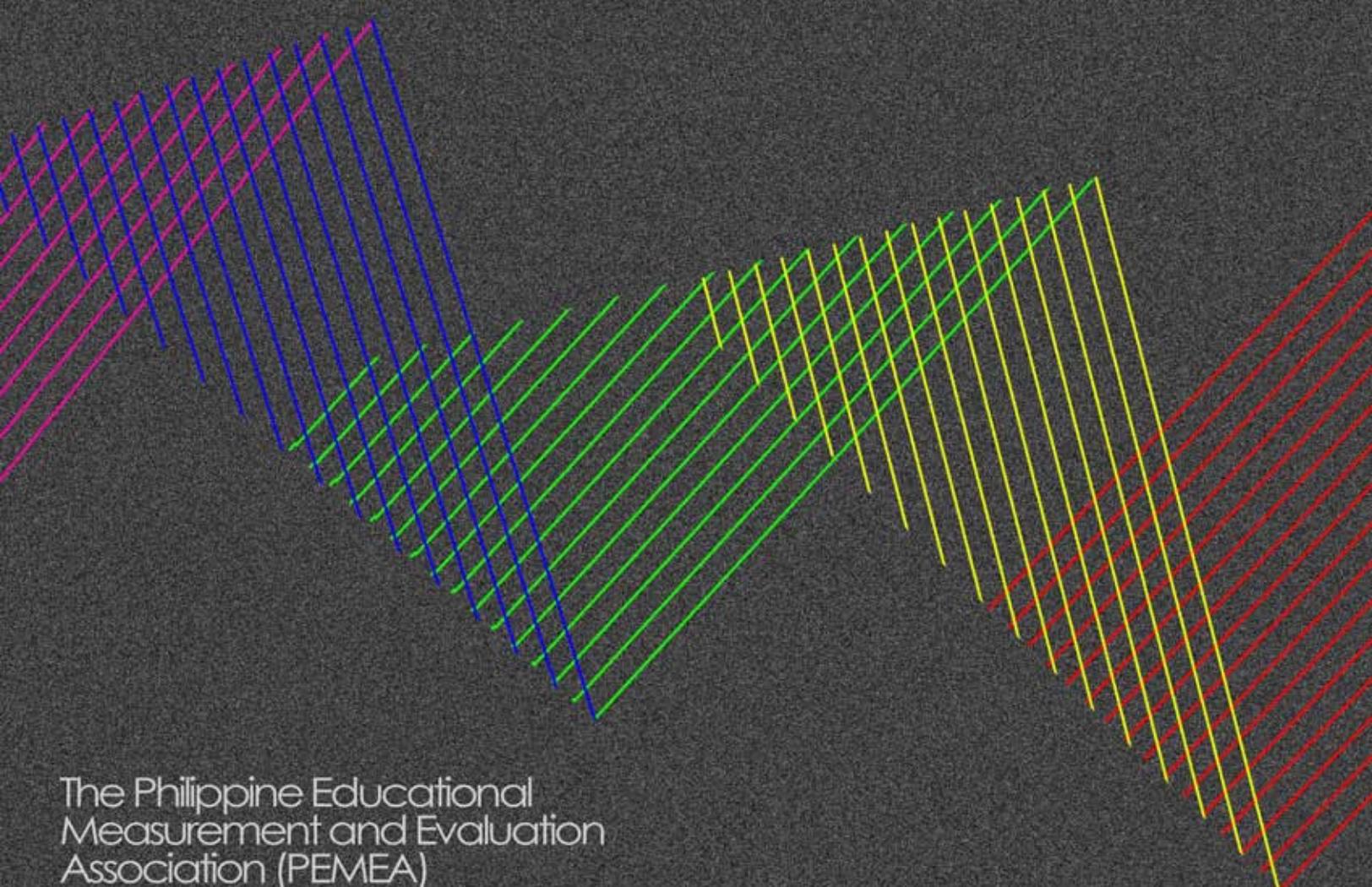


ISSN 2094 - 5876

The Educational Measurement and Evaluation Review

VOLUME 1 JULY 2010



The Philippine Educational
Measurement and Evaluation
Association (PEMEA)

The Educational Measurement and Evaluation Review is one of the official publications of the Philippine Educational Measurement and Evaluation Association (PEMMEA). The EMEReview publishes scholarly reports about contemporary theory and practice in the field of education and social science that highlights measurement, assessment, evaluation, psychometrics, psychological testing, and statistics. The journal is international, refereed, and abstracted. The journal is presently abstracted in the Asian Education Index, Social Science Research Network, Google Scholar, Open J-Gate, and NewJour.

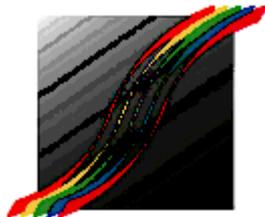
Copyright © 2010 by the Philippine Educational Measurement and Evaluation Association. Center for Learning and Performance Assessment, De La Salle-College of Saint Benilde, 2544 Taft Ave. Manila, Philippines

This journal is in copyright. Subject to statutory exception no reproduction of any part may take place without the written permission of the Philippine Educational Measurement and Evaluation Association.

No unauthorized photocopying

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior written permission of the Educational Measurement and Evaluation Review (EMEReview) educational board.

The articles in the EMEReview are open access at
<http://pemea.club.officelive.com/EMEReview.aspx>



Publication Division of PEMMEA
Philippine Educational Measurement and Evaluation Association

Educational Measurement and Evaluation Review

ISSN 2094-5876 <http://pemea.club.officelive.com/EMEReview.aspx> ♦ July 2010, Vol. 1

Editorial Guidelines of the Educational Measurement and Evaluation Review (EME Review) 1

Message for EMEReview
Richard Gonzales 3

Scientific directions in publication: An editorial note
Carlo Magno 4

Empirical Reports

The development and validation of the Emotionality in Teaching Scale (ETS)
Felicidad T. Villavicencio 6

The best practices of highly-rated college teachers
Pedrito A. Aton 24

The 2 x 2 achievement goal framework and intrinsic motivation among Filipino students: A validation study
Elmer dela Rosa 48

A confirmatory factor analysis of scores on Inventory of School Motivation (ISM), Sense of Self Scale, and Facilitating Conditions Questionnaire (FCQ): A study using a Philippine sample
Fraide A. Ganotice, Jr. 59

Development of Ethnic Identity Measure for Cordillera indigenous people
Cheryl F. Olvida 78

Examining the relationship of personality and burnout in college students: The role of academic motivation
Adonis P. David 90

A Short Form of the Subjective Well-Being Scale for Filipinos
Royce Hernandez 105

Determining the degree of inter-rater consistency in a high stakes, developing country, examination
Jack Holbrook 116

Short Reports

A brief history of educational assessment in the Philippines
Carlo Magno 140

An evaluation of the teaching-learning process: A transformative learning set-up
Violeta C. Valladolid 150

Organizing and reporting assessment results
Jennifer Ann L. Lajom 161

◆ **Editorial Board**

Rose Marie Salazar-Clemeña, *De La Salle University, Manila*

Milagros Ibe, *University of the Philippines, Diliman*

Alexa Abrenica, *De La Salle University, Manila*

Jack Holbrook, *University of Tartu, Estonia*

John Hattie, *The University of Auckland, Australia*

Anders Jonsson, *Malmö University, Sweden*

◆ **Managing Editors**

Carlo Magno, *De La Salle University, Manila*

Belen Chu, *Philippine Academy of Sakya*

Richard Gonzales, *Development Strategists International Consulting*

◆ **Editorial Staff for the First Volume**

Ma. Theresa Carmela E. Kanlapan

Copyeditor

Stephanie Dy

Cover Artist

◆ **Reviewers for this First Volume**

Royce Hernandez

Adonis David

Rita Ramos

Marife Mamauag

Carlo Magno

Belen Chu



Editorial Guidelines of the Educational Measurement and Evaluation Review (EMERreview)

Scope and Aim of the Journal

The Educational Measurement and Evaluation Review (EME Review) is one of the official publications of the Philippine Educational Measurement and Evaluation Association (PEMEA). It is international, refereed, and abstracted. The EME Review publishes scholarly reports about contemporary theory and practice in the field of education and social science that highlights measurement, assessment, and evaluation. It welcomes articles that are about test and scale development, quantitative models of a construct, evaluation studies, best practices in evaluation, issues in assessment, contemporary approaches in educational and psychological measurement, and other studies with direct implication to assessment in education, social science, and related fields.

Kinds of Scholarly Work Published

Scholarly and scientific reports can be in the form of the following:

1. Empirical reports such as studies in developing a construct, scale and test development, norming and IRT approaches.
2. Empirical essays such as those that deal with relevant issues and arguments in the practice of assessment and evaluation.
3. Literature reviews of scales or tests measuring a construct.
4. Metanalytic studies that synthesize the use of instruments and varied assessment methods.

Submission Process

The manuscript upon submission undergoes a blind review process. The editor in charge of the manuscript removes any identification of the author before sending it for review. The submitting author is notified upon receipt of the manuscript submitted. The author is also informed not to submit the manuscript to any journal for publication. The author waits for less than 2 months for the review.

The manuscript before submitted must be formatted to the latest edition of the Publication Manual of the APA. The manuscript should be single-spaced, use parenthetical citations, tables and figures are integrated within the text, and complete references should be found at the end of the report.

The manuscript is submitted on-line (through e-mail) at pemea2008@yahoo.com accompanying a cover letter attesting that the manuscript has not been published nor submitted to other journals for publication, and the overall work is original. The title page of the manuscript should indicate the complete name, affiliation, affiliation address, e-mail address, cell phone number, and telephone/fax number of the corresponding

author. In case there are multiple authors, one of them should be assigned as the corresponding author for the manuscript.

The Review Process

Stage 1: The manuscript is reviewed by the receiving editor for format and whether the requirements stated for submission are complete.

Stage 2: When the manuscript upon submission is complete and complied with all requirements, the identification of the author is removed and will be sent for blind review. The receiving editor keeps track of the identity of the author, timeline of the review, corresponds to the authors and reviewers, serve as a referee among authors and reviewers.

Stage 3: Upon completing the review, the editor sends the comments of all reviewers to the author.

Stage 4: The author upon receiving the manuscript should address all concerns provided in the review. Upon resubmission, the author writes back to the editor the details on how the comments were addressed and how the manuscript was improved. The author is given two to three weeks to resubmit the manuscript.

Stage 5. The editor reviews if all comments were addressed by the author. In case there are still concerns about the manuscript, it is sent back to the author. When all comments are addressed, the paper will be accepted for publication and the author received a formal notification.

Stage 6: When the manuscript is accepted for publication, the author signs a copyright transfer to the Publication Committee of the PEMEA.



Message for EMERreview

Richard DLC. Gonzales
Founding President and Chairman
PEMEA

The seminal idea of publishing the Educational Measurement and Evaluation Review (EMERreview) came about after the organization of the Philippine Educational Measurement and Evaluation Association (PEMEA) almost two years ago (2008). The idea was founded on the first theme of the National Conference of Education Measurement and Evaluation in 2008 where PEMEA was actually conceived and realised – “*Developing a Culture of Assessment in Learning Organizations.*”

As an initiative on its infancy stage, identifying members of editorial board and collecting and encouraging scholars and researchers to submit their papers in EMERreview were great challenges. However, with the determination to live up to the goals of PEMEA to be the primary organization in the country in promoting the best practices in educational measurement and evaluation and maintaining the highest standards of these practices and scholarship, this journal came to a reality.

It was a decision of the Managing Editors and the Board of Trustees of PEMEA to launch the maiden issue of EMERreview with both nationally and internationally known scholars and professors to be on its first batch of editorial board. Every effort has been made to carefully select papers to reflect the various dimensions and trends of educational measurement and evaluation in the Philippines as well as in other countries. We also gave much importance to the contributions of young and promising members of our organization. This way, we hope that scholarship and professional development are instilled among the members and trust that in the process, we influence others.

We are very grateful to our first set of Editorial Board and Managing Editors whose enthusiasm for taking the pains of going through the papers and putting them together gave us the courage to keep working on it.

We also thank the authors of this maiden issue for their contribution in propagating a culture of scholarship in educational measurement and evaluation.

We hope to sustain and further improve the quality and relevance of papers to be included in the future publications of this journal and we also encourage other scholars and professors to contribute their significant and novel studies to enrich the educational measurement and evaluation as a discipline.



Scientific directions in publication: An editorial note

Carlo Magno
PEMEA Board Member
EMERReview Managing Editor
De La Salle University, Manila

I would like to welcome the readers to this first volume of the Educational Measurement and Evaluation Review (EMERReview). This journal is one of the moves to accomplish the goals of the Philippine Educational Measurement and Evaluation Association (PEMEA), that is, to promote theory and practice of assessment in both international and local scene. The journal is specialized in the area of measurement, assessment, and evaluation studies that is committed to bring new theorizing and advances in the field of education, psychology, and other social sciences. Within five years, the editorial team envisions the journal to be abstracted in various indices that are known internationally including APA's PsycINFO and Thomson Reuters Information on Scientific Inquiry (ISI). This could be well-accomplished by promoting the journal and heightening the citations of the articles that are published in the journal.

In the process of conceptualizing and working on the journal's editorial guidelines, I ask myself what are the marks of a good journal? This can be answered by knowing the very nature of a scientific journal. A scientific journal is a periodical publication intended to further the progress of science, usually by reporting new research. The progress in science is made available in public through publishing scientific journals. A careful review process is needed to ensure the quality of studies published in the journal. People in the scientific community are conscious of the practices of peer-review and are open to the feedback provided by their peers (who is also within the same field). Providing quality studies can be controlled and several procedures can be ensured to produce good articles. The second part which involves the readers who needs to be enticed by reading and citing the articles in the journal is the less-controlled part. The journal needs to make an impact in the scientific community especially the ones within the fields of education and psychology in order to advance the knowledge presented in the journal. The articles in every issue of the journal are made open access at <http://pemea.club.officelive.com/EMERReview.aspx> in order to reach a wide variety of audience. Much can be accomplished in an open access movement: (1) Journals are made available to a larger audience making increasing the impact of the journal; (2) Possibility of abstracting in well-established indexing corpus; (3) Possibility of measuring the journals' impact by the number of times the authors were cited by others scholars.

In our effort to bring quality studies for the readership of the journal, 11 articles made it in the review process (I am proud to report that at least there are 20% rejection rates for the first volume). These articles were carefully selected to represent and provide a view about the current advancement of educational and psychological assessment in the Philippines. Authors from different regions in the country gave a glimpse on how they pursue scholarship in their lenses. The six articles by Villavicencio, Aton, Ganotice, Olvida, David, and Hernandez developed new scales and discovered new constructions of emotionality in teaching, best practices of teachers, ethnic identity, burnout, and well-being respectively. The deconstruction of these constructs reflects a typology that is appropriate for specific contexts where the study was conducted. The article by dela Rosa provided further evidence for a 2 X 2 framework of achievement goals instead of treating it with the traditional factors. Valladolid came up with an article showing the effects of a transformative type of learning in a higher education institution. Holbrook and Lajom provided useful practices in assessment. Specifically, Holbrook provided detailed guidelines for marking. Lajom also provided an overview on how assessment results are organized and reported. Magno, provided a brief account about the development of educational assessment in the Philippines.

The articles in this present volume covers a new way of looking at constructs, applicability of standards in higher education, call for improved practices in assessment, and a descriptive development of educational measurement in the Philippines. We are hoping that these articles will be of relevance to the readership of the journal. The works from experts and specialists in the field of educational and psychological measurement, assessment, and evaluation are invited to have their articles reviewed.



The development and validation of the Emotionality in Teaching Scale (ETS)

Felicidad T. Villavicencio

*Bulacan State University, Philippines
De La Salle University, Manila*

Abstract This study aims to develop and refine an instrument assessing teachers' emotions in teaching. The 120-item initial instrument was pilot tested from a sample of 547 teachers and the revised version was administered to a sample of 275 teachers. The dimensions of this construct were assessed through exploratory and confirmatory factor analyses. Confirmatory Factor Analysis was conducted to test the factor structure of the final version of the scale. The Emotionality in Teaching Scale (ETS) consists of five factors and 43 items which assess teachers' emotions in the context of teaching: Enjoyment, anger, boredom, guilt, and annoyance. The validity and reliability of this instrument in the two samples were found to be psychometrically adequate (Cronbach's $\alpha = .92$ and $.88$). This scale was designed to assess teachers' emotions that may influence motivation and teaching performance.

Keywords: *Emotions, teaching, emotionality in teaching*

Introduction

Teaching is an emotionally charged activity and changing the way teachers teach is extremely difficult (Scott & Sutton, 2009). As Nias (1996) noted, teaching involves human interaction and therefore has an emotional dimension. Emotions frequently shape cognitions, and some emotion theorists include changes in cognitive functioning as part of their definition of the emotion process (Mesquita, Frijda, & Scherer, 1997). Specifically, emotions can affect memory of teachers, their categorizing, thinking, and problem solving (Isen, 1993). Moreover, emotions may affect teachers' intrinsic motivation, attributions, efficacy beliefs, and goals. Negative emotions often reduce teachers' intrinsic motivation as "negative emotions tend to be incompatible with enjoyment as implied by interest and intrinsic motivation" (Pekrun, Goetz, Titz, & Perry, 2002, p. 97). Emotions may influence teachers' adoption of mastery versus performance, and approach versus avoidance goals (Kaplan, Gheen, & Midgley, 2002).

