ISSN 0119-5107

Philippine Normal University
CENTER FOR RESEARCH AND
DEVELOPMENT IN EDUCATION
Manila

RESEARCH SERIES

No. 88 October 2006

STUDENTS' RATINGS OF PROFESSORS' COMPETENCE: AN APPLICATION OF THE G THEORY

JESUS ARCE OCHAVE
Vice President for Planning, Research and Extension

EDNA LUZ RAYMUNDO-ABULON
Faculty Researcher

STUDENTS' RATINGS OF PROFESSORS' COMPETENCE: AN APPLICATION OF THE G THEORY

Jesus Arce Ochave
Vice President for Planning, Research and Extension

Edna Luz Raymundo-Abulon
Faculty Researcher

INTRODUCTION:

Student evaluation or student ratings of college teachers' performance is the most widely used technique to measure competence inside the classroom here and abroad. Aleamory (1998) offers the following arguments to support the use of student ratings:

1. Students are the main source of information about the learning environment, including teacher's ability to motivate students for continued learning, rapport or degree of communication between the instructor and students.

2. Students are more logical in evaluating the effectiveness of teacher and their satisfaction with course content, method of instruction, textbooks, home works and seat works.

3. Ratings given by students encourage communication between students and their instructor. The communication my lead to the kind of student and instructor involvement in the teaching-learning process that can raise the level of instruction.

On the other hand, some research findings claimed that the use of the ratings given by the students are unreliable in the sense that students lack the maturity and expertise to make sound judgment about the course content or instruction style. Most of the time, popularity was measured rather than ability of the instructor.

Some critiques argue that the ratings are widely affected by the personal style of the instructor rather than the instructor's ability to convey instructional materials. The research of Abrami, Leventhal and Perry (1996) concluded that considerable portion of the differences in instructor's student ratings is based on perceived personality characteristics rather than instructional effectiveness.

Fieldman (1997) elicited a moderate to high positive correlation between Instructor's personality and student ratings. A meta-analysis of a dozen of students involved in his study revealed that instructor's expressiveness had a substantial impact on student ratings but a small impact on student achievement.

Dowell and Neal (1993) contended that ratings given by students are inaccurate indicators of student learning. They can only be best regarded as indices of "consumer satisfaction" rather than "teaching effectiveness".

When it comes to the consistency of student ratings for a particular instructor, Peterson and Kauchak (1993) found out that student ratings are consistent and reliable from one year to the next. This was concurred in a study done by Aleamory (1998) which indicates that the correlation between student ratings of the same instructors and courses ranged from .70 to .87.

Research evidences further support the view that carefully constructed evaluation instrument with well-developed procedures for administration can yield high internal consistency reliabilities. Most of the research evidences supporting validity of forms in which student ratings were compared with other methods of evaluation indicated the existence of high to moderate positive correlation.

Zeroing into the current study, the focus is no longer on the argument that student ratings really measure teaching effectiveness. Admittedly, the prevailing measure of teaching effectiveness in universities in our country is still student ratings. Hence, the researchers deemed it more appropriate to focus the study on the following issues: How dependable is the scale in differentiating professors' competence? Is the scale sensitive enough to detect differences among the members of faculty in terms of their competence inside the classroom?

More explicitly, the issue is on the precision of the student-assigned ratings in measuring professors' competence. It can be pointed out that only when ratings on the scale are precise should the ratings be considered in accounting for competence, especially when the ratings are used for decisions like retention, promotion and advancement of professors.

Precision, or accuracy to others, refers to the sensitivity of an instrument to detect true differences. If the evaluation scale is precise, it can discriminate a professor who demonstrates more of the quality which is competence from another professor who possesses less of this quality.

Dependability of the scale or the precision of the student-assigned ratings in terms of detecting differences in professors' competence is indeed a gap in the existing body of knowledge in relation to behavioral measures.

