terrain given the range of ability and limitations of class

_____ 40. Places emphasis on keeping the class moving (i.e. provides optimal balance between talking and skiing)

_____ 41. Is able to make some progress with students in a variety of time limitations (i.e. is effective in teaching both single lessons and a week long series of lessons)

_____ 42. Gives adequate individual attention in group lessons when possible

_____ 43. Has ability to evaluate student's overall skiing ability and develop an individual program with flexibility to alter according to terrain and snow conditions (individual lessons)

Skiing Skill and Knowledge

_____ 44. Is knowledgable in the physical and technical aspects of skiing (e.g. body functioning, anatomy, etc.)

_____ 45. Has ability to appropriately use technical knowledge when teaching given a particular individual or group (i.e. knows when to be technical and when to be more intuitive in approach)

_____ 46. Has ability to understand and perform maneuvers at all levels of skiing ability

_____ 47. Has knowledge of a variety of corrective exercises at all levels of skiing ability

Personal Appeal

_____ 48. Is often requested by former students or people referred by former students

_____ 49. Has an attractive appearance

_____ 50. Is well-groomed

Please list any characteristics or abilities you consider important which are not included here and rate them using the scale on the first page.

Thank you very much for your participation!

APPENDIX B

ITEM POOL FOR INSTRUCTOR HIRING SCALE

Attitude

1. Seems to enjoy teaching

2. Invites comments and criticism of teaching ability

3. Display interest in teaching skiing (i.e. primary motivation is to teach skiing and help others learn as opposed to being an instructor for pure ego-fulfillment

4. Seeks out ways to improve teaching skills and abilities (e.g. is receptive to new teaching methods and techniques)

5. Seems dedicated to teaching skiing

6. Is willing to go beyond minimum requirements of the job (e.g. puts in extra hours if necessary, occasionally gives longer classes, etc.)

7. Shows interest in recent developments in teaching techniques

8. Is willing to give up own desires without begrudging

Personality and Interpersonal Skills

9. Has friendly attitude towards students

10. Relates to students as individuals (e.g. conveys a personal interest in them)

11. Has sense of humor

12. Is personable

13. Imparts enthusiasm

14. Treats students on equal level as self (i.e. does not exude an air of superiority)

15. Possesses adequate self-confidence

16. Is capable of creating a relaxed atmosphere and minimizes tension in students

Communication Skills

17. Has ability to speak clearly and distinctly in front of class

18. Has ability to grasp and hold class's attention when teaching

Error Correction

19. Has ability to identify errors

20. Has ability to analyze errors

21. Is effective in pointing out errors without criticizing student

22. Has ability to effectively correct error (i.e. employs the best corrective exercise or teaching method for an individual or group)

23. Can choose most appropriate maneuver for a given terrain and snow condition

24. Has ability to demonstrate maneuvers properly

25. Has ability to explain what has just been demonstrated

26. Understands teaching progressions and can implement them effectively

Skiing Skill and Knowledge

27. Is knowledgable in the physical and technical aspects of skiing (e.g. body functioning, anatomy, etc.)

28. Has ability to appropriately use technical knowledge when teaching a particular individual or group

29. Has ability to understand and perform maneuvers at all levels of skiing ability

30. Has knowledge of a variety of corrective exercises at all levels of skiing ability

Personal Appeal

31. Has an attractive appearance

32. Is well-groomed

33. Is dependable

APPENDIX C

PERSONALITY RESEARCH FORM SCALE DESCRIPTIONS (Jackson, 1967)