Although teachers may often hide their feelings, still students are often aware of and influenced by teachers' expression of both positive and negative emotions. Emotions may be expressed in several ways involuntarily and voluntarily (Sutton & Wheatley, 2003).

Literature suggests that empirical scales with good psychometric properties are not available particularly about teachers. Some available instruments that may be intended for teachers are: Maslach Burnout Inventory (MBI, Maslach & Jackson, 1981), Teacher Job Satisfaction Questionnaire (TJSQ, Lester, 1982), and Purdue Teacher Opinionnaire (PTO, Bentley & Rampel, 1980) but there is no existing instrument that would assess emotions of teachers. The Achievement Emotions Questionnaire (AEQ; Pekrun, Goetz, & Frenzel, 2005) however, was intended to assess students' emotions. Thus, the purpose of this study was to construct and validate appropriate scales with which to measure this important construct. The scales measuring emotionality in teaching developed in this study represent a contribution to the literature on academic motivation in general and to emotions in particular. The emotional experiences of teachers are of paramount importance to teachers themselves and students with whom they interact.

Dimensions of Emotions

Frijda (2004) considered the kinds of emotion to represent discrete emotion categories. They are often referred to as basic emotions (Izard, 1977). A typical set of such categories is the one found by Shaver, Wu, and Schwartz (1992): Anger, disgust, sadness, fear, joy, surprise, love, and pity. The sets of basic emotions found by such methods show considerable similarity between studies although languages do, however, exist, and some words (e.g., hope and jealousy) shift greatly from one cluster to another, and do not really belong to any one of them.

Mesquita and Frijda (1992) reviewed the psychological and anthropological literature on cultural variations in emotions. The literature has been interpreted within the framework of a cognitive-process model of emotions. They found that cross-cultural differences or similarities depend on an important degree on the level of description of the emotional phenomena. Furthermore, cultural differences in emotions appear to be due to differences in event types or schemas, in culture-specific appraisal propensities, in behaviour repertoires, or in regulation processes. Differences in taxonomies of emotion words sometimes reflect true emotion differences, but they may also just result from differences in which emotion-process phase serves as the basis for categorization.

Lutz (1988) argued that although most emotions are viewed as universally experienced "natural" human phenomena, emotions are anything but natural. Emotion, she contended, "can be viewed as cultural and interpersonal products of naming, justifying, and persuading by people in relationship to each other. Emotional meaning is then a social rather than an individual achievement—an emergent product of social life" (Lutz, 1988, p. 5).

As discussed (Scott & Sutton, 2009), many measurements of emotion are based on research in social psychology and rely on experiments with college and university students that may not be applicable to teachers.

Teachers' Emotionality

Emotionality signals that something has transpired that is important to the individual's self or self system (Markus & Wurf, 1987). More specifically, all emotionality experienced by teachers in their classroom has reference to their primary self and to their occupational selves – the current teaching self, and their "ought" and "ought-not" teaching selves (Higgins, 1987). As discussed, it is through the many facets of these changing psychological structures of selves that the individuals manifest cognitive and affective dimensions in their life. Thus, they claimed that teachers order their teaching lives in regard to who and what they are, their organized selves, and in regard to activities and feelings proper, and improper, for such selves. Thus, centrality of selves in all emotional experience implies that if we want to understand emotionality in teaching, we must understand the conditions of teacher's selves.

What kinds of conditions trigger teachers' emotionality? Previous psychological research (e.g., Frijda, 1993) identifies three major dimensions of emotionality: (1) Emotions proper, as reactions to particular, non-routine events; (2) emotions of relations, that is, sentiments of liking-hating, trusting-mistrusting; and (3) moods, like pessimism or optimism about what or where one is in life. Aligned with the foregoing discussion, this study considered these dimensions in the initial survey with teachers, which was conducted through the narratives of their emotional experiences.

Sutton and Wheatley (2003) reviewed the limited literature on the emotional aspects of teachers' lives. They described a multi-componential perspective on emotions and critiqued the existing literature on teachers' positive and negative emotions. The summary of literature suggests that, emotions influence teachers' and students' cognitions, motivation, and behaviours. Moreover, they recognize that emotions are an integral part of teachers' lives. Along with motivation and cognition, psychologists now recognize emotion as one of three fundamental classes of mental operations (Mayer, Salovey, & Caruso, 2000). Therefore, knowledge of teachers' emotions is essential in understanding teachers and teaching. Sutton and Wheatley (2003) also described the positive (e.g., love and caring, joy, satisfaction, pleasure, pride, excitement) and negative emotions (e.g., anger, frustration, anxiety, guilt, helplessness, and sadness) that teachers generally experience.

Recently, Scott and Sutton (2009) tested the previously-mentioned theories of how emotions relate to changes in practice and found that neither positive nor negative emotions were related to these changes. However, according to Pekrun (2006), emotions are to be viewed as involving sets of interrelated psychological processes (affective, cognitive, motivational, and physiological component processes).

Occurrence and Structure of Academic Emotions

In a series of exploratory studies, Pekrun and colleagues (2002) analyzed the emotions of school and university students by means of qualitative interviews and questionnaires. They used exploratory findings as an empirical basis for constructing theoretical taxonomies of the internal structures of different academic emotions. Their participants reported a broad range of affective, cognitive, physiological and motivational elements of their emotional experiences that were used for creating these taxonomies (Pekrun et. al., 2002, p. 94). As discussed, academic emotions are defined as emotions that are directly linked to classroom instruction, academic learning, tests and exams, and students' achievement in these academic situations (Pekrun, 2006; Pekrun, et. al., 2002).

Whereas, quantitative measures are needed to emotions, however, they are tailored to measurement of student emotions like the Achievement Emotions questionnaire (Pekrun, Goetz, & Frenzel, 2005). Minimal research has examined teacher emotion within the framework of an explicit theory of emotion (Sutton & Wheatley, 2003).

In Pekrun's (1992) model, discrete academic emotions are assumed to have specific effects on learning and achievement based on how they are classified within the conceptual schema. This model distinguishes between emotions that are positive-activating (enjoyment, pride, hope), positive-deactivating (relief, relaxation), negative activating (anxiety, anger, shame/guilt), and negative deactivating (boredom, hopelessness, disappointment). In addition to the valence and activation, academic emotions are also categorized according to the aspects of the situation (task- and self-related vs. social). The task-related emotions are further separated into process-related, prospective, and retrospective emotions (Goetz, Zirngibl, Pekrun, & Hall, 2003).

Anchored on Pekrun's theory (1992; 2006) and Sutton and Wheatley (2003) research (2003), the Emotionality in Teaching Scale has a valence of positive and negative emotions. Positive emotions consist of enjoyment, and pride which are activating. Negative emotions consist of irritation, anger, shame, guilt, annoyance (activating), and boredom (deactivating).

The objective of this paper was to develop a multi-item scale of self-report on emotionality in teaching as perceived by the teachers themselves. Defining emotions, distinguishing them from other states or traits, and more so measuring emotions in a comprehensive and meaningful way are indeed very challenging for researchers over a long period of time (Scherer, 2005). Without consensual conceptualization of exactly what phenomenon is to be studied, measurement of emotionality in teaching could hardly be realized.

Teachers' Emotions and Teaching

When asked, teachers usually share about the joy, satisfaction, and pleasure associated with teaching (Sutton, & Wheatley, 2003). Teachers' satisfaction occurs when students learn and make progress (Sutton, 2000; Hargreaves, 1998). They also find pleasure when their students strive to learn. In addition, teachers find joy

when students are responsive in class. They also find pleasure and pride when students cooperate with no major disruptions (Emmer, 1994; Sutton, 2000). In general, teachers experience positive emotions when they get everything done (Hatch, 1993).

Teachers also reported experiencing negative emotions. Every middle school teacher interviewed by Sutton (2000) talked about frustration or anger or both. As found in Sutton's study, frustration and anger arise from a number of sources related to goal incongruence. These include students' misbehavior and violation of rules (Emmer, 1994; Erb, 2002; Hargreaves, 2000; Sutton, 2000). Teachers are also likely to become angry when they believe that students' poor academic work is due to controllable factors such as laziness or inattention (Reyna & Weiner, 2001). Moreover, anger and frustration are exacerbated by exhaustion and stress (La Porte, 1996; Sutton, 2000) and losing one's temper can make teachers feel ashamed (Lortie, 1975). Interestingly however, some teachers consciously use "fake anger" to manage their students (Woods & Jeffrey, 1996), and some teachers report learning to control their anger with experience (Golby, 1996; Sutton & Conway, 2002).

Furthermore, students' misbehavior that elicits negative emotions in teachers is distracting and diverts attention from instructional goals. High anxiety can reduce limited resources of working memory because it impairs task-relevant processing (Ashkraft & Kirk, 2001). In contrast to the negative effects of anxiety, positive emotions of joy, interest, and pride may broaden the memory capacity, that is, more thoughts and actions come to mind (Fredrickson & Branigan, 2001). This suggests that teachers who experience more positive emotions may generate more teaching ideas and strategies.

The present study was conceived based on the literature reviewed so far which provides an overview of the range of emotions that teachers experience and the conditions under which these emotions are experienced.

Method

Establishing Construct Definition and Content Domain

The content validity is focused in line with its theoretical dimensionality. Literature review clearly identifies the most appropriate definition and theory of emotions, specifically the emotions associated in teaching. A survey was conducted for domain sampling to be able to establish theoretical assumptions about items for emotionality in teaching. Teachers were asked to report their emotions both positive and negative, activating or deactivating in teaching. They were also asked to specify the situations which led them to experience enjoyment, anger, irritation, as well as boredom and annoyance. A total of 154 initial items was generated from the participants.

Generating and Judging Measurement Items

The 154 original items were rephrased in such a way that it will fit the response format. Likert five-point scale was used (1 as strongly disagree to 5 strongly agree). The initial item pool was reviewed by experts to determine if the phenomenon of interest has generated a pool of suitable items and an appropriate response format for these items has been selected. Item critiquing was performed by three experts to further establish its content validity. They were provided with the working definition of emotionality in teaching. The three experts were each provided a copy of the Emotionality in Teaching Scale consisting of 154 items, grouped into five subscales: subscale 1-enjoyment, enthusiasm, fulfilment (43 items); subscale 2- anger, irritation (30 items); subscale 3: boredom, annoyance (25 items); subscale 4- pride (28 items); and subscale 5- shame, guilt, embarrassment (28 items). The items were evaluated by the experts because they were asked to make a decision if the item is to be accepted, rejected, and revised. Some items were eliminated based on a priori criteria, such as lack of clarity, questionable relevance, or undesirable similarity to other items (DeVellis, 1991). A few insightful comments about why certain items are ambiguous, clarified on how items measure the construct. After incorporating the comments given by the experts, 34 items were eliminated. Thus, the *Emotionality in Teaching Scale* consists of 120 items for pilot testing.

Pilot testing

A preliminary form of the ETS consisting of 120 items was administered to a sample of 547 teachers. Teacher participants came from diverse populations. Data were cross-sectional in nature with 414 female and 133 male teachers ranging from 21 to 64 years old ($M = 34.5$; $SD = 9.79$). They were teaching in public ($N=382$) and private ($N=165$) schools in different levels (grade school, high school, college, and graduate school in Manila, Bulacan, Pampanga, and other provinces). Pilot testing was done as an item-trimming procedure and to establish its psychometric properties.

Data Analysis

The data obtained from pilot testing were subjected to a factor analysis using principal factor analysis with iteration and an orthogonal (varimax) rotation. A set of selection criteria was then applied to the items to determine which items would be retained. Items were retained that met all of the following criteria: A factor loading greater than .40 on only one of the five factors, and a high item-total correlation. Aside from eigenvalues, scree plot was used to determine the number of factors that will be retained.

The psychometric properties of the scale were determined. Initial item analyses were conducted via exploratory factor analysis (EFA), specifically, the principal components analysis. The EFA retained 45 items with five scales. The internal consistency of the scales was also determined using Cronbach's alpha.

Confirmatory factor analysis (CFA) was conducted with structural equation modelling (SEM) packages that use Maximum Likelihood Estimation (MLE). MLE involves an iterative process in which the observed covariance matrix (among the items) is compared with an implied or theoretical matrix to minimize the differences (or residuals) between the observed and implied matrices. The SEM package iterates until these differences are minimized. When the differences can no longer be reduced further, convergence occurs. Thus, an initial step in model evaluation is to determine if the CFA solution converges without any offending estimates (out-of-range parameter estimates), and to assess the model fit.

Validation of the Scale

In order to obtain confirmatory data for the pattern of factors, the 45-item form was administered to graduate students who were teaching in several schools in Bulacan and Pampanga. Two hundred seventy-five teachers (275) participated, 245 were female, and 30 were male (M age = 33.64; SD = 9.65). These teachers came from public schools (156) and private schools (119) handling pre-school, grade school, high school, and college students.

Results

Descriptive statistics

Means, standard deviations, minimum and maximum values, skewness, and kurtosis values were determined. Examination of the data shows that there were 39 items with poor skewness and kurtosis values. All values greater than 2 were already removed prior to factor analysis, with 81 items left. Five factors emerged on the basis of item pool generation when the 154 items were categorized resulting to five underlying factors.

Exploratory Factor Analysis

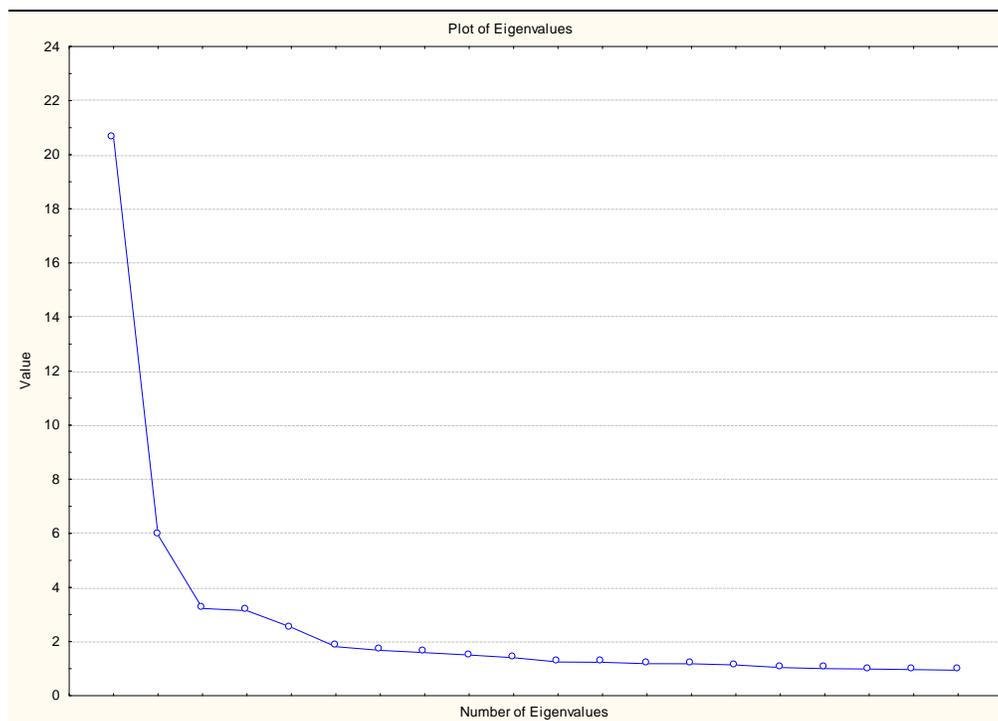
Exploratory factor analysis was conducted in Statistica 7 software. Table 1 presents the eigenvalues of five factors extracted from the scale. Determining the number of factors to retain in any EFA is a critical decision (Thompson, 2004). According to the eigenvalue-greater-than-one rule, the number of factors is equal to the number of eigenvalues greater than one (Netemeyer, Bearden, & Sharma, 2003). If this was observed, over-extraction is possible, wherein various items might load across several factors. Thus, in the present study however, a more robust eigenvalues of 2 and above was considered. After performing the varimax rotation, five factors were extracted. A coefficient value of .40 was used as the cut-off score. Items with coefficient value below .40 were eliminated. The initial five factors extracted consist of: Factor 1 (18 items), Factor 2 (12 items), Factor 3 (9 items), Factor 4 (11 items), and Factor 5 (10 items). A total of 60 items remained after deleting 21 cross-loading items. However, 15 items were eliminated because they seem not to be grouped on the five factors.