The Generalizability Theory

The Generalizability Theory has been described as a liberalization of classical test theory because it gets away from a tedious separate calculation of reliability coefficients for equivalence, internal consistency and stability. Most especially, it has the capability to identify multiple sources of errors in measurement.

Knowing the sources and magnitude of errors in measurement is very important since if these errors are too large, they threaten the precision of an instrument.

Analysis through the use of the generalizability theory (Cronbach, 1972) suggests instrument improvement by its ability to specify and estimate sources of errors, the information much needed in our system of performance evaluation to arrive at optimum desired precision and dependability.

PURPOSE

The primary purpose of this research is to find out whether the current faculty evaluation scale being used at the Philippine Normal University is able to detect differences in professors' competence inside the classroom.

In carrying out this purpose, the researchers applied the system of the generalizability theory as used in a study done during the last quarter of the 20th century in the University of Toledo, Ohio, USA (Ochave, 1977).

In the study of Ochave (1977), the index of the ability of the scale to differentiate faculty in terms of teaching effectiveness is called a generalizability coefficient. Different generalizability coefficients were calculated to suit the varied purposes of a user of the scale. The process of calculating these coefficients suggest how an obtained observed average class rating is to be correctly interpreted.

The generalizability coefficients calculated in the study are as follows:

1. Generalizability Coefficient $E\rho^2$ (osi), generalizing over occasions (i.e. semesters), students and items. This is an index of the precision of the observed rating (average class rating) to differentiate professors' competence and is interpreted without reference to the occasions, the students and the items employed in obtaining the rating.

2. Generalizability Coefficient $E\rho^2$ (so), generalizing over students and occasions. This is an index of the precision of the rating to detect differences in professors' competence given the items used (i.e. item facet is fixed) regardless of students and occasions (student and occasion facets are random). Generalizing over students and occasion is useful to the user of the scale whose interest lies on the teaching effectiveness as rated given the items used in the scale without reference to any specific sample of students, or rather, with reference to all the students who could have rated the faculty and to any occasions.

3. Generalizability Coefficient $E\rho^2$ (i), generalizing over items. This is an index of the rating scale's capacity to detect differences between faculty as regards to their

competence inside the classroom given the students and occasions employed. Generalizing over items is of interest to a user of the scale who wishes to determine how well the items used in the scale represent those items that could have been included in the scale.

4. Generalizability Coefficient $E\rho^2$ (s), generalizing over students. This is an index of the precision of the scale to detect differences in professors' competence given the items used and the occasion when the ratings were elicited (i.e., items and occasions are fixed) regardless of any particular set of students who responded to the scale (student facet is random). Generalizing over students is of interest to those who would like to know the precision of the ratings assigned to Professors given the items used in the scale. It is of interest also to users who want to know how the Professors are rated by independent sets of students at fixed occasion.

5. Generalizability Coefficient $E\rho^2$ (o), generalizing over occasions. This has something to do with the precision of the scale to detect differences in professors' competence using the scale (items are fixed) and given the students who responded to the scale (students are fixed) regardless of occasions. This is of interest to a user of the scale who wants to be informed if the scale is precise across repeated administration given a particular scale and the same set of students.

   In the actual data used in this study however, it did not involve retesting for the same set of students at different occasions. The generalizability coefficient, $E\rho^2$ (o), and other coefficients wherein the student facet was considered fixed, was justified by assuming a finite

population of students from which subsets of students were sampled.

6. Generalizability Coefficient $E\rho^2$ (oi), generalizing over occasions and items. This refers to the precision of the scale given the responding set of students (students fixed facet) regardless of items used and occasions responded to. This will be useful to those who are interested on using alternate forms of the scale or if the interest lies on how well a given rating represents all other ratings assigned by the same set of students across several different occasions.

7. Generalizability Coefficient $E\rho^2$ (si), generalizing over students and items. This refers to the precision of the scale to detect differences in professors' competence given an occasion (fixed) for random sets of items used in the scale which are responded to by independent random samples of students. This coefficient is of practical interest to a user of the scale who needs information on how well Professors are rated given an occasion (semester), by different sets of students, all of whom are represented by the students who responded to the scale, and who would have responded to randomly selected scales.