Scale	Description of high scorer	Defining trait adjectives
Abasement	Shows a high degree of humility; accepts blame and criticism even when not deserved; exposes himself to situations where he is in an inferior position; tends to be self-effacing	Meek, self-accusing, self-blaming, obsequious, self-belittling, surrending, resigned, self-critical, humble, apologizing, subservient, obedient, yielding, deferential, self-subordinating
Achievement	Aspires to accomplish difficult tasks; maintains high standards and is willing to work towards distant goals; responds positively to competition; willing to put forth effort to attain excellence.	Striving, accomplishing, capable, purposeful, attaining, industrious, achieving, aspiring, enterprising, self-improving, productive, driving, ambitious, resourceful, competitive.
Affiliation	Enjoys being with friends and people in general; accepts people readily; makes efforts to win friendships and maintain associations with people	Neighborly, loyal, warm, amicable, good-natured, friendly, companionabbe, genial, affable, cooperative, gregarious, hospitable, sociable, affiliative, good-willed.
Aggression	Enjoys combat and argument; easily annoyed; sometimes willing to hurt people to get his way; may seek to "get even" with people whom he perceives as having harmed him.	Aggressive, quarrelsome, irritable, argumentative, threatening, attacking, antagonistic, pushy, hot-tempered, easily-angered, hostile, revengeful, belligerent, blunt, retaliative
Autonomy	Tries to break away from restraints, confinement, or restrictions of any kind; enjoys being unattached, free not tied to people, places, or obligations; may be rebellious when faced with restraints.	Unmanageable, free, self-reliant, independent, autonomous, rebellious, unconstrained, individualistic, ungovernable, self-determined, non-conforming, uncompliant, undominated, resistant, lone-wolf.

Change	Likes new and different exper-iences; dislikes routine and avoids it; may readily change opinions or values in different circumstances; adapts readily to changes in environment	Inconsistent, fickle, flexible, unpredictable, wavering, mutable, adaptable changeable, irregular, variable, capricious, innovative, flighty, vacillating, inconstant
Cognitive structure	Does not like ambiguity or un-certainty in information; wants all questions answered complet-ly; desires to make decisions based upon definite knowledge rather than guesses or pro-babilities	Precise, exacting, definite, seeks certainty, meticulous, perfectionistic, clarifying, explicit, accurate, rigor-ous, literal, avoids ambi-guity, defining, rigid, need structure.
Defedence	Readily suspects that people mean him harm or are against him; ready to defend himself at all times; takes offense easily; does not accept criticism readily	self-protective, justify-ing, denying defensive, self-condoning, suspicious, secretive, has a 'chip on the shoulder', resists inquiries, protesting, wary, self-excusing, rationalizing guarded, touchy.
Dominance	Attempts to control his environment, and to influ-ence or direct other people; expresses opinions forcefully, enjoys the role of leader and may assume it spontaneously	governing, controlling, commanding, domineering, influential, persuasive, forceful, ascendant, leading directing, dominant, asserti powerful, supervising
Endurance	Willing to work long hours; doesn't give up quickly on a problem; perservering, even in the face of great difficulty; patient and unrelenting in his work habits	Persistent, determined, stea fast, enduring, unfaltering, perservering, unremitting, relentless, tireless, dogged energetic, has stamina, stur zealous, durable
Exhibition	Wants to be the center of attention; enjoys having an audience; engages in behavior which wins the notice of others; may enjoy being dramatic or witty	Colorful, entertaining, unus spellbinding, exhibitionisti conspicuous, noticeable, expressive, ostentatious, immodest, demonstrative, flashy, dramatic, pretentiou showy.

Harmavoidance	Does not enjoy exciting activities, especially if danger is involved, avoids risk of bodily harm; seeks to maximize personal safety	Fearful, withdraws from danger, self-protecting, pain-avoidant, careful, cautious, seeks safety, timorous, apprehensive, precautionary, unadventurous, apprehensive, avoids risks, attentive to danger, stays out of harm's way; vigilant
Impulsivity	Tends to act on the 'spur of the moment' and without deliberation; gives vent readily to feelings and wishes; speaks freely; may be volatile in emotional expression	Hasty, rash, uninhibited, spontaneous, reckless, irrepressible, quick-thinking, mercurial, impatient, incautious, hurried, impulsive, foolhardy, excitable, impetuous.
Nurturance	Gives sympathy and comfort; assists others whenever possible, interested in caring for children, the disabled, or the infirm; offers a 'helping hand' to those in need, readily performs favors for others.	Sympathetic, paternal, helpful, benevolent, encouraging, caring, protective, comforting, maternal, supporting, aiding, ministering, consoling, charitable, assisting.
Order	Concerned with keeping personal effects and surroundings neat and organized; dislikes clutter, confusion, lack of organization; interested in developing methods for keeping materials methodically organized.	Neat, organized, tidy, systematic, well-ordered, disciplined, prompt, consistent, orderly, clean, methodical, scheduled, planful, unvarying, deliberate.
Play	Does many things "just for fun", spends a good deal of time participating in games, sports, social activities, and other amusements; enjoys jokes and funny stories; maintains a light-hearted, easy-going attitude toward life.	Playful, jovial, jolly, pleasure seeking, merry, laughter-loving, joking, frivolous, prankish, sportive, mirthful, fun-loving, gleeful, carefree, blithe.
Sentience	Notices smells, sounds, sights, tastes, and the way things feel; remembers these sensations and believes they are important part of life; is sensitive to many forms of experience; may maintain an essentially hedonistic or aesthetic view of life	Aesthetic, enjoys physical sensations, observant, earthy, aware, notices environment, feeling, sensitive, sensuous, open to experience, perceptive, responsive, noticing, discriminating, alive to impressions.