Table 1
Eigenvalues of Five Factors Extracted from the Scale

	Eigenvalue	Percent Total	Cumulative Eigenvalue	Cumulative Percent Total
1	24.53	20.45	24.53	20.45
2	14.85	12.37	39.38	32.82
3	3.97	3.31	43.35	36.13
4	3.36	2.80	46.71	38.93
5	2.94	2.45	49.65	41.38

According to Thompson (2004), determining the number of factors to extract or retain requires judgment. Scree plot of eigenvalues is illustrated in Figure 1. Although the elbow of the scree plot suggests 3, 4 and 5 factors, 5 factors was considered because deleting a factor well below this elbow will show little loss of explained variance. Factor 1 refers to emotions of teachers related to irritation and anger (12 items). Factor 2 refers to emotions related to pride and enjoyment (12 items). Factor 3 refers to emotions of guilt and shame (9 items). Factor 4 refers to boredom in teaching (8 items). Factor 5 refers to annoyance experienced by teachers (4 items). The five-factor solution for a factor analysis of the 45 items, based on the sample of 547 and using principal factor analysis with iteration plus an orthogonal rotation, is presented in Appendix.

Figure 1
Scree Plot of Eigenvalues for Five Factors



Reliability of the Scale

Cronbach's alpha (α) reliability of the ETS was determined in the initial version of 45 items. The reliability of each factor was also determined. The reliability coefficient suggests that the scale and the five subscales have very good internal consistency. Summary statistics for each scale is shown in Table 2.

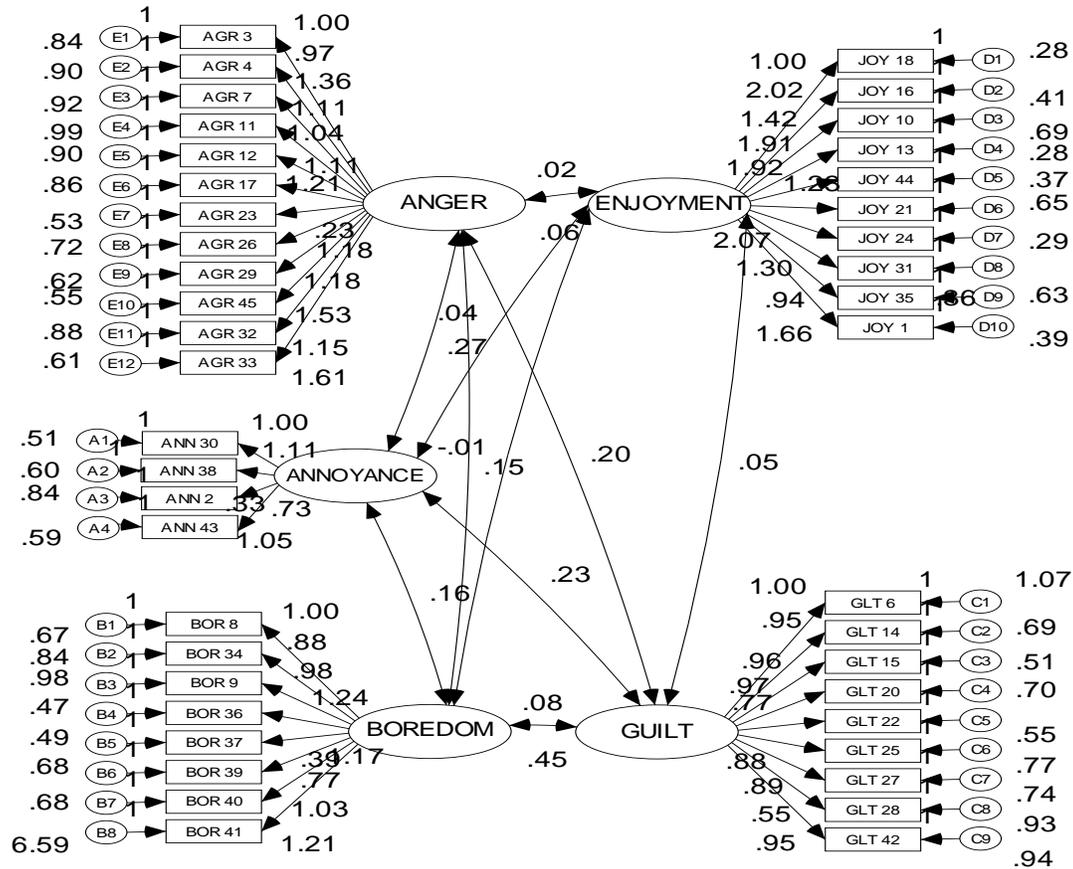
Table 2
Descriptive Statistics of the 45-Item Scale

Try-out Version of the Scale	No. of items	<i>M</i>	<i>SD</i>	Cronbach's alpha reliability
Factor 1 (Irritation and Anger)	12	44.20	9.38	.92
Factor 2 (Pride and Enjoyment)	10	52.42	5.49	.86
Factor 3 (Guilt and Shame)	9	32.64	6.39	.84
Factor 4 (Boredom)	8	22.00	5.79	.78
Factor 5 (Annoyance)	4	15.02	3.21	.75
ETS (Whole Scale)	45	166.29	20.94	.92

Confirmatory Factor Analysis (CFA)

The validity of the five factors in ETS was examined through confirmatory factor analysis. The measurement model that was tested in the study is presented in Figure 2. In this model, there were five latent constructs (anger, enjoyment, guilt, boredom, and annoyance) and the manifest variables which are the items. The parameter estimates show that the five factors of the scale are interrelated and the items which constitute each factor. The items (manifest variables) were specified as indicators for the five latent variables. Consequently, these five latent factors contribute to emotionality associated in teaching.

Figure 2
Measurement Model of the Study



Results of confirmatory factor analysis (CFA) indicate that the solution has converged normally with the following fit indices: ($\chi^2 = 1862.53$; $df = 935$; $\chi^2/df = 1.99$) In addition, the scale has a Steiger-Lind Root Mean Square Error approximation (RMSEA) index of .06 which suggests an acceptable fit. RMSEA is considered as a stand-alone or absolute measure of fit (Netemeyer, et al., 2003).

Validation of the Emotionality in Teaching Scale

The 45-item ETS was validated by administering to a sample of 275 teachers. The reliability of the whole scale in its second administration shows a Cronbach’s alpha reliability of .88 with subscale reliabilities of .66 to .84. Two enjoyment items were eliminated due to its insignificant correlation with other items. Thus, the validated version of the Emotionality in Teaching Scale consists of 43 items. Table 3 shows the descriptive statistics of the final version of the scale.

Table 3
Descriptive Statistics of the Final Version of the ETS

Final Version of the Scale	No. of items	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	Cronbach's alpha reliability
Factor 1: Irritation and Anger	12	42.22	7.56	-.47	-.07	.83
Factor 2: Pride and Enjoyment	10	41.72	4.45	-.41	-.42	.77
Factor 3: Guilt and Shame	9	31.29	5.92	-.42	.26	.80
Factor 4: Boredom	8	19.91	6.26	1.72	1.57	.71
Factor 5: Annoyance	4	14.38	2.75	-.29	.10	.66
ETS (Whole Scale)	43	149.53	18.29	-.21	.29	.88

Discussion

The results are consistent with theory (Hargreaves, 2001; Sutton & Wheatley, 2003) which suggest that teachers tend to experience different emotions. Moreover, these studies shed light on the nature of the distinction between positive and negative emotions although only one factor corresponding to positive emotions emerged as compared to four factors of negative emotions. The five factors extracted in the analyses further imply that emotionality in teaching is a multidimensional construct.

The items belonging to Factor 1 pertain to irritation and anger generally experienced by teachers. Items in this scale suggest that teachers get angry (e.g., “I get angry when a student cheats during exams”) and get irritated (e.g., I feel irritated when students do not pay attention to the discussion”) in relation to students’ behavior. Aligned with an earlier study, teachers are likely to become angry when they believe that students’ poor academic work is due to controllable factors such as laziness or inattention (Reyna & Weiner, 2001). As claimed by Hargreaves (1998; 2000; 2001), anger is related to teachers’ goals. Teachers tend to get angry when they see that their goals are not attained. Specifically, students who are not prepared, provoke a teacher to anger. In line with Stein, Trabasso, and Liwag (2000), when teachers fail to attain or maintain desired goal in teaching, or when they cannot prevent undesired goal from occurring or continuing, they experience negative emotions like anger or irritation (e.g., “I get irritated when no one knows the correct answer”).

On the other hand, items in Factor 2 pertain to positive emotions of enjoyment and pride. Based on these items, the conditions under which teachers often experience positive emotions are related to their performance of duties (e.g., “I am proud of my ability to design lessons that are relevant, meaningful, and fun”) and to student performance (e.g., “I appreciate students who monitor their own thinking during class discussion”). These confirm earlier findings (Sutton, 2003) that teachers tend to experience positive emotions when they see students make progress and students who are responsive in class. Consistent with Lewis (2000),

teachers' enjoyment in teaching is elicited by specific situations identified by teachers themselves. Similarly, pride is elicited depending on cognitions related to the self (Lewis, 2000). Specifically, pride occurs when one makes a comparison or evaluate one's behavior vis-à-vis some standard, rule, or goal and finds that one has succeeded (e.g., "I am proud of my ability to integrate technology in the classroom"). Teachers experience pride in their accomplishments, the pride in being successful in fulfilling a particular goal or activity. These conditions trigger enjoyment and pride in teaching.

Factor 3 consists of emotional experiences of guilt. An examination of the items suggests that guilt experiences of teachers are associated with their performance (e.g., "I feel guilty if I don't achieve all the objectives of the course."). As discussed by Lewis (2000), guilt occurs when an individual's evaluation leads to the conclusion that one has failed. The determination of success and failure in teaching resides within the teacher and on the nature of the standard that is set. Guilt experienced by teachers is not guilt related to the whole self, but rather guilt related to one's action (e.g., "I feel guilty when majority of my students failed in the exam").

Factor 4 refers to boredom experienced by teachers which may be task-related (e.g., "Constructing tests/exams is boring for me") or student-related ("I feel bored in a class of passive students"). Clearly, teachers' boredom is caused by repetitive or routine work. Thus, in teaching, teachers also carry a value judgment. Boredom is elicited when teachers engage in activity for some time that are no longer interesting and fulfilling for them.

Factor 5 pertains to teachers' feeling of annoyance which can be attributed to students (e.g., "I am annoyed when no one could answer my question in class"). Annoyance is evoked when teachers appraise an event as unpleasant, difficult, or they are unable to control the situation. Moreover, teachers get annoyed when they fail in their expectations (e.g., "I am annoyed when students did not do their assignments").

Theoretically, teachers experience a diversity of emotions in teaching. To date, there has been no instrument about emotionality in teaching that was developed which utilized qualitative and quantitative methods in its development. In this research, qualitative method was employed in exploring the occurrence and structure of emotions associated to teaching. Then, a quantitative self-report instrument measuring five emotions experienced by teachers in the context of teaching was developed and validated.

The scale developed in this study has strong empirical qualities which represents a contribution to the assessment and measurement literature specifically emotions associated in teaching. Both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted to determine and confirm the factorial stability of the Emotionality in Teaching Scale (ETS). The five emotions in this scale were stringed together from exploratory factor analysis. Items in the final version of the scale were based on item statistics, reliabilities, and measures of fit obtained from confirmatory factor analysis. The resulting scale integrates primary emotions of enjoyment and anger, self-conscious emotions of pride and shame, as well as less frequently reported emotion of annoyance.

Teaching is an emotionally intense experience encompassing the full gamut from enjoyment to anger. These two emotions have received considerable attention (Lewis, 2000). However, shame, boredom, and annoyance have received little attention. Thus, ETS captures this string of emotions generally experienced in teaching. Emotions of students are categorized according to valence and activation (Pekrun et. al., 2002). In ETS however, emotions are categorized according to valence: whether the emotion is positive or negative, and agency: whether the teacher or the student or both are responsible for the emotion. According to appraisal theorists, emotions are evoked by events that are relevant to the individual's concerns, motives, major goals, or well being (Frijda, 1993; Lazarus, 1991). The five emotions contained in ETS focused on the personal significance and meaning of events experienced by teachers in everyday teaching.

In the Achievement Emotions Questionnaire for students (Pekrun et. al., 2005), emotions are measured in domain-specific ways and was designed to assess college students' academic emotions. In the ETS however, emotions of teachers across domains and levels (grade school, high school, college, and graduate school) can be assessed. The items for each emotion represent the wide range of emotions generally experienced by teachers in academic setting, both pleasant and unpleasant.

A five-factor solution consisting of five different but interrelated emotions in teaching best accounted for the data. The factors had eigenvalues greater than two and are considered subscales of the ETS. The irritation and anger subscale consists of 12 items. The 10 items in the pride and enjoyment subscale describe the positive emotions experienced by teachers. The guilt and shame subscale consists of 9 items, boredom scale has 8 items and annoyance scale has 4 items. For pride and enjoyment subscale, higher mean scores correspond to higher levels of positive emotional experiences. In contrast, higher mean scores of irritation and anger, guilt and shame, boredom and annoyance suggest high levels of negative emotions of teachers. Such a correlation is in accord with theoretical expectations that there are positive and negative dimensions of teachers' emotionality. The five subscales showed good internal reliabilities.

The development and validation of the Emotionality in Teaching Scale was based on the need for an instrument to assess emotionality associated with teaching. Since there is no existing instrument related to this construct, this study provides initial empirical evidence on how emotions of teachers associated with teaching can be assessed. Future studies may allow us to achieve a better understanding of the pleasant and unpleasant emotions commonly experienced by teachers. This can be adopted by future researchers in investigating other motivational constructs or any factors related to teaching performance. An understanding of teachers' emotions in teaching may help future research examine the links between teachers' emotions and their choice of teaching strategies (Hargreaves, 2000); their emotions and goal orientation, as well as their classroom management and discipline, and teaching performance.

In summary, the purpose of this research was to develop and validate the psychometric properties of the ETS among teachers. Taken together, the results represent further evidence that the ETS and its five subscales have adequate levels

of reliability and validity in two samples. Overall, the results of the present study suggest that emotionality in teaching can now be assessed using the present scale.

References

- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224-237.
- Bentley, R. R., & Rampel, A. M. (1980). *Manual for the Purdue Teacher Opinionnaire*. West Lafayette, IN: Purdue Research Foundation.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. California: Sage Publications.
- Emmer, E. T. (1994, April). *Teacher emotions and classroom management*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Erb, C. S. (2002, May). *The emotional whirlpool of beginning teachers' work*. Paper presented at the annual meeting of the Canadian Society of Studies in Education, Toronto, Canada.
- Fredrickson, B. L., & Branigan, C. A. (2001). Positive emotions. In Mayne, T. J., & Bonnano, G. A. (Eds.), *Emotions: Current Issues and Future Directions* (pp. 123-151). Guilford Press, New York.
- Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition and Emotion, 7*, 357-387.
- Frijda, N. H. (2004). Emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 272-288). New York: Guilford Press.
- Goetz, T., Zirngibl, A., Pekrun, R., & Hall, N. (2003). Emotions, learning and achievement from an educational-psychological perspective. In P. Mayring & C. V. Rhoeneck (Eds.), *Learning emotions: The influence of affective factors on classroom learning* (pp. 9-28). Frankfurt am Main, Germany: Peter Lang.
- Golby, M. (1996). Teachers' emotions: An illustrated discussion. *Cambridge Journal of Education, 26*, 423-435.
- Hargreaves, A. (1998). The emotional politics of teaching and teacher development: With implications for educational leadership. *International Journal of Leadership Education, 1*, 315-336.
- Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teacher and Teacher Education, 16*, 811-826.
- Hargreaves, A. (2001). The emotional geographies of teaching. *Teachers' College Record, 103*, 1056-1080.
- Hatch, J. A. (1993). Passing along teacher beliefs: A good day is ... *Education Horizons, 71*, 109-112.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319-340.
- Isen, A. M. (1993). Positive affect and decision making. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 261-277). New York: Guilford Press.

- Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.
- Kaplan, A., Gheen, M., & Midgley, C. (2002). Classroom goal structure and student disruptive behavior. *British Journal of Educational Psychology*, *72*, 191-211.
- Lazarus, R.S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, *8*, 819-834.
- La Porte, E. (1996). Teaching: Getting it right. *Cambridge Journal of Education*, *26*, 361-366.
- Lester, P. E. (1982). *Teacher job satisfaction questionnaire*. Long Island University. Brookville: New York.
- Lewis, M. (2000). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis & A. Haviland (Eds.), *Handbook of emotions* (pp. 623-636). New York: Guilford Press.
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press
- Lutz, C. (1988). *Unnatural emotions: Everyday sentiments on a Micronesian atoll and their challenge to Western theory*. University of Chicago Press: Chicago.
- Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*, 299-337.
- Maslach, C., & Jackson, S.E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior*, *2*, 99-113.
- Mayer, J. D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 396-420). New York: Cambridge University Press.
- Mesquita, B., & Frijida, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112*, 179-204.
- Mesquita, B., Frijida, N. H., & Scherer, K. R. (1997). Culture and emotion. In J. W. Berry., P. R., Dasen, & T. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology: Basic processes and human development* (pp. 255-29). Boston: Allyn and Bacon.
- Netemeyer, R. G., Bearden, W.O., & Sharma, S. (2003). *Scaling procedures. Issues and applications*. California: Sage Publications.
- Nias, J. (1996). Thinking about feeling: The emotions in teaching. *Cambridge Journal of Education*, *26*, 293-306.
- Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive motivational mediators. *Applied Psychology: An International Review*, *41*, 359-376.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91-105.
- Pekrun, R., Goetz, T., & Frenzel, A.C. (2005). *Academic Emotions Questionnaire-Mathematics (AEQ-M) User's manual*. Department of Psychology, University of Munich.

- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315-341.
- Reyna, C., & Weiner, B. (2001). Justice and utility in the classroom: An attributional analysis of the goals of teachers' punishment and intervention strategies. *Journal of Educational Psychology, 93*, 309-319.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 695-729.
- Scott, C., & Sutton, R. (2009). Emotions and change during professional development for teachers: A mixed methods study. *Journal of Mixed Methods Research, 3*, 151-171.
- Shaver, P., Wu, S., & Schwartz, J. C. (1992). Cross-cultural similarities and differences in emotion and its representation: A prototype approach. In M. Clark (Ed.), *Review of personality and social psychology* (pp. 175-212). Beverly Hills: Sage.
- Stein, N. L., Trabasso, T., & Liwag, M. D. (2000). A goal appraisal theory of emotional understanding: Implications for development and learning. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 436-455). New York: Guilford Press.
- Sutton, R. E. (2000, April). *The emotional experiences of teachers*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Sutton, R. E., & Conway, P. (2002, April). *Middle school teachers' day-to-day experiences of anger and frustration*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review, 15*, 327-329.
- Thompson, B. (2004). Exploratory and confirmatory factor analysis: understanding concepts and applications. Washington DC: American Psychological Association.
- Woods, P., & Jeffrey, B. (1996). *Teachable moments: The art of teaching in primary schools*. Buckingham, England: Open University Press.

Appendix

Factor Loadings of 45 items

	Scale items	Factor 1 Irritation and anger	Factor 2 Pride and enjoyment	Factor 3 Guilt and shame	Factor 4 boredom	Factor 5 annoyance
1	I feel irritated when a student does not listen during discussion and asks questions later because he does not understand.	.71				
2	I feel irritated when my students are talking with their seatmates while I'm discussing.	.77				
3	I feel irritated when students are making noise.	.76				
4	I feel irritated when students ask questions they already know the answer to.	.55				
5	I feel irritated when students do not pay attention to the discussion.	.77				
6	I feel irritated whenever a student asks irrelevant questions.	.57				
7	I get angry if my student is doing something else in class.	.72				
8	I get angry when a student arrives late in class.	.67				
9	I get angry when a student cheats during exams.	.68				
10	I get angry when a student sleeps in class.	.62				
11	I get disappointed when no one knows the correct answer.	.52				
12	I get irritated when students ask questions that I have just answered.	.70				
13	I am proud of my ability and skills in teaching	.	.78			
14	I am proud of my ability to design lessons that are relevant, meaningful, and fun.		.77			
15	I am proud of my ability to incorporate various strategies to address the needs of all students and different modalities.		.76			
16	I am proud of my ability to integrate technology in the classroom.		.71			
17	I am satisfied with the result of student evaluation.		.45			
18	I appreciate students who monitor their own thinking during class discussion.		.55			
19	I enjoy checking the course requirements submitted by my students especially when they did it well.		.46			
20	I enjoy preparing my lesson plan or class activity.		.45			
21	I enjoy talking to intelligent students in my class.		.45			
22	I feel happy when someone compliments my effort to teach.		.48			
23	I think I can be proud of my outstanding teaching performance.		.69			
24	I am so proud that I can teach with maximum proficiency.		.73			
25	I feel ashamed whenever I realize that my classroom is far from ideal.			.43		
26	I feel defeated when a student consistently gets a failing score in the test.			.43		
27	I feel embarrassed when I call my student with another's name.			.46		
28	I feel embarrassed when I could not answer a question of my students.			.49		
29	I feel guilty if I don't achieve all the objectives of the course.			.70		
30	I feel guilty when I arrive late in class.			.78		
31	I feel guilty when I did not perform well in my teaching.			.77		
32	I feel guilty when I give a failing course grade.			.51		
33	I feel guilty when majority of my students failed in the exam.			.59		
34	Constructing tests/exams is boring for me.				.45	

35	I am bored during test administration.	.46
36	I am bored talking to students who communicate poorly.	.52
37	I feel bored in a class of passive students.	.46
38	I feel bored when dealing with difficult students.	.54
39	I feel impatient when students refuse to accept a scientifically proven concept.	.57
40	I feel tired explaining a simple subject matter but too difficult for my students.	.56
41	I get bored teaching the same subject (course) for years.	.46
42	I am annoyed when a student is consistently taking my subject for granted despite constant reminders.	.57
43	I am annoyed when a student keeps on committing the same mistake.	.64
44	I am annoyed when my students did not do their assignments.	.59
45	I am annoyed when no one could answer my question in class.	.49

Author Note

For questions and permission to use the ETS please address correspondence to Dr. Fely T. Villavicencio, e-mail: dr_fely@yahoo.com



The best practices of highly-rated college teachers

Pedrito A. Aton

Trinity University of Asia, Philippines

Abstract This study has a threefold objective: (1) To identify the best practices of the highly-rated college teachers, (2) to determine the domains of those practices, and (3) to come-up with a new scale to measure teacher effectiveness. A questionnaire composed of 95 attributes and strategies of effective teachers which adapted from several studies and it was distributed to highly-rated college teachers and their respective students in a Catholic University in Southern Mindanao. From the 125 accomplished questionnaires, 51 common attributes and strategies of highly-rated college teachers were identified. The top 25 are considered the best practices considering the mean rating and the variability per item. The leading best practices include teacher confidence, respect for individual student, clarity and organization, encouragement of students' participation, and good communication skill. The data gathered which were factor analyzed yield four factors -loading 70 out of 95 items with an internal consistency of .93. The four factors are identified under four dimensions such as (1) competence, organization and respect for individuality with 24 items, (2) classroom management with 16 items, (3) dynamic and reflective teaching with 15 items, and (4) affective teaching with 15 items. A new scale to measure teacher effectiveness based on the results of the factor analysis was developed for studies and validation.

Keywords: *Teachers, best practices, teacher effectiveness*

Introduction

In the educational process, the role of teachers in student learning is a great responsibility. More than ensuring that students should be able to give back what are expected from them to learn, teaching calls for effective teaching skills and practices needed to motivate students to learn.

Student motivation naturally has to do with students' desire to participate in the learning process. According to Brophy (1987), motivation to learn is a competence acquired "through general experience but stimulated most directly through modeling, communication of expectations, and direct instruction or socialization by significant others". This statement clearly stressed the role of parents and most of all, the teachers.

According to Lumsden (2006), the beliefs of the teachers about teaching and learning and the nature of the expectations they hold for students also exert a powerful influence. In addition, Lumsden (2006) notes that to a very large degree, students expect to learn if their teachers expect them to learn.

Although students' motivational histories accompany them into each new classroom setting, it is essential for teachers to view themselves as "active socialization agents capable of stimulating . . . student motivation to learn" (Brophy, 1987).

How teachers manage the classroom climate is very important in the learning process. If students experience the classroom as a caring, supportive place where there is a sense of belonging and everyone is valued and respected, they will participate more fully in the process of learning.

In the western perspective, Bain (2004) stipulated that teaching is like parenting which should lead to sustain substantial and positive change in students, without harming them all too much. Those lines suggest the need for a teacher to become effective. What should it take for a teacher to become effective? Bain (2004) further states that "the best teachers focus on the students and what they learn, instead of focusing on the teachers and what they teach". He explains that no depth of insight introduced by the teacher really matters if it does not lead to a new perspective adopted by the student. Teacher has to be just one or two steps ahead of the pupil oftentimes because students get bored if teachers are so much of steps ahead or several steps behind them, no matter how important the subject matter is. Teachers practicing this kind of approach with a child would share fewer experiences with students which is beneficial to both. This practice employs a results-oriented approach, defining what students would be able to do with the knowledge they gain. At the end of the day (or the course), the key question is what difference it made for the students, in the way it made them change.

To be successful in helping diverse students advance to the next level of developing a worldview, the best teachers pay close attention to individual students' learning styles; e.g. students who are passive receptors of information and think that teacher is always right, students who are just adept at giving the answers that teachers expect, without substantially altering their own views, or students who are "committed knowers"- that are developing their own views and ready to defend them (Bain, 2004). In effect, best teachers help students shift their focus from making the grade to thinking about personal goals of development and it should reflect on changes within them.

The best teachers adapt their teaching to different groups of learners (Bain, 2004). They should present the material to match the students' learning styles (seminars for listeners, written assignments for readers and so on). On the other hand, they suggest enrichments for the learning methods to better suit the material being taught (such as helping students with reading techniques, which is valuable for all age groups). Beyond knowing their material, the best teachers ruminate their own thinking and the background philosophy within their disciplines. This helps the teacher be aware of how different students may be in approaching the material. Teaching has to be individualized. The best teachers are good at creating an academic environment in which dissimilar students thrive.

To bring about sustained change, the best teachers appeal to students' sense of accomplishment. They use positive stimuli, often offering material through honors program rather than through remedial assignments. Social recognition and verbal praise go a long way as well. In addition, they engage students' sense of ownership of their own education. It is believed that when students feel in charge of the decision to learn, they would learn better. Because of the idea that students learn best when they answer their own questions, best teacher helps students formulate their own questions. Only through rising up to owning their education can students grow into committed knowers (Bain, 2004).

Regardless of what teaching pedagogy the individual teachers imbibe in the practice of their profession, the effectiveness of their approaches needs to be evaluated and reviewed. To aide teachers review the effectiveness of their teaching strategies, teacher evaluations are being administered to provide feedback to their performance in support to the extent of students' achievements. In most of the schools, teacher performance is reflected in his or her general evaluation ratings. Like any other tool in measurement, the results of any performance evaluation differentiate an effective teacher from the teachers who need to enhance their teaching strategies as indicated by the high and the low ratings.

What should a teacher consider to become effective in motivating student to learn, considering the diverse background and personality of students in the classroom? What teaching techniques and strategies which are worth recognizing do consistent highly-rated teacher employ? Are these techniques and strategies appropriate for effective teaching? With a desire to contribute facts about some concerns in the field of educational psychology, the researcher was motivated by these questions.

According to Good and Brophy (1994), effective teaching is synonymous with active teaching and it is exemplified by a teacher who individualized instruction, maximize instructional time, and create opportunities for students to apply skills and concepts. In addition, effective teachers ask questions requiring higher-order problem solving skills, convey high expectations, monitor students' performance, and provide relevant feedback and reinforcement (Berliner, 1985). Personal traits that have been associated with efficacious teachers include knowledge of the subject content, flexibility, enthusiasm in their delivery of instruction, and strong classroom management skills (Demmon-Berger, 1986).

In the search to discover some important teacher characteristics related to effectiveness in teaching, Fitch (1997) in his master's thesis studied college students' perceptions of the characteristics of effective teachers of deaf and hard-of-hearing students in Rochester Institute of Technology. By administering the structures response questionnaire to 28 college students in Business Technology, he identified the following five core domains of characteristics of effective teacher: (1) Communication of course content and expectations, (2) independent thinking and learning, (3) classroom management, (4) affective, and (5) teacher knowledge.

Another study related to characteristics of effective teachers based on the perception of their colleague was undertaken by Wilson, Dienst, and Watson (1973). Based on the description of 119 faculty members of the University of California, five factors were produced in the principal component analysis which include: (1) Research activity and recognition, (2) participation in the academic

community, (3) intellectual breath, (4) relations with students, and (5) concern for teaching. The results generated were used by researchers in developing a scale to be used in quantifying judgment of teaching effectiveness.

The collaborative study of Witcher et al. (2003) determined the characteristics of effective college teachers based on the perception of the students. Their phenomenological analysis revealed nine characteristics that the students considered for an effective college teaching. These characteristics include the following: (1) Student-centered, (2) knowledgeable of the subject matter, (3) professional, (4) enthusiastic about the teaching, (5) effective communicator, (6) accessible, (7) competent instructor, (8) fair and respectful, and (9) provider of adequate performance feedback.

Drummond (1995) compiled the best practices in college teaching which focus on the aspects of classroom teaching competence. The best practices are labeled in 12 headings which are: (1) Lecture practices which include 7 effective ways to present new information orally to fit differences in learning styles; (2) group discussion triggers which include 8 effective ways to present common experience to engage a group in a discussion; (3) thoughtful questions which include effective ways to formulate questions that foster engagement and confidence; (4) reflective response to learner contributions which include 3 effective ways to establish mutually beneficial communication by reflective listening; (5) rewarding learner participation which include 7 effective ways to support learner actions with well-timed, encouraging positives; (6) Active learning strategies which include 10 effective ways to foster active, constructive participation among students; (7) cooperative group assignments which include 6 ways to assign formal cooperative tasks; (8) goals to grade connections which include establishing a logical agreement of goals and objective, flowing to measures of performance, criteria, and grading; (9) modeling which represents openness, continuous learning, and trust; (10) double loop feedback which include 3 steps in facilitating mutual awareness of how one learns to learn; (11) climate setting which include 4 ways of regulating the physical and mental climate of the students; and (12) fostering learner self-responsibility which include 4 ways of allowing learners to plan and evaluate much of their learning.

In the Philippine setting, Reyes (2002) conducted a survey using an open-ended questionnaire to 134 students. Results revealed that effective teaching constitutes mastery of the topic and appropriate choice of teaching strategy. On the personality side, an effective teacher should have a positive regard and constructive empathy for and towards students. On the philosophical side, effective teachers should promote the ideals of academic excellence, integrity, fairness and equity, and social relevance and transformation.

Reyes (2002) published her study entitled unveiling teaching expertise. It is a descriptive-developmental study showcasing the sixty nine (69) coming from higher education institutions (HEIs) where Metrobank Outstanding Teachers employed, Centers of Excellence, and with Level-3 accredited programs. This highly ethnographic study came up with a Filipino Model of Teaching Expertise. The pyramidal model of teaching expertise has two domains which are the effective teaching and responsible teaching. Effective teaching refers to teachers' instructional practices and behaviors that lead to the attainment of educational objectives which

promotes students' academic achievement. This includes six (6) sub-domains such as matter expertise, instructional expertise, classroom management expertise, diagnostic expertise, communication expertise, and relational expertise. On the other hand, responsible teaching refers to the teacher's ability to develop within the students the desirable attitudes and values towards assuming responsibility for learning. Reyes (2002) concluded that without an effective and responsible teaching, there can never be an expert teacher.

The characteristics of an effective teacher described above and the given results from those studies of Fitch (1997), Wilson et al. (1973), Witcher et al. (2003), and those best practices in college teaching compiled by Drummond (1995) are reflective of what an effective teacher should be or should do on the bases of the perceptions of the teachers to their colleagues as well as of the students to their teachers in the western setting. While those studies provided bulk of information about teacher effectiveness, the present study brings in those studies in the Philippine setting. Based on the general perception of both the teachers and the students, this study seeks to identify the best practices of the highly-rated college teachers in an attempt to contribute to the literature on studies like Reyes (2002) on effective and expert teaching.

The foreign studies presented above generally focus on effective teaching practices. While the limited local studies follows seemingly the same tract except for Reyes (2002) which narrows down the studies from effective to expert teaching, there was no attempt to include span of time in teaching profession as contributory factor to teaching effectiveness or expertise. Reyes (2002) however identified continuing professional growth as precursors of teaching expertise which gives an idea that teacher experiences along these matters are factors. Experiences in general have an element of time which may bring a difference in performance between those who have been teaching for longer years than those who are just new in the profession. This condition triggers the researcher to include the difference in perception between the noble and the expert highly rated teachers. Highly rated teachers in this study include those who are consistently rated high in the teacher evaluation for three consecutive evaluation periods.

The result of this study may provide the academic administrators with the ideas on what teaching practices may be developed and be put into practice in order to effectively stimulate students' interest to learn and so would result to a more productive teaching-learning environment.

This study has a threefold objective. First, it aims to identify the best practices of the highly-rated college teachers which stimulate student's interest in learning. Second is to determine the domains of those practices, and third is to come-up with a new scale to measure teacher effectiveness. To address these objectives, answers to the following questions are sought:

- (1) What common teaching strategies do the highly-rated college teachers use in their respective classes?
- (2) Which of these teaching strategies are the best practices of the highly-rated college teachers?
- (3) What are some of the differences in the best practices between the experts and the novice teachers?
- (4) What dimensions could be extracted from these practices?

Method

This is descriptive study which employs a survey method using a structured response questionnaire designed for this purpose. This study is highly quantitative.

Setting

The questionnaire was administered at Notre Dame of Dadiangas University, a sectarian school managed by the Marist Brothers of the Philippines. It is a PAASCU Level 3 accredited and has been granted by CHED an autonomous status from 2005 to 2010. The school offers traditional courses like Education, Arts and Sciences programs, Engineering, Nursing, Business and allied courses in Information Technology. The school is periodically evaluating its faculty members every semester as part of its administrative program. The result of the evaluation is communicated to all the teachers at the end of the semester. It is also used as one requirement for ranking and as basis for determining faculty development needs.