The meaning of each of the generalizability coefficients is a determination of precision vis-à-vis the facets of generalization. The Professors as the unit of analysis in the study, the universe of generalization are the facets of occasions, students and items. These are the facets of generalization because they are *a priori* thought of to be sources of variance.

It is worthy to note that the course facet was not included in the universe of generalization since the data available could not

warrant at its present form. However, the course facet was involved in the research design but only for the control of its general effect on the precision of the instrument and not as a facet of generalization.

Research Problems

The seven research problems addressed by the study are the following:

1. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence where the universe of generalization is composed of the occasion, student, and item facets?

2. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence given the conditions of the items used, and the universe of generalization is composed of occasion and student facets?

3. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to the specific students and occasions employed and generalization over items?

4. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to items used and a given set of occasions while generalization is over students?

5. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to the set of students who

responded to the particular scale while the universe of generalization is over occasion?

6. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to the set of students from whom ratings were obtained while the universe of generalization is composed of the item and occasions facets?

7. What is the precision of the PNU faculty evaluation scale in detecting differences in professors' competence given the occasion used while the universe of generalization is composed of the item and student facets?

RESEARCH HYPOTHESES

With the seven research problems posed in the study, the research hypotheses listed below were tested. The professors as the facet of differentiation and the universe of generalization is composed of the facets of occasion, student and item. The obtained rating variance hence, is *a priori,* composed of the variance component due to faculty plus the separate variance components due to occasion, student, item, and the unexplained residual.

1. The PNU faculty evaluation scale is able to detect differences in professors' competence where the universe of generalization is composed of the facets of occasion and student and items.

2. The PNU faculty evaluation scale is able to detect differences in professors' competence for the items used with the universe of generalization composed of occasion and student facets.

3. The PNU faculty evaluation scale is able to detect differences in professors' competence given students who responded to the scale in one occasion with the universe of generalization as the item facet.

4. The PNU faculty evaluation scale is able to detect differences in professors' competence for the items used and the occasion employed while the universe of generalization is the student facet.

5. The PNU faculty evaluation scale is able to detect differences in professors' competence for the given set of students responding to the particular items employed and the universe of generalization as the occasion facet.

6. The PNU faculty evaluation scale is able to detect differences in professors' competence given the set of students who responded to the scale while the universe of generalization is composed of the item and occasion facets.

7. The PNU faculty evaluation scale is able to detect differences in professors' competence given the occasions used while the facets of students and items defined the universe of generalization.

Each of the above stated hypotheses assumes that there is more precision when the relevant generalizability coefficient approaches the value of 1.000

IMPORTANCE OF THE STUDY

If the instrument being used to rate the professors' competence inside the classroom is considered precise and dependable, ratings would have more credibility. This will be of paramount concern in every educational institution since this aids the management for sound decision-making especially for retention, promotion and advancement of faculty.

DEFINITION OF TERMS

The following terms used in the study are operationally defined as guided by the study of Ochave (1977):

1. Dependability. This refers to the accuracy of generalizing from a person's observed score on a test/scale or measure to the average score that person would have received under all the possible conditions that the test/scale would be equally likely to accept. Dependability is relative to the "universe" to which we wish to generalize. The universe is a term in the G Theory which pertains to the sources of variation , or the facets.

2. Fixed Facet. When all the conditions of a facet are exhausted and used in the study, it is considered fixed. Whenever a facet is considered fixed, as in some generalizability coefficients obtained in this study, there is no sampling error and generalization beyond the conditions used can not be claimed. The variance component due to the facet becomes part of universe score variance.