Social Recognition	Desires to be held in high esteem by acquaintances; concerned about reputation and what other people think of him; works for the approval and recognition of others	Approval seeking, proper, well-behaved, seeks recognition, courteous, makes good impression, seeks respectability, accommodating, socially proper, seeks admiration, obliging, agreeable, socially sensitive, desirous of credit, behaves appropriately.
Succorance	Frequently seeks the sympathy, protection, love, advice, and reassurance of other people; may feel insecure or helpless without such support; confides difficulties readily to a receptive person.	Trusting, ingratiating, dependent, entreating, appealing for help, seeks support, wants advice, helpless, confiding, needs protection, requesting, craves affection, pleading, help-seeking, defenseless.
Understanding	Wants to understand many areas of knowledge; values synthesis of ideas; verifiable generalization, logical thought, particularly when directed at satisfying intellectual curiosity.	Inquiring, curious, analytical, exploring, intellectual, reflective, incisive, investigative, probing, logical, scrutinizing, theoretical, astute, rational, inquisitive.
Desirability	Describes self in terms judged as desirable; consciously or unconsciously, accurately or inaccurately, presents favorable picture of self in responses to personality statements.	
Infrequency	Responds to implausible or pseudo-random manner, possibly due to carelessness, poor comprehension, passive non-compliance, confusion, or gross deviation.	

APPENDIX D

INFORMATION SHEET

Name _____ Age _____ Sex _____

1. How many years have you been a ski instructor? _____

2. How many years have you been employed at this resort? _____

3. What is your current level of certification? _____ uncertified
                                                 _____ associate
                                                 _____ fully certified

4. Are you currently a part time or full time instructor?
                                                 _____ part time
                                                 _____ full time

5. Do you consider ski instructing to be your primary occupation
   during skiing season?   ____ yes
                                                 ____ no

   If not, what is your primary occupation at this time?

   _____

6. What is your primary off-season occupation? _____

7. Please indicate the highest level of education you have reached

                                 _____ High school
                                 _____ Some college
                                 _____ B.A. or B.S.
                                 _____ Some graduate work
                                 _____ M.A. or M.S.
                                 _____ PhD

APPENDIX E

TEACHER EFFECTIVENESS SCALE

1. Seems comfortable in teaching students of all ability levels
2. Has ability to grasp and hold class's attention when teaching
3. Knows when class is not understanding him and takes corrective steps (e.g. changes approach or explains more carefully)
4. Effectively communicates own ideas about philosophy of skiing and teaching
5. Is effective in pointing out errors without criticizing student
6. Has ability to effectively correct error (i.e. employs the best corrective exercise or teaching method for a given individual or group
7. Can choose most appropriate maneuver for a given terrain and snow condition
8. Has ability to demonstrate maneuvers properly
9. Understands teaching progressions and can implement them effectively
10. Has ability to accurately and quickly assess group's skiing abilities and deficiencies
11. Has ability to evaluate student's overall skiing ability and develop an individual program with flexibility to alter according to terrain and snow conditions
12. Handles class diplomatically
13. Has ability to speak clearly and distinctly in front of class
14. Is able to balance encouragement and support with criticism (i.e. is not overly supportive or overly critical of student)
15. Gives adequate individual attention in group lessons when possible
16. Seeks out ways to improve teaching skills and abilities (e.g. is receptive to new teaching methods and techniques)
17. Shows interest in recent developments in teaching techniques
18. Places emphasis on keeping class moving (i.e. provides optimal balance between talking and skiing)
19. Has ability to appropriately use technical knowledge when teaching given a particular individual or group