Respondents

Twenty teacher respondents who are identified as consistently highly-rated in the Teacher Behavior Inventory (TBI) answered the questionnaire. Eight (8) or 40% of the teachers have 2-3 years teaching experience that represent the novice teacher, while 60% or 12 of the teacher respondents have been teaching from 4 to 17 years representing the expert teachers. One hundred five (105) students successfully answered the questionnaire. Fifty six (56) of them are from the two classes of identified novice teacher, while 49 of them are from the classes of the expert teachers. A total of 125 respondents successfully accomplished the questionnaire ($N=125$).

Instrumentation

Two sets of structured response questionnaire were designed intended for the teacher respondents and for the students. These two sets of structured response questionnaires contained the same items but only differ in instructions. The 95 items of the structured response questionnaire were adapted from the different sources such as those that were used in the studies of Fitch (1997) and identified by Wilson, Dienst and Watson (1973). More items were adapted from Kenneth Eble's (1971) "the recognition and evaluation of teaching"; Edward Sheffield's (1974) "teaching in the universities- no one way"; Lea Ebro's (1977) "instructional behavior patterns of distinguished university teachers"; and Karron Lewis et al. "the large class analysis project" conducted by CTE. Since majority of the adapted items are results of qualitative studies, items were restated in consistent with the rest. The questionnaire is composed of 4 columns. The first column is the item number, the second column is the item statement, the third column is a preliminary question with yes or no response. It asked the respondents whether the item stated is manifested by their teacher or not. If the answer is yes, then the respondent will

rate the performance of the teacher in the particular statement. The rating scale to choose is from 1 to 6, 1 means very effective and 6 means not effective.

The survey questionnaire was validated by the experts in scale development and was tried out before its actual administration. There were 135 original items subjected for validation and the final form has only 95- items after being validated. The questionnaire has an internal reliability of .93 Cronbach Alpha.

Procedure

To identify the highly rated teachers, the search is coordinated with the Officer in Charge (OIC) of the Guidance Services who handles the evaluation administration. The highly-rated teachers both expert and novice teachers were identified in coordination with the school academic department head who looked into the files of the teacher performance evaluation results.

Two (2) representative classes from the expert teachers and two (2) representative classes from the novice teachers were randomly selected. Each representative class was administered with the same questionnaire to determine the attributes of their teachers and the strategies commonly used by their teachers. The class was also asked to rate each strategy in terms of how it motivates students to learn.

Expert college teachers are those who have been teaching for three or more years and those who were rated-highly in the performance rating evaluation for three consecutive semesters. On the other hand, novice teachers are those college teachers who are in their first or second year in teaching and who were rated highly for two consecutive semesters.

Analysis

Data processing included computation of the mean rating and factor analysis. The mean rating was computed to determine the common teaching strategies used by the teachers, the best practices, and the difference in practices between the regular and the probationary teachers. Factor analysis was applied to determine dimensions of the best practices.

For inclusion in the common and the best practices, an item must have a mean that is not more than 2.00 and standard deviation of not more than 1.00. In factor analysis, the limited loading is set at .40.

Results

Common Attributes and Strategies of the Highly-rated College Teachers

Of the 95 items representing the teacher attributes and strategies used by effective teacher, the teacher and student respondents identified at least 51 items that are commonly used by the highly rated teacher, based on their perception as shown in Appendix A.

Among the top 10 items are: (1) showing confidence in what they are doing; (2) respecting students as person; (3) clear and well-organized in presenting ideas;

(4) encouraging students' participation; (5) speaking clearly; (6) emphasizing important points in the class; (7) enjoys teaching; (8) interested and concern in the quality of his/her teaching; (9) enthusiastic about the students; and 10) involving students in learning activities.

The items which are not much considered as attributes and strategies of effective teacher are presented in appendix B. These are ranked according to Mean rating and variability of responses (*SD*).

Among the top 10 items that are considered as not much an attribute and strategy of an effective teacher are the following: (1) giving advices not directly related to the course; (2) providing feedback to other teachers; (3) allowing students to interrupt at any time; (4) using students' background data to adapt teaching to individual needs; (5) walking around as they talk; (6) getting right down to business; (7) raring to discipline (govern) to eliminate unnecessary talking, etc.; (8) introducing humor spontaneously; (9) inviting criticism of his/her own ideas; and (10) using gestures frequently .

Best Practices of Highly-rated College Teachers

Table 1
Best Practices of Highly-Rated College Teachers (*N*=125)

	Orig			
	No.	Attribute	<i>M</i>	<i>SD</i>
1	76	Shows confidence in him/herself and what he/she is doing.	1.26	0.46
2	74	Respects students as persons.	1.30	0.51
3	51	Is clear and well-organized in presenting his/her ideas.	1.38	0.56
4	20	Encourages students' participation.	1.38	0.56
5	78	Speaks clearly.	1.27	0.60
6	15	Emphasizes important points in the class.	1.33	0.62
7	23	Enjoys teaching.	1.30	0.64
8	38	Has interest and concern in the quality of his/her teaching.	1.44	0.64
9	53	Is enthusiastic about the subject.	1.46	0.64
10	45	Involves students in learning activities.	1.41	0.65
11	58	Is well prepared.	1.46	0.67
12	77	Shows mastery of his/her subject, competent.	1.36	0.68
13	17	Encourages class discussion.	1.45	0.68
14	5	Cares about the progress and welfare of the students.	1.64	0.69
15	33	Has a genuine interest in students.	1.46	0.69
16	65	Promotes active student learning.	1.44	0.70
17	71	Relates to students as individuals.	1.46	0.70
18	70	Relates the course material to experiences of the students.	1.65	0.73
19	39	Helps with communication among students.	1.58	0.73

Cont. Table 1

20	61	Lectures are well prepared and orderly.	1.38	0.74
21	47	Is able to pace the class lectures to the needs of the students.	1.59	0.74
22	28	Gives immediate response to student question or answer.	1.64	0.74
23	31	Has caring attitude about students and their progress.	1.65	0.75
24	24	Explains clearly.	1.46	0.76
25	55	Is friendly toward students.	1.55	0.77

Based on the respondents' perception, Table 1 shows the teacher attributes and strategies that are considered as the "best practices". The mean response was used as an indicator to determine for the top 25 in rank, 1.0 is the highest possible rate with *SD* not more than 1.00 while 4.0 is the lowest rate. Leading among those attributes and strategies include teacher confidence, respect for individuality, clarity and organization, encouragement of students' participation, and good communication skill.

Differences in Practices between the Experts and Novice Teachers

Table 2 next page presents the differences in perception between the "experts" and the "novice" teacher in terms of what they manifested as "best practices" towards their students.

Table 2 shows that 18 out of the top 25 best practices are the same to both groups (experts and novice) at 72% rate of congruency. To some degree at 28% (7 out of 25), the expert and the novice teachers manifested practices that are considered effective exclusive of the other group. The experts tend to be more nurturing (friendly, clearly explains, careful, helps, and praises) while the novice teachers tend to be more free-flowing, enthusiastic and innovative (enjoy, encourage independence, and variety).

In order to extract dimensions from those identified attributes and strategies of effective teacher, the researcher considered all the items to be included in the factor analysis because of variability in responses. One respondent would rate a particular item as not an attribute and strategy of effective teacher while most respondents rated it the other way and even rated it high.

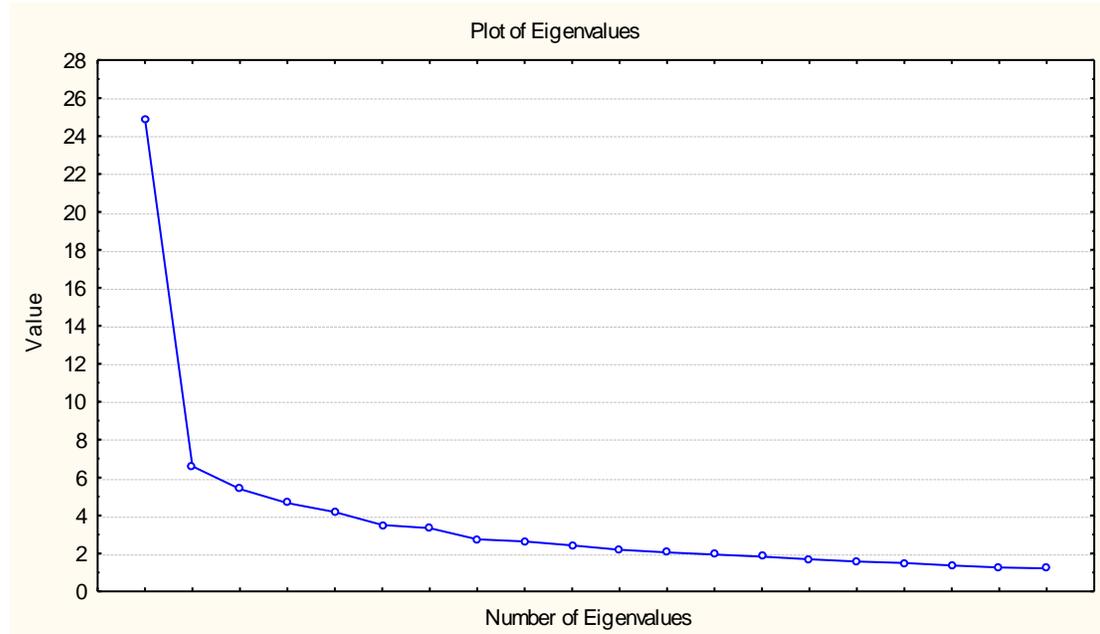
Table 2
Differences in “Best Practices” Between the “Experts” and “Novice” Teachers
(N=125).

	Experts	<i>M</i>		Novice	<i>M</i>
1	Speaks clearly.	1.13	1	Shows confidence in him/herself and what he/she is doing.	1.25
2	Emphasizes important points in the class.	1.17	2	Has high self-confidence.	1.27
3	Is clear and well-organized in presenting his/her ideas.	1.22	3	Enjoys teaching.	1.35
4	Enjoys teaching.	1.24	4	Respects students as persons.	1.35
5	Has a friendly, warm professional attitude.	1.24	5	Shows mastery of his/her subject, competent.	1.38
6	Respects students as persons.	1.24	6	Speaks clearly.	1.38
7	Shows confidence in him/herself and what he/she is doing.	1.26	7	Encourages students' participation.	1.42
8	Explains clearly.	1.28	8	Involves students in learning activities.	1.42
9	Encourages students' participation.	1.33	9	Is enthusiastic about the subject.	1.42
10	Lectures are well prepared and orderly.	1.33	10	Lectures are well prepared and orderly.	1.42
11	Shows mastery of his/her subject, competent.	1.33	11	Emphasizes important points in the class.	1.45
12	Able to communicate knowledge effectively.	1.35	12	Encourages students to learn independently.	1.45
13	Involves students in learning activities.	1.39	13	Has interest and concern in the quality of his/her teaching.	1.45
14	Promotes active student learning.	1.39	14	Seems to enjoy teaching.	1.45
15	Encourages class discussion.	1.41	15	Is well prepared.	1.46
16	Has a genuine interest in students.	1.41	16	Communicates expectations and assignments clearly to students.	1.48
17	Presents facts and concepts from related fields.	1.43	17	Encourages class discussion.	1.48
18	Has an interesting style of presentation.	1.43	18	Presents facts and concepts from related fields.	1.48
19	Has interest and concern in the quality of his/her teaching.	1.43	19	Promotes active student learning.	1.48
20	Relates to students as individuals.	1.43	20	Relates to students as individuals.	1.48
21	Is well prepared.	1.44	21	Is clear and well-organized in presenting his/her ideas.	1.49
22	Is careful and precise in answering questions.	1.46	22	Has a genuine interest in students.	1.51
23	Helps with communication among students.	1.46	23	Uses a variety of instructional strategies.	1.51
24	Praises correct answers and explains why the answer is correct.	1.46	24	Has an interesting style of presentation.	1.54
25	Identifies what he/she considers important.	1.48	25	Encourages students' questions and opinions.	1.55

Note: Items not shaded were chosen by both groups as best practices, while Items shaded are identified as differently by the comparative groups.

The scree plot in Figure 1 indicates at least four to five possible factors that could be extracted from the data, thus, from the factor analysis, 4 factors are extracted. The item loading is set at .40. The summary of the results of the factor analysis is shown in Table 5 that follows.

Figure 1
The Scree Plot



Applying the Factor Analysis

There are four factors extracted after the rotated factor analysis with Cronbach alpha of .93, .84, .86, and .88, respectively. Factor 1 is a cluster of 24 items with an eigenvalue of 24.89; Factor 2 includes 16 items with eigenvalue of 6.59; Factor 3 includes 15 items with an eigenvalue of 5.39; and Factor 4 has 15 items with an eigenvalue of 4.65.

The whole 70 items compose of the four factors have a reliability index (Cronbach alpha) of .93, while the original questionnaire with 95 items has also a reliability index of .93 Cronbach alpha.

Table 3
Rotated Factor Analysis Results (N=125).

Extracted Factor	Number of Items	Cronbach Alpha	Eigenvalue	%Total
F1	24	.93	24.89	26.21
F2	16	.84	6.59	6.94
F3	15	.86	5.39	5.68
F4	15	.88	4.65	4.89
Overall	70	.93	45.72	
Original Questionnaire	95	.93		

Table 4 presents the 24 teacher attributes and strategies that were clustered in the first factor with a Cronbach's alpha of .93. Most of the items are focused on the teacher's communication of knowledge, presentation of facts, ideas and concepts, teacher preparation and organization of carrying out his/her tasks and respect for individual differences. This factor can be identified as a dimension for "competence, organization, and respect for individuality".

Table 4
Items loaded in Factor 1: Competence, Organization, and Respect for Individuality

	Origl No.	
1	1	Able to communicate knowledge effectively.
2	14	Emphasizes conceptual understanding.
3	18	Encourages freedom of expression of ideas and opinions in the classroom.
4	24	Explains clearly.
5	28	Gives immediate response to student question or answer.
6	31	Has a caring attitude about students and their progress.
7	32	Has a friendly, warm professional attitude.
8	33	Has a genuine interest in students.
9	36	Has an interesting style of presentation.
10	39	Helps with communication among students.
11	46	Is a dynamic and energetic person.
12	51	Is clear and well-organized in presenting his/her ideas.
13	55	Is friendly toward students.
14	56	Is knowledgeable of world events.
15	58	Is well prepared.
16	61	Lectures are well prepared and orderly.
17	63	Presents facts and concepts from related fields.
18	70	Relates the course material to experiences of the students.
19	71	Relates to students as individuals.
20	74	Respects students as persons.
21	76	Shows confidence in him/herself and what he/she is doing.
22	77	Shows mastery of his/her subject, competent.
23	78	Speaks clearly.
24	95	Warm, kind, sympathetic to students.

Table 5
Items Loaded in Factor 2: Classroom Management

	Origl No.	
1	10	Dares to discipline (govern) to eliminate unnecessary talking, etc.
2	19	Encourages group work.
3	20	Encourages students' participation.
4	21	Encourages students' questions and opinions.
5	22	Encourages students to learn independently.
6	35	Has a sense of humor.
7	42	Introduces humor spontaneously.
8	43	Invites criticism of his/her own ideas.
9	44	Invites students to share their knowledge and experiences.
10	57	Is on time for class.
11	60	Knows if the class understands him/her or not.
12	66	Provides (specific) feedback to the students.
13	89	Uses probing questions.
14	90	Uses students' background data to adapt teaching to individual needs.
15	92	Uses visual materials (overheads, blackboard, handouts, etc).
16	94	Walks around as they talk.

The second factor which has a Cronbach alpha of .84 is a cluster of 16 teacher attributes and strategies that are more focused on classroom discipline, encouragement for class discussions, group works, participations and independent learning. It also includes items that focus on critical thinking, art of questioning, and bench marking. In addition, it also includes items that focus on the importance of sense of humor, used of visual materials and reaching out each student in the class when talking. This factor can be identified as a dimension for “classroom management”.

Table 6
Items Loaded in Factor 3: Dynamic and Reflective Teaching

	Origl No.	
1	9	Contrasts implications of various theories.
2	11	Discusses points of view other than his/her own.
3	13	Does not intimidate students.
4	26	Gives advices not directly related to the course.
5	40	Identifies what he/she considers important.
6	52	Is concerned for students' progress and needs.
7	54	Is flexible (not overly strict).
8	67	Provides corrective feedback.
9	68	Provides feedback to others teachers.
10	72	Requires students to apply concepts to demonstrate understanding.
11	73	Respects students as individuals who can contribute to the course.
12	81	Summarizes major points.
13	82	Tests to measure pre-requisite skills.
14	83	Treats students as adults.
15	84	Understands cultural differences among students.

Factor 3 which has a Cronbach alpha of .84 is composed of 12 items that focus on teaching students to learn by considering the ideas of others, openness, enthusiasm, proper use of feedback to help student learning, evaluation, and trying to get into the level of their students. This factor can be identified as a dimension for “dynamic and reflective teaching”.

Table 7 shows the 15 items that were clustered for Factor 4. This group of 15 items has a Cronbach alpha of .88. These items are focused on caring for the welfare of the students, clarifying reasons for students' questions, having clear communication of the expectations and assignments, having emphasis on learning the course vocabulary to students, directing students to productive learning, having flexibility of teaching methods and being open to students' needs, and having an extensive use of non-verbal communications for better understanding. This factor can be identified as dimension for “affective teaching”.