3. Generalizability Coefficients. They are indices of the precision of an instrument in detecting variance between the objects of measurement. Generalizability Theory allows for choices of alternative generalizability coefficient depending on the purpose of the user of the scale. These are similar to the reliability coefficients except that they are able to deal formally with the facet that a given rating can be generalized in various ways, like over items, over students or raters, over occasion, and the like.

4. Precision. It is used interchangeably with the concept of "dependability" of the instrument under study. This refers to the sensitivity of an instrument to detect true differences.

5. Random Facet. When many other conditions of a facet are not exhausted in the study, it is considered a random facet. Unless otherwise specified in the generalizability coefficient, the conditions for every facet used were assumed to be randomly sampled from infinite number of conditions. While not infinite, the number of conditions should be definitely large in relation to the number of conditions actually employed.

6. Professors' Competence. As used in the study, this construct was measured by the PNU Faculty Evaluation Scale. Ratings of students provide the index of professors' competence inside the classroom.

7. Universe of Generalization. In general, a universe of generalization is composed of the random facets included in the study. These facets are included because they are *a priori* thought of to be sources of error variance if generalization over all these facets was desired. In this study, the universe of generalization was composed of occasion, student and item facets.

8. Universe Score. There are alternative universe scores because they are, within the constraint of the universe of generalization, different ways by which an observed professor's competence rating can be interpreted. An example of this is the coefficient generalizing over occasions, students and items which pertains to the average of all admissible observations. That is, the competence rating assigned to a given Professor averaged across all the possible ratings if all the corresponding conditions (levels) of the facets such as occasion, student and items were obtained and the mean across these conditions was calculated. The computed average, hence, is a universe score.

9. Variance Components. These are the raw data for determining generalizability coefficients. They are the expected mean squares corresponding to a source of variance (i.e. facet) in an analysis of variance design. It is through the components of variance analysis that the generalizability theory liberated classical test theory.

METHODS AND PROCEDURES

In the application of the Generalizability Theory, a research design that embodied to answer the seven specific research questioned was employed.

Aptly called as S:O:T:C by I design provided the magnitude of separate variance components which were preliminary data for answering the research problems posed in the study.

The symbols S,O,T,C, and I stand, respectively, for students, occasions (in this study it pertains to semesters), teachers or professors, courses and items. This is a

hierarchically nested design and only the item facet crosses all other facets. The unit of analysis was the teacher or the professor, alternatively being referred to as the facet of differentiation. To control the course effect, course content and course level being possible sources of differential rating assigned, the professors involved were nested within courses. Figure I illustrates the research design and shows the relationships between the facets.

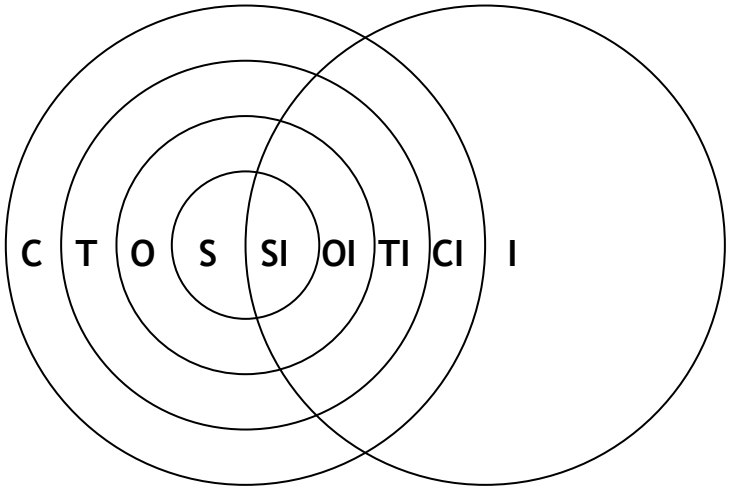

Figure 1. An illustration of (S:O:T:C) by I Design

In the Venn Diagram illustrated in Figure 1, the following symbols were utilized for the sources of observed score variance: C, T, O, S. and I, respectively for the courses, teachers or professors, occasions or semesters, students and items. The symbols SI, OI, TI, and CI are the first order interactions obtainable through the design.

Table 1 presents through symbols the sources of variance and their corresponding variance components. One can better appreciate the data in Table 1 by referring to Figure 1, the Venn diagram which illustrates the design. Since the professor is the facet being differentiated in order to determine the precision of the faculty evaluation scale, attention must be focused on the teacher or professor (T(c)) source of variation and its expected mean squares or variance components. The symbol δ^2 stands for the magnitude of a variance component. The letters s, o ,t, c and I stand for the number of students, occasions (semesters), teachers (professors), courses, and items, respectively. The new symbol or letter which is "n" stands for the number of replications or cases per cell in the design. Since in this study n=1, one can drop that coefficient.

Table 1. Sources of Variance and Expected Variance Components
Design (S:O:T:C) x I

Sources	df	Expected Variance Components
C	C-1	$n\delta^2_{is\,(cto)} + s\delta^2_{io\,(ct)} + i\delta^2_{s\,(cto)}$ $+ os\delta^2_{it\,(c)} + si\delta^2_{o\,(ct)} + to\delta^2_{ci}$ $+ iso\delta^2_{t(c)} + soti\delta^2_{c}$
I	k-1	$n\delta 2$ is (cto) $+ s\delta 2$ io (ci) $+ os\delta 2$ it (c) $+ tos\delta 2$ si $+ sotc\delta 2$ i
T(C)	C(T-1)	$n\delta 2$ is (cto) $+ s\delta 2$ io (ct) $+ i\delta 2$ s (cto) $+ os\delta 2$ it (c) + si$\delta 2$ o (ct) $+ iso\delta 2$ t (c)

Sources	df	Expected Variance Components
CI	(C-1) (k-1)	$n\delta^2_{is(cto)} + s\delta^2_{io(ct)} + os\delta^2_{it(c)} + tos\delta^2_{ci}$
O(CT)	CT(0-1)	$n\delta^2_{is(cto)} + s\delta^2_{io(ct)} + i\delta^2_{s(cto)} + si\delta^2_{o(ct)}$
TI (C)	C(T-1) (K-1)	$n\delta^2_{is(cto)} + s\delta^2_{io(ct)} + os\delta^2_{it(c)}$
S(CTO)	OTC(s-1)	$n\delta^2_{is(cto)} + io\delta^2_{s(cto)}$
IO(CT)	(k-1) (0-1CT)	$n\delta^2_{is(cto)} + S\delta^2_{io(cto)}$
IS (CTO)	(k-1) (S-1) CTO	$n\delta^2_{is(cto)}$

Calculation of Variance Components, Design: (S:O:T:C) x I

The following formulas were used in computing for the variance components:

1. Estimated variance component for the teacher/faculty:

$$\delta^2 T(c) = \frac{MS_{t(c)} - MS_{o(ct)} - MS_{it(c)} + MS_{io(ct)}}{iso}$$

2. Estimated variance component for the occasion nested in teacher/faculty and course:

$$\delta^2 o(cT) = \frac{MS_{o(cT)} - MS_{s(cTo)} - MS_{io(cT)} + MS_{is(cTo)}}{si}$$

3. Estimated variance component for the teacher/faculty and item interaction:

$$\delta^2 \text{ it(c)} = \frac{MSi_{t(c)} - MSi_{o(cto)}}{so}$$

4. Estimated variance component for the student nested in occasion, teacher/faculty and course:

$$\delta^2 \text{ s(cto)} = \frac{MS_{s(cto)} - MS_{is(cto)}}{i}$$

5. Estimated variance component for the occasion and item interaction:

$$\delta^2 \text{ io(ct)} = \frac{MS_{io(ct)} - MS_{is(cto)}}{s}$$

6. Estimated variance component for the student and item interaction:

$$\delta^2 \text{ is(cto)} = \frac{MSi_{s(cto)}}{n} = \frac{MS_{is(cto)}}{1}$$

Calculation of the Generalizability Coefficients

The following symbolic formulas were used to calculate the seven generalizability coefficients. Each of these coefficients corresponds to the specific problems stated earlier.