Table 7
Items Loaded in Factor 4: Affective Teaching

	Origl No.	
1	5	Cares about the progress and welfare of the students.
2	6	Clarifies thinking by identifying reasons for questions.
3	7	Communicates expectations and assignments clearly to students.
4	16	Emphasizes important vocabulary in the course.
5	25	Gets right down to business.
6	30	Gives references for more interesting and involved points.
7	47	Is able to pace the class lectures to the needs of the students.
8	48	Is accessible to students outside of the class.
9	49	Is approachable, friendly and available to student consultations.
10	69	Recognizes and greets students outside of the class.
11	75	Seems to enjoy teaching.
12	80	Stays with their subjects.
13	86	Uses facial expressions.
14	88	Uses of eye-contact extensively.
15	93	Varies the speed and tone of his/her voice.

Items Included for the New Set of Questionnaire

In putting together all the 70 loaded items after the rotated factor analysis, (24 items for Factor 1, 16 items for Factor 2, 15 items for Factor 3, and 15 items for Factor 4) the new set of items is presented in Table 8 with the newly assigned item number. The items were arranged alphabetically just for the sake of organization.

If a scale to measure teacher effectiveness is to be developed by using the 70 above items, the arrangement of items will appear in the newly assigned number as presented in Table 8 arranged alphabetically.

Table 8
Overall 70 Loaded Items

No.	Item	Factor
1	Able to communicate knowledge effectively.	1
2	Cares about the progress and welfare of the students.	4
3	Clarifies thinking by identifying reasons for questions.	4
4	Communicates expectations and assignments clearly to students.	4
5	Contrasts implications of various theories.	3
6	Dares to discipline (govern) to eliminate unnecessary talking, etc.	2
7	Discusses points of view other than his/her own.	3
8	Do not intimidate students.	3
9	Emphasizes conceptual understanding.	1
10	Emphasizes important vocabulary in the course.	4
11	Encourages freedom of expression of ideas and opinions in the classroom.	1
12	Encourages group work.	2
13	Encourages student's participation.	2
14	Encourages students' questions and opinions.	2
15	Encourages students to learn independently.	2
16	Explains clearly.	1
17	Gets right down to business.	4
18	Gives advices not directly related to the course.	3
19	Gives immediate response to student question or answer.	1
20	Gives references for more interesting and involved points.	4
21	Has a caring attitude about students and their progress.	1
22	Has a friendly, warm professional attitude.	1
23	Has a genuine interest in students.	1
24	Has a sense of humor.	2
25	Has an interesting style of presentation.	1
26	Helps with communication among students.	1
27	Identifies what he/she considers important.	3
28	Introduces humor spontaneously.	2
29	Invites criticism of his/her own ideas.	2
30	Invites students to share their knowledge and experiences.	2
31	Is a dynamic and energetic person.	1
32	Is able to pace the class lectures to the needs of the students.	4
33	Is accessible to students outside of the class.	4
34	Is approachable, friendly and available to student consultations.	4
35	Is clear and well-organized in presenting his/her ideas.	1
36	Is concerned for students' progress and needs.	3
37	Is flexible (not overly strict).	3
38	Is friendly toward students.	1

Cont. Table 8

39	Is knowledgeable of world events.	1
40	Is on time for class.	2
41	Is well prepared.	1
42	Knows if the class understands him/her or not.	2
43	Lectures are well prepared and orderly.	1
44	Presents facts and concepts from related fields.	1
45	Provides (specific) feedback to the students.	2
46	Provides corrective feedback.	3
47	Provides feedback to others teachers.	3
48	Recognizes and greets students outside of the class.	4
49	Relates the course material to experiences of the students.	1
50	Relates to students as individuals.	1
51	Requires students to apply concepts to demonstrate understanding.	3
52	Respects students as individuals who can contribute to the course.	3
53	Respects students as persons.	1
54	Seems to enjoy teaching.	4
55	Shows confidence in him/herself and what he/she is doing.	1
56	Shows mastery of his/her subject, competent.	1
57	Speaks clearly.	1
58	Stays with their subjects.	4
59	Summarizes major points.	3
60	Tests to measure pre-requisite skills.	3
61	Treats students as adults.	3
62	Understands cultural differences among students.	3
63	Uses facial expressions.	4
64	Uses of eye-contact extensively.	4
65	Uses probing questions.	2
66	Uses students' background data to adapt teaching to individual needs.	2
67	Uses visual materials (overheads, blackboard, handouts, etc).	2
68	Varies the speed and tone of his/her voice.	4
69	Walks around as they talk.	2
70	Warm, kind, sympathetic to students.	1

Discussion

Looking back at the results of the identified common attributes and strategies of highly-rated teachers, one could see much similarities between the western perspective of an effective teacher (Bain, 2004; Good & Brophy, 1994; Witcher, et al., 2003; Berliner, et al., 1985) and the findings of the studies of Fitch (1997) and Wilson et al. (1973), on domains of characteristics of effective teacher in the Philippine perspective. For some, it includes showing confidence, respect for students, clarity and organization, encouraging students' participation and good communication skill (see Appendix A).

Some differences in the common teaching practices were also noted which can be explained by cultural variables (see Appendix B). An example for this is allowing students to interrupt any time in the class which is considered as provision for “warm classroom climate” (Ebro, 1977). The said characteristic may not be much accepted in the seemingly conservative culture in the Philippines which is rooted by tradition. To the western culture, this may be part of their practices which may be observed some International schools in the country where students are encouraged to speak out their minds.

Moreover, strategies like giving advices not directly related to the course; providing feedback to other teachers; allowing students to interrupt at any time; using students' background data to adapt teaching to individual needs; and walking around as they talk may not be commonly employed locally but have been a trend in education.

The result on best practices (refer to Table 1) among the highly-rated college teachers which has themes of teacher confidence, respect for individuality, clarity and organization, encouragement of students' participation, and good communication skill perfectly matched with most of the 9 characteristics of effective teacher identified by Witcher, et al. (2003). These are termed essentials according to the Filipino model of teaching expertise by Reyes (2002).

Generation gap which may be differentiated the length of time spend in teaching may contribute to the differences in approach to teaching between the experts and the novice teachers. However, they are still considered effective by the students. Students may also have different expectations from teachers who have been teaching for quite some time against those new teachers by profession. These can be noted in the differences in practices considered effective by the respondents especially by the students. This may be explained by the findings of Reyes (2002) that the experts usually advise the novice teachers to undergo further studies and exposure even if they effective already in their craft. They still have to learn other strategies to keep up-dated by the trends.

In the course of doing the factor analysis, the original plan was to include only the items that are selected by the respondents as attributes and strategies of highly-rated college teacher. In the process, the researcher noticed the variability of the respondents' opinion in a particular item, e.g., one particular item is considered “NOT” an attribute or strategy of a highly-rated teacher by some respondents, but is also considered by some in the other way and was even rated very highly effective. Looking back at the items, content wise, there might be a reason for validity of both responses. Because of this discrepancy, the researcher decided to include all the original 95 items in the factor analysis. Surprisingly, the result of factor analysis vindicated the researcher's decision because there are 28 out of the 37 items that were not considered as “NOT” attribute or strategy of highly-rated teacher that were loaded into the domains. On the other hand, there are 23 out of the 58 items considered attribute or strategy of highly-rated teacher that were not loaded into the domains.

More so with the results of the factor analysis, Factor 1 (refer to Table 4) as dimension for “competence, organization, and respect for individuality” is similar and can be explained in consonance to Witcher et al.'s (2003) knowledge of the subject matter, student-centered, competent instructor, professional, or flexibility,

and Wilson, Dienst, and Watson's (1973) intellectual breath, relationship with students, and concern for teaching. Student-centered however is considered as enhancers in the local study of Reyes (2002), under the domain of responsible teaching, this is a teacher quality that develops responsibility for their learning.

Factor 2 (refer to Table 7) as dimension for "classroom management" includes facilitation of class discussions, group activity, individual participation and expression of ideas and concepts is comparable to Witcher et al. (2003) classroom management, a characteristic of efficacious teacher. In Reyes (2002), these factors belong to the upper level in the pyramidal model of teaching expertise which are under the domain of responsible teaching.

Factor 3 (refer to Table 8) as dimension for "dynamic and reflective teaching" includes items that focus on teaching students to learn by considering the ideas of others, openness, enthusiasm, proper use of feedback to help student learning, evaluation, and trying to get into the level of their students shows teacher dynamism (Eble, 1971) and the ability to give feedback to students and authorities regarding how the students might learn sharing the same explanation with Drummond's (1997) reflective response to learner contributions. In Reyes (2002), all these characteristics can be under the classification of enhancers, the sub-domains of responsible teaching.

Factor 4 (refer to Table 9) as dimension for "affective teaching" has a theme that focuses on caring for the welfare of the students, getting into the reasons for students' questions, having clear communication of the expectations and assignments, giving emphasis to student learning in the course and guiding students to productive learning, having flexibility to students' needs, and having the extensive use of non-verbal communications for better understanding shares the same explanation with Fitch's (1997) domains of characteristics of effective teacher. To Reyes (2002) most of these are considered as essentials to teaching expertise, where all the teachers should master in order to achieve the academic objectives.

In all, a total of 70 items are accepted representing the 4 dimensions such as: (1) Competence, organization, and respect for individuality; (2) classroom management; (3) dynamic and reflective teaching; and (4) affective teaching. If a new scale to measure teacher effectiveness on the basis of the result of factor analysis will be validated, it should contain these 70 accepted items which was arranged alphabetically for organization purposes.

Based on the results, the leading attributes and strategies of the highly-rated teacher in this particular school are focused on effective communication of knowledge, consideration and flexibility to adapt to the individual differences among students, care about the progress and welfare of the students, effort to clarify thinking by identifying reasons for students' questions; and clear communication of expectations and assignments to students.

The attributes and strategies that are not much an attribute and strategies of highly-rated teachers in this particular school include allowing students' to interrupt at any time; giving advices not directly related to the course; providing feedback to other teachers; using students' background data to adapt teaching to individual needs; and walking around the classroom as they talk.

The best practices of the highly-rated college teachers include showing confidence with him/herself and with what he/she is doing; speaking clearly;

enjoying teaching; respecting students as persons; and emphasizing important points in the class. The difference in the best practices between the experts and the novice highly-rated college teachers is that the experts tend to be more nurturing while the “novice” teachers tend to be more liberated, and practiced the value of teaching with enthusiasm and employing variety of techniques.

The four dimensions of the best practices of the highly-rated college teachers are identified such as: Factor 1 is competence, organization, and respect for individuality; Factor 2 is classroom management; Factor 3 is dynamic and reflective teaching; and Factor 4 is affective teaching.

The 74% (70/95) acceptance rate of the items loaded in factor analysis shows that the western perspective of an effective teacher highly shares with the Philippine perspective. However, there are some effective teacher attributes and strategies in the western context that are may not be acceptable in the local context for reasons of culture and tradition.

On the basis of the results, the following recommendations are given: (1) The same questionnaire will be administered to a larger scale to other accredited colleges and universities where teacher evaluation is a part of the curriculum program; (2) larger number of students respondents representing classes of highly-rated college teachers should be involved; (3) inclusion of some open-ended questions in the questionnaire to determine some culturally unique best practices of highly-rated college teacher in the Philippine perspective should be considered; (4) confirmatory factor analysis on the basis of the identified best practices of the highly-rated teachers and comparison of the results with the results of this study should be done, and (5) interpretation of this study should be taken with great precautions.

References

- Bain, K. (2004). *What the best college teacher do*. CA: Harvard University Press.
- Brophy, J. (1986). *On motivating students*. Occasional Paper No. 101. East Lansing, Michigan: Institute for Research on Teaching, Michigan State University.
- Berliner, D. C. (1985). Effective classroom teaching: The necessary but not sufficient condition for developing exemplary schools. In G. R. Austin & H. Garber (Eds.), *Research in exemplary schools* (pp. 127-151). Orlando, FL: Academic Press.
- Demmon-Berger, D. (1986). *Effective teaching: Observations from research*. Arlington, VA: American Association of School Teachers.
- Drummond, T. (1995). *Compilation of best practices in college teaching*. Center for Teaching & Learning, USA: UNC Charlotte.
- Eble, K. (1971). *The recognition and evaluation of teaching*. Washington, DC: American Association of University Professors.
- Ebro, L. (1977). *Instructional behavior patterns of distinguished university Teachers*. Unpublished doctoral dissertation, Ohio State University.
- Fitch, C. (1997). *College students' perceptions of the characteristics of effective teachers of deaf and hard-of-hearing students*. Unpublished master's thesis, Rochester Institute of Technology, NY.

- Good, T. L., & Brophy, J. E. (1994). *Looking in classrooms* (6th ed.). New York: Harper Collins.
- Lewis, K., et. al. (1982). *The large class analysis project*. Austin, TX: University of Texas at Austin.
- Lumsden, L. S. (2006). Tenure policies and practices of American evangelical colleges and universities part 2: Institutions granting tenure. *Christian Higher Education*, 5, 341-364.
- Reyes, F. C. (2002). *Unveiling teaching expertise: A showcase of sixty-nine outstanding teachers in the Philippines*. Manila: De La Salle University Press, Inc.
- Wilson, R. C., Dienst, E. R., & Watson, N. L. (1973). *Characteristics of effective college teachers as perceived by their colleagues*. *Journal of Educational Measurement*, 10(1), 31-7
- Witcher, A. E., Onwuegbuzie, A. J., Collins, K. M., Filer, J. D., Wiedmaier, C. D., & Moore, C. (2003). *Students' perceptions of characteristics of effective college teachers*. Technical Report 143, US Department of Education.

Appendix A
Common Attributes and Strategies of Highly-rated College Teachers (N=125)

	Origl No.	Attribute	<i>M</i>	<i>SD</i>
1	76	Shows confidence in him/herself and what he/she is doing.	1.26	0.46
2	74	Respects students as persons.	1.30	0.51
3	51	Is clear and well-organized in presenting his/her ideas.	1.38	0.56
4	20	Encourages students' participation.	1.38	0.56
5	78	Speaks clearly.	1.27	0.60
6	15	Emphasizes important points in the class.	1.33	0.62
7	23	Enjoys teaching.	1.30	0.64
8	38	Has interest and concern in the quality of his/her teaching.	1.44	0.64
9	53	Is enthusiastic about the subject.	1.46	0.64
10	45	Involves students in learning activities.	1.41	0.65
11	58	Is well prepared.	1.46	0.67
12	77	Shows mastery of his/her subject, competent.	1.36	0.68
13	17	Encourages class discussion.	1.45	0.68
14	5	Cares about the progress and welfare of the students.	1.64	0.69
15	33	Has a genuine interest in students.	1.46	0.69
16	65	Promotes active student learning.	1.44	0.70
17	71	Relates to students as individuals.	1.46	0.70
18	70	Relates the course material to experiences of the students.	1.65	0.73
19	39	Helps with communication among students.	1.58	0.73
20	61	Lectures are well prepared and orderly.	1.38	0.74
21	47	Is able to pace the class lectures to the needs of the students.	1.59	0.74
22	28	Gives immediate response to student question or answer.	1.64	0.74
23	31	Has a caring attitude about students and their progress.	1.65	0.75
24	24	Explains clearly.	1.46	0.76
25	55	Is friendly toward students.	1.55	0.77
26	1	Able to communicate knowledge effectively.	1.55	0.78
27	49	Is approachable, friendly, available to student consultations.	1.56	0.79
28	36	Has an interesting style of presentation.	1.49	0.79
29	14	Emphasizes conceptual understanding.	1.70	0.79
30	18	Encourages freedom of expression of ideas and opinions in the classroom.	1.57	0.80
31	64	Presents origins of ideas and concepts.	1.67	0.82
32	91	Uses teaching aids effectively.	1.65	0.83
33	32	Has a friendly, warm professional attitude.	1.49	0.83
34	85	Uses a variety of instructional strategies.	1.50	0.83
35	95	Warm, kind, sympathetic to students.	1.66	0.83
36	7	Communicates expectations and assignments clearly to students.	1.56	0.84
37	21	Encourages students' questions and opinions.	1.56	0.84
38	62	Praises correct answers and explains why the answer is correct.	1.53	0.85

Cont. Appendix A

39	22	Encourages students to learn independently.	1.63	0.87
40	2	Adapts to students' individual differences.	1.78	0.87
41	8	Connects subject to life /practical experiences.	1.68	0.87
42	27	Gives clear lectures.	1.65	0.87
43	6	Clarifies thinking by identifying reasons for questions.	1.60	0.88
44	44	Invites students to share their knowledge and experiences.	1.70	0.88
45	34	Has a good sense of humor.	1.72	0.89
46	63	Presents facts and concepts from related fields.	1.46	0.94
47	52	Is concerned for students' progress and needs.	1.58	0.94
48	60	Knows if the class understands him/her or not.	1.67	0.96
49	50	Is careful and precise in answering questions.	1.54	0.96
50	40	Identifies what he/she considers important.	1.52	0.96
51	30	Gives references for more interesting and involved points.	1.86	0.99