Problem number one:

$$E\rho^2\,(osi) = \frac{\delta^2\ t(c)}{E\ (\delta)\ ^2x_1}$$

Problem number two:

$$E\rho^2\,(os) = \frac{\delta^2\ t(c) + 1/k'\quad \delta^2\ it(c)}{E\ (\delta)^2\ x_1}$$

Problem number three:

$$E\rho^2\,(i) = \frac{\delta^2\ t(c) + 1/1'\quad \delta^2\ o(ct) + 1/n'\ \delta^2\ s(cto)}{E\ (\delta)^2\ x_1}$$

Problem number four:

$$E\rho^2\,(s) = \frac{\delta^2\ t(c) + 1/1'\quad \delta^2\ o(ct) + 1/k'\ \delta^2\ it(c)}{E\ (\delta)^2\ x_1}$$

Problem number five:

$$E\rho^2\,(o) = \frac{\delta^2\ t(c) + 1/n'\quad \delta^2\ s(cto) + 1/k'\ \delta^2\ it(c)}{E\ (\delta)^2\ x_1}$$

Problem number six:

$$E\rho^2\,(oi) = \frac{\delta^2\ t(c) + 1/n'\quad \delta^2\ s(cto)}{E\ (\delta)^2\ x_1}$$

Problem number seven:

$$E\rho^2\,(is) = \frac{\delta^2\ t(c) + 1/1'\quad \delta^2\ o(ct)}{E\ (\delta)^2\ x_1}$$

Sources of Data

The Faculty Evaluation Scale

As prescribed in the Faculty Manual of the Philippine Normal University, performance appraisal shall be conducted at the end of each term, that is, semester and summer. For this purpose, a University-wide Faculty Evaluation Scale was developed and validated. The Office of the Vice President for Academics handled the administration of the scale to the classes.

It consists of twenty-four items written in English with Filipino translation. There are four sub-scales namely: commitment, knowledge of the subject matter, teaching for independent learning and management of student learning. For the twenty-four items, the students are asked to rate their teacher using the following scale: (5) = Always; (4) = Most of the time; (3) = Sometimes; (2) = Rarely; and (1) = Never. The students are also asked to provide additional comments/observations on the course and on the faculty in-charge of the class.

This twenty-four-item-rating scale served as the item facet (I) under study. This instrument served also as the measure of the Professors' competence inside the classroom.

Target Population

For the Professor/teacher facet (T), a total of eight (8) professors [two (2) from the College of Arts and Social Sciences, two (2) from the College of Science, two (2) from the College of Education and two (2) from the College of Languages, Linguistics and Literature] were sampled. For the purpose of controlling the course effect, each set of two Professors was nested in a common course taught.

For the occasion facet (O), the results of the faculty evaluation during the first and second semesters of School Year 2004-2005 were used.

For the student facet (S), those enrolled in the courses that were sampled were involved in the study.

Calculations

Calculations of the mean squares, the variance components corresponding to each source of variance in the design and the seven generalizability coefficients were made possible via the SAS Programming Language available at the Computer Laboratory of the School of Statistics, University of the Philippines – Diliman Campus.

Calculations of the variance components as well as the seven generalizability coefficients were computed also by hand following the formulas as adopted in the Study of Ochave (1977), shown in the preceding sections of the study.

RESULTS AND DISCUSSION

Teacher/Professor: The Unit of Analysis

Table 2 represents the mean squares necessary for the calculation of the variance components. The mean square is the weighted sum of the variance components. In this study, there were six variance components and these components are due to the teacher (professor), the occasion, the item-teacher interaction, the students, the item-occasion interaction. These components are assumed to be independent of each other.