Note: Items with Mean of not more than 2.0 with SD of not more than 1.00 are included

Appendix B
Items not much Considered Attributes and Strategies of Highly-rated Teacher
(N=125)

	Orig No.	Attributes and Strategies	<i>M</i>	<i>SD</i>
1	26	Gives advices not directly related to the course.	3.87	2.48
2	68	Provides feedback to other teachers.	3.44	2.46
3	3	Allows students to interrupt at any time.	4.26	2.25
4	90	Uses students' background data to adapt teaching to individual needs.	2.88	2.12
5	94	Walks around as they talk.	2.70	2.00
6	25	Gets right down to business.	2.46	1.97
7	10	Dares to discipline (govern) to eliminate unnecessary talking, etc.	2.46	1.92
8	42	Introduces humor spontaneously.	2.41	1.92
9	43	Invites criticism of his/her own ideas.	2.48	1.86
10	87	Uses gestures frequently.	2.56	1.85
11	13	Do not intimidate students.	2.13	1.66
12	73	Respects students as individuals who can contribute to the course.	1.94	1.57
13	9	Contrasts implications of various theories.	2.27	1.52
14	67	Provides corrective feedback.	1.93	1.50
15	83	Treats students as adults.	2.06	1.49
16	66	Provides (specific) feedback to the students.	2.02	1.46
17	80	Stays with their subjects.	1.93	1.44
18	41	Interacts with students during, as well as before and after class.	1.90	1.42
19	57	Is on time for class.	1.91	1.42
20	29	Gives lectures that are easy to outline.	2.00	1.39
21	16	Emphasizes important vocabulary in the course.	1.90	1.39
22	35	Has a sense of humor.	1.98	1.36
23	11	Discusses points of view other than his/her own.	2.23	1.36
24	75	Seems to enjoy teaching.	1.54	1.35
25	86	Uses facial expressions.	1.99	1.35



The 2 x 2 achievement goal framework and intrinsic motivation among Filipino students: A validation study

Elmer dela Rosa

Central Luzon State University, Philippines

Abstract In Western studies, the applicability of the 2 x 2 achievement goal framework (Elliot & McGregor, 2001) was established from within diverse samples (e.g., Cury, Elliot, Fonseca, & Moller, 2006; Witkow & Fugilni, 2007). However, there is not much evidence for the theoretical and psychometric validity of the 2 x 2 framework among Asian populations, where some research suggest problems with some of the basic conceptual categories underlying the model. This study explored the applicability of the four-factor achievement goal model (mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance) among Filipino high school students, with particular emphasis on the item analysis, reliability and validity of the measure; and examined the predictive utility of the framework on intrinsic motivation. Results from item analysis provided a relatively strong support that the 2 x 2 achievement goal measure was internally consistent. Exploratory factor analysis showed only three distinct factors as against the hypothesized four-factor model. All items representing avoidance goals (mastery-avoidance and performance-avoidance) significantly loaded on a single factor. Confirmatory factor analysis was performed using both absolute and incremental fit indices. Results indicate that the data did not fit the model under investigation. Nevertheless, all of the achievement goals except performance-avoidance goals were significant predictors of student's intrinsic motivation.

Keywords: *Achievement goal, mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance, intrinsic motivation*

Introduction

One of the most significant variables in the motivational research in educational milieu is the type of achievement goals pursued by students. Achievement goals reflect the purpose of an individual's achievement pursuits (Dweck & Leggett, 1988), and have emerged as the most prominent account of individual's affect, cognition, and behavior in competence-relevant settings (Ames, 1984a; Elliot & Church, 1997). In understanding student's motivation, it focuses on

why individuals are motivated rather than whether individuals possess or lack motivation (Ames, 1992a). Achievement goal theorists commonly defined achievement goals as the purpose of task engagement (Maehr, 1983), and the specific type of goal is conceived to create a framework for how individuals interpret, experience, and act in their achievement pursuits (Dweck, 1986). Achievement goals, which are either defined normatively (demonstrating competence relative to others) or self-referentially (developing competence or skills) are considered to be important determinants of people's motivational orientation toward competence (Harackiewicz & Elliot, 1993). As such, achievement goals are viewed as important predictors of a host of achievement-relevant processes and outcomes.

Generally there are two goal types: mastery goals and performance goals. These goals have been differentiated by their linkages to contrasting patterns of motivational processes. Central to mastery goal is a belief that effort and outcome co-vary, and it is this belief pattern that maintains achievement-directed behavior over time (Weiner, 1986). The focus of attention is on the intrinsic value of learning (Meece & Holt, 1990). With students pursuing mastery goals, they are oriented toward developing new skills, trying to understand their work, improving their level of competence, or achieving a sense of mastery based on self-referenced standards (Ames, 1992b). Thus, the motivation of the individual is focused on mastering and understanding the content of the lesson and shows more willingness to engage in the learning process. On the other hand, the central focus of performance goals is on one's ability and sense of self-worth (Dweck, 1986). In this goal orientation, ability is evidenced by doing better than others, and surpassing normative-based standards, or by achieving with little effort (Ames, 1984b). In this performance orientation, learning is viewed only as a way to achieve a desired goal and student's attention is directed towards achieving normatively defined success. When students adopt performance goals, they perceive ability-outcome linkages and their behavior are guided by their perception of their own ability to perform (Covington & Beery, 1976). Thus, the motivation of the individual is focused on demonstrating their superior skills to other students in his class. According to this dichotomous conceptualization, the extent to which student endorse mastery goals is negatively associated to the extent to which they endorse performance goals, and this pattern have been illustrated in a number of empirical studies (i.e., Nicholls, Cobb, Wood, Yackel, & Patashnick, 1990; Rhodewalt, 1994).

Traditional research evidence suggests that it is mastery goal orientation that promotes a motivational pattern likely to promote long-term and high-quality involvement in learning (Ames, 1992a). However, this omnibus dichotomy (the mastery-performance distinction) was unable to explain why at times, mastery and performance goals resulted in similar levels of objective achievement despite different patterns of engagement (Elliot, 1999). Furthermore, empirical examination of the mastery-performance framework revealed that performance goals were not consistently associated with maladaptive outcomes (Elliot & Church, 1997). Researcher therefore revised the original goal theory by dividing performance goals into approach and avoidance orientation, which was labeled as the trichotomous goal framework. The performance goal dichotomy is distinguished by either a desire to achieve success, associated with performance-

approach, or to avoid failure, associated with performance-avoidance (Elliot & Church, 1997; Elliot & Harackiewicz, 1996). In a study testing a hierarchical model of approach and avoidance achievement motivation (Elliot & Church, 1997), it was found that mastery goals facilitated intrinsic motivation but the same goals did not have a reliable effect on graded performance. Performance-avoidance goals on the other hand proved negative to both intrinsic motivation and graded performance while performance-approach goals manifested a null relationship with graded performance. The offering of a trichotomous achievement goal framework provided empirical evidence that the two functionally separate goals lead to different outcomes, with performance-approach goals linked with some adaptive outcomes especially achievement (Elliot, 1999; Harackiewicz, Baron, & Elliot, 1998), and performance-avoidance goals linked with maladaptive outcomes (Elliot, 1997; Elliot & Church, 1997; Elliot & Harackiewicz, 1996). With this revised trichotomous framework, it is not performance goals in general, but performance-avoidance goals in particular that are consistently maladaptive (Harackiewicz, Barron, Pintrich, Elliot, & Trash, 2002). Both mastery and performance demonstration goals are viewed as approach-oriented because they involve regulation in relation to potential positive outcomes, while performance-avoidance are viewed as avoidance-oriented because they involve regulation in relation to potential negative outcomes (Elliot, 1997; Elliot & Church, 1997; Elliot & Harackiewicz, 1996).

Using undergraduate students in an introductory-level psychology class from a New York university as samples in the investigation, Elliot & McGregor (2001) presented a further revision of the trichotomous framework by splitting mastery goals to create mastery-approach and mastery-avoidance achievement goals. The rationale of the proposed framework is based on how competence, which is at the conceptual core of the achievement goal construct, is defined or valenced. Competence could either be defined as absolute/intrapersonal (the requirement of the task itself or developing own understanding and mastery) or normative (the performance of others). On the other hand, competence could either be positively valenced (approaching success) or negatively valenced (avoiding failure). This is the 2 x 2 achievement goal framework. It consists of four distinct goals and these goals are the individuals' representation of competence-based outcomes which they strive to attain or avoid. These goals include mastery approach (focused on attaining intrapersonal competence), performance-approach (focused on attaining normative competence), mastery-avoidance (focused on avoiding intrapersonal competence) and performance avoidance (focused on avoiding normative incompetence). Figure 1 shows the representation of the 2 x 2 achievement goal framework.

Figure 1
The 2 x 2 Achievement Goal Framework
Definition and Valence Represent the two Dimensions of Competence (Elliot & McGregor, 2001).

		Definition	
		Absolute/intrapersonal (mastery)	Normative (performance)
Valence	Positive (Approach)	Mastery-Approach Goal	Performance-Approach Goal
	Negative (Avoid)	Mastery-Avoidance Goal	Performance-Avoidance Goal

In Western studies, the fit or the utility of the 2 x 2 framework was established from within a diverse sample (e.g., Cury, et al., 2006; Elliot & McGregor, 2001; Finney, Pieper, & Baron, 2004; McGregor & Elliot, 2002; Witkow & Fulgini, 2007). Meanwhile, studies among Filipino samples (i.e. Bernardo, 2003) linked achievement goals and learning among Filipino samples, but focused on adopting the mastery-performance distinction. Thus, the purpose of this study was to explore the fit or applicability of the 2 x 2 achievement goals framework among Filipino students. Furthermore, it is sought to determine whether the achievement goal framework is predictive of intrinsic motivation, defined as the inherent tendency to seek out challenges to explore and to learn (Deci & Ryan, 1985).

Method

Sample and Sampling

A total of 682 high school students (388 females and 294 males) from two national secondary schools (both $N=341$, 170 juniors and 171 seniors) served as participants. The samples' ages ranged from 13 to 22 years (mean age = 15.04 years, $SD = .91$)

Measures

Items for all measures were worded so that students would refer to their Mathematics class when answering the items.

Achievement goals. The items were lifted from the instrument used by Elliot and McGregor (2001). Sample items are "it is important for me to

understand the content of this course as thoroughly as I'd like" for mastery-approach; "I worry that I may not learn all that I possibly could in this class" for the mastery-avoidance; "It is important for me to do better than other students in this class" for performance-approach; and "I just want to avoid doing poorly in this class" for performance-avoidance. Respondents indicated the extent to which they thought each item was true of them on a 1(not all true of me) to 7 (very true of me) scale.

Intrinsic motivation. Seven items were used to assess participants' intrinsic motivation toward their Mathematics class (Elliot & Harackiewicz, 1996). Sample items include "I think this class is interesting". Participants indicated their responses using a 1 (strongly disagree) to 7 (strongly agree) scale. The scale reliability coefficient was .83.

Data Analysis

The analysis involved dividing the data sets into two halves, in which the first half was intended for exploratory factor analysis and the second half was utilized for the confirmatory factor analysis. Descriptive statistics like mean and *SD* were used to report student's responses on the achievement goal questionnaire and intrinsic motivation. Pearson (*r*) was used to establish the intercorrelations of the different variables. Item-scale correlation was used to describe the discrimination indices of the items in each subscale. Cronbach alpha was computed to estimate the reliability of each of the subscales of the measure. Factor Analysis was conducted on the twelve achievement goal items using principal components extraction with varimax rotation. Confirmatory factor analysis was performed using Statistica 8.0 to test the fit of the four-goal model. For this procedure, three fit indices were used. These included: (1) chi-square fit index (2) chi-square over degrees of freedom index, (3) root-mean-square error of approximation (RMSEA), (4) Comparative Fit Index (CFI), and (5) the Non-Norm Fit Index (NNFI) or the Tucker-Lewis Index (TLI). Finally, regression analyses were performed to examine the predictive utility of the achievement goals to student's intrinsic motivation.

Results

Descriptive Statistics, Item Analysis and Reliabilities

The mastery-approach scale had the highest mean (5.94) while mastery-avoidance had the lowest (4.92). The correlations among the measures indicate, that mastery-approach goals were positively correlated with performance-approach goal ($r = .41, p < .01$); but it is also correlated with performance-avoidance ($r = .38, p < .01$). Similarly, performance-approach goals were positively correlated with performance-avoidance ($r = .33, p < .01$) but were also correlated with mastery-avoidance ($r = .41, p < .01$). Correlations also show that all scales except the mastery-avoidance scale were significantly correlated with intrinsic motivation. It suggests that students who are most likely to adopt mastery-approach ($r = .36$) and

performance-approach ($r = .27$) goals have higher levels of intrinsic motivation in their Math classes.

Analyses showed that of the four scales of the achievement goal questionnaire, the highest average inter-item correlation (.40) was found among the items for the performance-approach and mastery-avoidance (Table 2). It can be observed that there are no much difference in the reliability coefficients of the scales and the item-scale correlations. Nevertheless, coefficients were only moderate.

Table 1
Descriptive Statistics and Intercorrelations Among Variables

	<i>M</i>	<i>SD</i>	1	2	3	4	5
(1) Mastery-approach goals	5.94	.85	---				
(2) Mastery-avoidance goals	4.96	1.15	.18*	---			
(3) Performance-approach goals	5.23	1.13	.41*	.17*	---		
(4) Performance-avoidance goals	5.67	1.09	.38*	.43*	.33*	---	
(5) Intrinsic motivation	5.69	.90	.36*	.01	.27*	.18*	---

* $p < .01$

Table 2
Item-scale Correlation and Reliability Coefficients

Achievement Goal	Item-total Correlation	Alpha if item is deleted
Mastery-approach		
Item 1	.50	.42
Item 5	.39	.58
Item 9	.41	.56
Scale Mean=17.83	M=.36	Cronbach's alpha=.63
Mastery-avoidant		
Item 2	.47	.57
Item 6	.46	.58
Item 10	.49	.55
Scale Mean=14.87	M=.40	Cronbach's alpha=.66
Performance-approach		
Item 3	.50	.49
Item 7	.42	.60
Item 11	.46	.55
Scale Mean=15.70	M=.40	Cronbach's alpha=.65
Performance avoidance		
Item 4	.42	.61
Item 8	.52	.48
Item12	.45	.57
Scale Mean=17.02	M=.39	Cronbach's alpha=.65

Exploratory Factor Analysis

An EFA was conducted on the 12 achievement goal items using principal component extraction with varimax rotation. Only three factors had eigenvalues exceeding unity (Kaiser criterion eigenvalue >1, which accounted) for a total of 51.88% of the variance explained (Table 3).

Table 3
Eigenvalues of the Extracted Factors

Factor	Eigenvalue	% Total	Eigenvalue Cumulative	Percentage Cumulative
1	3.48	28.99	3.84	28.99
2	1.69	14.05	5.16	43.04
3	1.06	8.84	6.23	51.88

All the avoidance items (mastery and performance) loaded significantly on a common factor. The summary of factor loadings of the actual items is presented in Table 4. For the performance-approach scale, all the three items significantly loaded on a single factor (loadings ranged from .71 to .77), and likewise for the mastery-approach scales; all of the three items loaded on a single factor (loadings ranged from .67 to .77)

Table 4
Factor Loadings for Achievement Goals

Achievement goal items	F1	F2	F3
I just want to avoid doing poorly in this class		.51	
My goal in this class is to avoid performing poorly		.49	
My fear to performing poorly in this class is what motivates me to study		.55	
I worry that I may not learn all that I possibly could in this class		.71	
Sometimes, I am worried that I do not understand the content of lesson as thoroughly as possible.		.72	
I am often concerned that I may not learn all about there is to learn in this class		.71	
It is important for me to do better than other students in this class	.72		
It is important for me to do well compared to others in this class	.74		
My goal in this class is to get a better grade than most of the other student	.77		
I want to learn as much as possible from this class			.77
It is important for me to understand the content of this course as thoroughly as possible			.67
I desire to completely master the lessons presented in this class.			.70

Note. N=341. All factor loadings > .40. Factor loadings were obtained using principal components extraction with varimax rotation. F1=Performance-approach, F2=Avoidance-approach, F3=Mastery approach

Confirmatory Factor Analysis

Results of the confirmatory factor analysis show that the data did not satisfy most of the conventional criteria for a good fitting model. The CFA confirms the finding in the earlier factor analysis that the items for the mastery-avoidance and performance scales loaded on a single latent variable. With a sample size of $N=341$ the fit of the model was adequate, $\chi^2/df=3.54$, $RMSEA=.07$, $CFI=.83$, and $TLI=.78$.

Achievement Goals as Predictors of Intrinsic Motivation

The regression summary in Table 5 shows that both mastery-approach [$F(4.677) = 36.41, p < .01 (\beta = .33)$] and performance-approach goals ($\beta = .15$) were positive predictors of intrinsic motivation, while mastery-avoidance goals ($\beta = -.12$) were negative predictors of intrinsic motivation. These apply that students who pursue approach goals orientation tend to be intrinsically motivated in their Math class while students who are trying to avoid intrapersonal incompetence (mastery-avoidance goals) tend to have lower levels of intrinsic motivation.