The numeric value of the mean square considering the general professors' competence inside the classroom (T(C)) is 184.128576. This mean square indicates immediately observable differences in professors' competence. However, it is worthy to note that the difference is a result of separate sources or effects, with these effects being the components of variance that make up the mean square.

As concurred with the study of Ochave (1977), employing the generalizability theory, as a conceptual frame of reference, the observed ratings on Professors' competence upon which the above mean squares were calculated are really the results of the separate and combined conditions under which the ratings on teaching effectiveness were obtained. These conditions are deliberately selected on the *a priori* ground that they have affected the observed rating. Therefore, the next logical step is to estimate the magnitude of these effects. It is in this phase that the components of variance analysis is needed.

Table 2. Sources of Variance Design (S:O:T:C) by I Overall Teaching Effectiveness

Sources	A. DF	Mean Square
Course	2	63.787914
Items	23	3.172878
T(C)	3	184.128576
CI	46	1.347701
O(CT)	6	84.379440
IT(C)	69	0.915670
S(CTO)	156	5.472487
OI(CT)	138	0.584250
SI(CTO)	3588	0.374027

As can be gleaned in Table 3, the variance components that accounted for most variance was the item-student interaction (δ^2 is(cto)).

Table 3. Estimated Expected Variance Components General Twenty-four Item Scale

Variance Components	Value
Estimated variance component for the teacher/faculty - δ^2 T(c)	**0.14794303**
Estimated variance component for the occasion nested in teacher/faculty and course- δ^2 o(ct)	**0.234217946**
Estimated variance component for the teacher/faculty and item interaction - δ^2 it(c)	**0.011836643**
Estimated variance component for the student nested in occasion, teacher/faculty and course – δ^2 s(cto)	**0.212435833**
Estimated variance component for the occasion and item interaction - δ^2 io(ct)	**0.015015929**
Estimated variance component for the student and item interaction - δ^2 is(cto)	**0.374027**

The presence of greater than zero variance components for other sources is indicative that the conditions under which the overall Professors' competence ratings obtained were the causes for the observed rating differences among instructors. The item-student interaction has the highest proportion of variance explained by the size of .37402 / .90, or 41% of the observed variance. This may suggest that the items were responded too differently by different students.

The second highest variance component was that for the occasion nested in teacher/professor and course or δ^2 o(ct). The proportion of variance explained was 0.234217946 / .90 or 26%. If the interpretation of the rating is made with respect to the occasion when the ratings given, precision of the scale would at least be .26. This precision could be increased if the levels in other facets were increased. Increasing the number of occasions

involved in the study would effect the lowering of the δ^2 o(ct) error, thus, increasing the level of precision.

The third highest variance component was that for the student nested in occasion, teacher/professor and course (δ^2 s(cto)). The proportion of variance explained was 0.212435833 / .90 or 24%. This variance is not considered error if generalization over students is not desired. That is, if the interpretation of rating is made with respect to the students who rated the professors, precision of the scale would at least be .24%. This precision could also be increased if the number of students involved shall be increased.

The errors contributed by the rest of the components were considered minimal. This implies that a fewer number of levels or conditions corresponding to the other facets could be used without jeopardizing the precision of the scale.

Generally, the computed variance components appear to be not that large to affect the precision of the instrument or the Faculty Evaluation Scale of PNU.

The seven generalizability coefficients calculated as presented in Table 4 answered the seven research questions posted in the study.

Table 4. Calculated Generalizability Coefficients General Twenty-four Item Scale

$E\rho^2$ (osi)	**.524352102**
$E\rho^2$ (os)	**.526100088**
$E\rho^2$ (i)	**.993197839**
$E\rho^2$ (s)	**.941164903**
$E\rho^2$ (o)	**.57988101**
$E\rho^2$ (oi)	**.578133024**
$E\rho^2$ (is)	**.939416917**

The level of precision of the PNU Faculty Evaluation Scale in detecting differences in professors' competence where the universe of generalization is composed of the occasion, student, and item facets is .524352102.

The level of precision of the PNU faculty evaluation scale in detecting differences in professors' competence given the conditions of the items used, and the universe of generalization is composed of occasion and student facets is .526100088

The level of precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to the specific students and occasions employed and generalization over items is .993197839

The level of precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to items used and a given set of occasions while generalization is over students is .941164903

The level of precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to the set of students who responded to the particular scale while the universe of generalization is over occasion is .57988101

The level of precision of the PNU faculty evaluation scale in detecting differences in professors' competence with reference to the set of students from whom ratings were obtained while the universe of generalization is composed of the item and occasions facets is .578133024

The level precision of the PNU faculty evaluation scale in detecting differences in professors' competence given the occasion used while the universe of generalization is composed of the item and student facets is .939416917

Considering the calculated coefficients that answered the seven research questions were from .50 to .99, the seven research hypotheses are considered tenable. It was postulated that there is more precision when the generalizability coefficient approaches the value of 1.000 (Winer, 1971). The PNU Faculty Evaluation scale, thus, can detect differences in Professors' competence encompassing all the facets of generalization.

CONCLUSION

The PNU Faculty Evaluation Scale possessed moderate to high degree of precision or dependability. It has the ability thereby to detect professors' competence inside the classroom. Specific conclusions vis-à-vis the facets of generalization used in the study are the following:

1. The scale is generally able to differentiate professors' competence without reference to occasions or to a particular semester that the scale was administered, to particular students who rated the faculty and to the items employed in obtaining the rating.

2. The scale is able to differentiate professors' competence given the 24 items used regardless the set of students and occasions employed.

3. The scale, with its 24 items, is able to differentiate professors' competence given the students and occasions employed. Hence, the items used in the scale were well represented by those items that could have been included in the scale.

4. The scale is able to differentiate professors' competence given the items used and the occasion regardless of any particular set of students who responded to the scale.

5. The scale is able to differentiate professors' competence given the students who responded to the scale regardless of occasions. Hence, the scale is considered precise across repeated administration given a particular scale and the same set of students.

6. The scale is able to differentiate professors' competence given the responding set of students regardless of items used and occasions responded to. Thus, the average rating given by a class could represent all other ratings assigned by the same set of student across several different occasions.

7. The scale is able to differentiate professors' competence given a particular occasion for random sets of items in the scale which are responded to by independent random samples of students.

RECOMMENDATION

While there was evidence that the scale had moderate to high degree of precision, it is recommended that a regular conduct of the faculty evaluation be done for every course taught by each faculty member in every semester so that a follow up study with increased number of professors, semesters/occasions and students be undertaken. This follow-up study can further dwell on the magnitude of variance accounted for by the variance component due to the course effect.

Construct validation of the faculty evaluation scale of the university is also recommended.

References

1. Abrami, Leventhal & Perry (1996) Student Ratings and Instructors' Personal Style as cited in www.vitaide.com/png/ERIC/Student-Evaluation.html

2. Aleamory (1998) Student Ratings as cited in www.vitaide.com/png/ERIC/Student-Evaluation.html

3. Cronbach, Lee (1972) The Dependability of Behavioral Measures. Stanford University, New York. 113-159

4. Dowell and Neal (1993) Faculty Performance Indicators as cited in www.vitaide.com/png/ERIC/Student-Evaluation.html

5. Fieldman (1997) Instructor's Personality and Student Ratings Journal of Educational Psychology 45 (10) 1183-1184

6. Ochave, J. (1977) Generalizability Theory: An Application in the Study of the Reliability of Student Ratings of Instructor Teaching Effectiveness. Unpublished Dissertation. University of Toledo, Ohio, USA

7. Peterson & Kauchak (1993) www.vitaide.com/png/ERIC/Student-Evaluation.html

8. Winer, B.J. (1971) Statistical Principles in Experimental Design (2nd ed). New York.