Table 5
Achievement Goals as Predictors of Intrinsic Motivation

	Beta	SE	B	SE	t	p
Mastery-approach	.33**	.038	.37	.04	8.68	.001
Mastery-avoidance	-.12**	.038	-.10	.032	-3.05	.002
Performance-approach	.15**	.040	.12	.031	3.81	.001
Performance-avoidance	.06	.041	.054	.034	1.56	.120

** $p < .01$

Discussion

The research was conducted to investigate the applicability of the 2 x 2 achievement goal framework on Filipino students. Results from the item analysis provided a relatively strong support that the items of the achievement goal measure have good discrimination indices and was internally consistent; although showing lower coefficients compared to the findings of Eliot and McGregor (2001) and Cury et al., (2006), where reliability coefficients ranged from .83 to .92, and from .89 to .91, respectively. Nonetheless, reliability coefficients and average inter-scale correlations fell within a moderate range.

The analysis shows that all the scales are correlated with each other, even if the scales do not share the same definition or valence, as the model posits. It suggests that students do not only adopt multiple goals, but also goals that are contrary to one another. Thus, it seems students cannot differentiate approach from avoidance goals. The results of the factor analysis revealed that all the avoidance items (mastery and performance) loaded on a single latent construct, but confirmatory factor analysis did not support this model. In the achievement goal literature, it was mastery-avoidance that has likely have been overlooked until the

inception of the 2 x 2 framework came. Most assume that mastery goals represent an approach form of regulation. In this study, mastery avoidance, in which incompetence is the focal point of regulatory attention, was validated not as an independent achievement goal.

Intercorrelations of the achievement goals revealed two unexpected results. Mastery-approach was found to be correlated with performance-avoidance and mastery-avoidance correlated with performance approach. This implies that a significant portion of the sample respondent are concern with performing and doing better than others and are also focused on worrying about learning very little of the materials presented in the class, or even at the same time trying to avoid doing poorly in the class. There are also students who are focused on attaining mastery of the lesson and are also concerned with avoiding not performing poorly in the class (Pintrich, 2000). These situations might have contributed to the observed intercorrelations of the variables that theoretically should not correlate with one another.

The confirmatory factor analysis revealed non-fit of the data to the 2 x 2 achievement goal framework. Theoretically, achievement goals that share the same definition or valence would load on a single latent variable. In this case, however, items on a positive valence (mastery-approach) and normative definition (performance-avoidance) load on a single factor. Reliability coefficients revealed an internal consistency of the measure but factor analytic procedures did not provide valid indices of the fitness of the measures on the 2 x 2 framework. Nevertheless, mastery-approach goals have been documented to be highly related with intrinsic motivation, and are found to be a positive predictor of the intrinsic motivation.

The finding indicated that the 2 x 2 framework did not operate distinctly among the samples used in this study. It is however premature to conclude that the framework is not cross-culturally valid. The study only involved samples from two secondary schools instead of examining and combining and/or comparing different samples. The limitations of language used might also have caused the non-satisfaction of the hypothesis. Though English is the medium of instruction used in the class, Filipino is mostly used as medium of communication. The translation was not considered because the psychometric properties of the measure would be different, and establishing equivalence is beyond the scope of this research. It is therefore very imperative to conduct further investigation to validate the finding of this research. In so far as the measure is concerned, the items showed considerably favorable psychometric properties like reliability coefficients. It is recommended, however, to find out if translations in the existing goal questionnaire would account for the generality of the framework. Despite the weak assumption, that the achievement goal framework works well with Filipino, the researcher believes that the measure can still be used for the purpose of differentiating the achievement goals of individual student so that we can help them realize their goals and at the same time, perform at high levels.

References

- Ames, C. (1984a). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Education Psychology, 76*, 478-487.
- Ames, C. (1984b). Competitive, cooperative, and individualistic goal structures: A motivational analysis. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (pp. 177-207). San Diego, CA: Academic Press.
- Ames, C. (1992a). Classrooms: Goals, structures, and student motivation, *Journal of Educational Psychology, 84*, 261-271.
- Ames, C. (1992b). Achievement goals and classroom motivational climate. In J. Meece & D. Schunk (Eds.), *Students' perceptions in the classroom* (pp. 327-348). Hillsdale, NJ: Erlbaum.
- Bernardo, A. B. I. (2003). Do Filipino youth really value education? Exploring Filipino adolescents' beliefs about the abstract and pragmatic value of education and its relationship to achievement goals and learning strategies. *Philippine Journal of Psychology, 36* (2), 49-67.
- Cury, F., Elliot, A., Fonseca, D., & Moller, A. (2006). The social-cognitive model of achievement and the 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 90*, 666-679.
- Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.
- Dweck, C., & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.
- Elliot, A. (1997). Integrating the "classic" and "contemporary" approaches to achievement motivation: a hierarchical model of approach and avoidance achievement motivation. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (pp.143-179). Greenwich, CT: JAI Press.
- Elliot, A. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologists, 34*, 169-189.
- Elliot, A., & Harackiewicz, J. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 76*, 628-644.
- Elliot, A., & Church, M. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218-232.
- Elliot, E., & Mc Gregor, H., (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519.
- Harackiewicz, J., & Elliot, A. (1993). Achievement goals and intrinsic motivation. *Journal of Personality and Social Psychology, 65*, 904-915.
- Harackiewicz, J., Barron, K., & Elliot, A. (1998). Rethinking achievement goals: When they are adaptive for college students and why? *Educational Psychologists, 33*, 1-21.

- Harackiewicz, J., Barron, K., Pintrich, P., Elliot, A., & Trash, T. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*, 638-645.
- Maehr, M. (1983). On doing well in science: Why Johnny no longer excels; Why Sarah never did. In E. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 179-210). Hillsdale, NJ: Erlbaum.
- Meece J., & Holt, K. (1993). A pattern analysis of students' achievement goals. *Journal of Educational Psychology, 85*, 582-590.
- Mc Gregor, H., & Elliot, A. (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Personality and Social Psychology, 94*, 381-395.
- Nicholls, J., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: Individual and classroom differences. *Journal of research in Mathematics Education, 21*, 109-122.
- Pintrich, P. (2000b). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*, 544-555.
- Rhodewalt, F. (1994). Conceptions of ability, achievement goals, and individual differences in self-handicapping behavior: On the application of implicit theories. *Journal of Personality, 62*, 67-85.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.
- Witkow, M., & Fulgini, A. (2007). Achievement goals and daily school experiences among adolescents with Asian, Latino, and European American backgrounds. *Journal of Educational Psychology, 99*, 584-596.



A confirmatory factor analysis of scores on Inventory of School Motivation (ISM), Sense of Self Scale, and Facilitating Conditions Questionnaire (FCQ): A study using a Philippine sample

Fraide A. Ganotice, Jr.
Palawan State University, Philippines

Abstract This study was conducted to explore the validity of Inventory of School Motivation (ISM), Sense of Self Scale (SOS), and Facilitating Conditions Questionnaires (FCQ), developed by Dennis McInerney. A Confirmatory Factor Analysis (CFA) was conducted on the results of the items in the questionnaires involving 378 high school Filipino students. The ISM has four dimensions: mastery, performance, social, and extrinsic goals; while the SOS has also four dimensions: sense of purpose, sense of reliance, negative self esteem, and positive self esteem. On the other hand, FCQ is composed of eleven dimensions: positive peer influence, school valuing, parent support, teacher support, peer help, leave school, pride from others, negative parent influence, affect to school, negative peer influence, and university intention. Based on the responses, the internal reliability of the instruments were all adequate. The results indicate that confirmatory factor analyses conducted for the ISM, Sense of Self, and FCQ scales indicated a good fit between the models and the data using the goodness-of-fit index. Thus, the CFA approach used in this study provided a strong validation, with the Filipino sample. Discussions focused on the possible research explorations in the future involving bigger Filipino participants with the use of the said validated instruments.

Keywords: *Motivation, sense of self, facilitating conditions*

Introduction

Interest in cross-cultural comparisons has grown over the last 20 years (Van de Vijver & Hambleton, 1996). Seemingly, researchers do not only end their investigations with single culture but extend these to include cross-cultural studies in order to account for diversity and convergences of other cultures. With the advent of cross-cultural research, validation of instrument has become an important methodological aspect of cross-cultural research studies. Research on samples from different cultural groups need to consider whether the scores obtained are

comparable. Equivalence and bias or differential item functioning are important issues that need to be addressed before meaningful cross-cultural comparisons can be made (Van de Vijver & Leung, 1997; Fisher, 2004).

Agreeably, the application of an instrument in a new cultural group involves more than simply producing text in another language, administering the translated instrument, and comparing the results. The following questions may serve as reminders for reflection before cross-cultural researchers start to cross the boundaries of other cultures specifically with instrumentation: “Does the construct apply to the target group or does it show ethnocentric bias? Are the behaviors associated with the construct similar in the source and target groups? Is the measurement procedure (e.g., stimulus and response format) adequate for application in the target groups? Are the language loadings within the vocabulary level of the target groups? Is the instrument free from item bias?” These are only some of the questions that need to be addressed in order to minimize the impact of bias in cross-cultural measurement.

While it is an acceptable practice to adopt and/or adapt a foreign-made instrument, caution is necessary on its application if it is to be used with other groups other than its intended population (Hambleton, 2001). Validation of the instrument is an essential phase designed to ascertain the psychometric properties of a foreign-made test when applied to local setting. In the Philippine context, students’ cultural backgrounds abound that need to be considered when a researcher does cross-cultural investigation. It is therefore necessary for him/her to be reminded of the importance of validation of the instruments when it is to be used within the confine of Philippine classrooms.

One of the foreign-made instruments whose psychometric properties need an evaluation when applied to the Philippine setting is the Inventory of School Motivation (ISM). As McInerney (1991) stressed, this instrument is used to define scales that are relevant to examining students’ motivation in cross-cultural contexts. Hence, if it is to be used in other countries like the Philippines, it is imperative to conduct further validation study in order to shed light on issues concerning its validity and applicability to the distinct Filipino culture.

Together with the ISM, McInerney and his group developed the Sense of Self scale and Facilitating Conditions Questionnaire to complete the set of questionnaires. These instruments were tailored fit within the framework of Personal Investment Theory (Maehr & Braskamp, 1986; see also Maehr & McInerney, 2004), one of the interesting theories on motivation concerned with the significant role played by social and cultural contexts in determining motivational patterns in performing achievement tasks of learners. Hence the personal investment theory seems very well suited to provide a theoretical framework for comparing between-group differences across social and cultural groups. Specifically, the theory designates three basic components of meaning as critical to determining personal investment in specific situations: (1) Sense of self, (2) perceived goals of behavior in given situations, and (3) perceived facilitating conditions for pursuing these goals. Personal Investment (PI) theory is particularly relevant in investigations into how individuals of varying social and cultural backgrounds relate to differing achievement situations. This is because it does not assume that people from a given culture or group will invest effort in the same

achievement situations or, if they do, for the same reasons, as those from other cultures and groups. PI theory also emphasizes the role played by social and cultural contexts in determining motivational patterns in performing achievement tasks.

Specifically, the Inventory of School Motivation (McInerney, Roche, & Marsch, 1997; McInerney & Sinclair, 1991; 1992) was designed as an exploratory instrument through which a range of motivation-salient constructs drawn from Maehr's personal investment model could be identified in educational settings across diverse groups. In addition to this, because performance goals and achievement are other referenced, self-worth is determined by an individual's perception of ability to perform and to compete successfully. Hence, when a student tries hard without being completely successful (in terms of established norms) his/her sense-of-self worth is threatened. Guided by this framework, McInerney (in press) included Sense of Self scale which is the second part of ISM. On the other hand, because of the potent impact of significant others in the motivation of students, the Facilitating Conditions Questionnaire (FCQ) was constructed with the assumption that although individuals' internal motivations (such as their goals and expectancies) may promote positive academic behavior and achievement, external (or at least externally referenced) forces in students' social environments may facilitate or inhibit translation of these internal motivations into actual behavior (McInerney, Dowson, & Yeung, 2005).

The ISM has been validated in countries involving sample Australian, American, African, Chinese students (McInerney & Ali, 2006). In the Philippines, the reliability of the ISM has been initially explored by Watkins, McInerney, & Boholst (2003) utilizing 86 fourth year Filipino secondary school students in the central Philippines. Generally, the results of the research support the internal consistency reliability and construct validity of responses to the ISM by the said 86 students. Their study ended with a recommendation for a confirmatory factor analysis based on the responses to the ISM by a much larger sample of Filipino students that is expected to strengthen their claim.

In a separate study involving elementary and high school students in Arizona, McInerney, Dowson, and Yeung (2005) explored the construct validity and applicability of the Facilitating Conditions Questionnaire (FCQ). The result of their study supports the FCQ's applicability to these elementary and secondary students. The goodness-of-fit and alpha reliability measures for the models tested with the whole sample provided preliminary confidence that the FCQ may yield valid scores with both the elementary and the secondary subsamples.

Although the psychometric properties of ISM and FCQ scales have been examined to some extent through exploratory factor analysis, stronger validation of the instrument requires an approach such as confirmatory factor analysis using samples from other populations. The utility of the said instruments may encompass the Philippine landscape where, like in the case of other countries, people's cultures are diverse. However, before these instruments can be used for research purposes, it is necessary to verify the psychometric usefulness, factor structures among others which are prerequisite for an appropriate interpretation of the ISM and FCQ scores in a diverse culture. Specific to the present study, this reports if the said instruments capture the array of tests constructs, which the

authors have established using other populations, when applied to the Philippine setting. Moreover, this study tested the factor structures of the mentioned instruments using the Philippine student samples through the use of confirmatory factor analysis.

Overview of the Present Study

The purpose of this study was twofold: (a) to conduct a confirmatory factor analysis to assess the factor structure of the scale items obtained from 43-item Inventory of School Motivation (ISM); 26-item Inventory of School Motivation: Sense of Self Scales; and 55-item Facilitation Conditions Questionnaire (FCQ); and (b) to investigate the estimates of internal consistency and construct validity of the ISM and FCQ scores.

Method

Participants

Participants of the study were 378 (160 males and 218 females) first year to fourth year high school students in public and private schools in Metro Manila and a laboratory school in the province. The age of the participants ranged from 12 to 20 years old, with a mean age of 14.62 and *SD* of 1.39. These students were assured that their responses would remain confidential and would not, in any way, influence their grade.

Measures

For all the 124 items considered in this study, a 5-point Likert-type scale with a response scale ranging from 1 (strongly agree) to 5 (strongly disagree) was used. The reliability, mean and standard deviation of each scale (coefficient alpha) in the present study are presented in Table 5.

Inventory of School Motivation (ISM). There are four dimensions of ISM which include mastery goal, performance goal, social goal, and extrinsic goal. These achievement goals were assessed with 43 items of the ISM. Achievement goals are cognitive representations of the different purposes that students may have in different achievement situations and are presumed to guide students' behavior, cognition, and feelings as they become involved in academic work (McInerney, et al., 1997).

Mastery goal is defined as the degree to which a student is motivated by intrinsic self-referenced goals. There are 11 items which comprised the mastery goal (Cronbach's $\alpha=.81$). It is further grouped into task (e.g., "I like being given the chance to do something again to make it better.") and effort (e.g., "I don't mind working a long time at school that I find interesting.")

Performance goal is the degree to which a student is motivated by competitive other-referenced goals (Cronbach's $\alpha=.81$). This dimension of ISM has 12 items distributed between competition (e.g., "Winning is important to me.");

and social power (e.g., “I work hard at school that I will be put in charge of the group.”).

In addition, Extrinsic goal is the third dimension of the ISM (Cronbach’s $\alpha=.88$). This is defined as the degree to which a student is motivated by praise and rewards. There are 12 items which are measured this dimension as praise (e.g., “Praise from my teachers for my good schoolwork is important to me.”); and token (e.g., “I work best in class when I can get some kind of reward.”). Social goal is the fourth dimension of the ISM (Cronbach’s $\alpha=.73$). This is described as the degree to which a student is motivated by competitive other-referenced goals. Social goal is composed of 8 items distributed into social concern (e.g., “It is very important for students to help each other at school.”); and affiliation (e.g., “I do my best work when I am working with others.”). The described composition of ISM questionnaire is based on the hierarchical, multidimensional model of goal orientations designed by McInerney and Yeung (2000).

Inventory of School Motivation. Sense of self refers to the more or less organized collections of perceptions, beliefs, and feelings related to who one is. It is presumed to be composed of a number of components such as sense of competence, sense of autonomy, and sense of purpose, each also contributing to the motivational orientation of the individual and interacting with the motivational previously described. There are 4 sense-of-self scales, for a total of 26 items. They include the following dimensions: sense of purpose (Cronbach’s $\alpha=.85$), sense of reliance (Cronbach’s $\alpha=.84$), negative self esteem (Cronbach’s $\alpha=.79$), and positive self esteem (Cronbach’s $\alpha=.76$).

Sense of purpose is defined as the degree to which a student values school for the future (e.g., “I try hard to do well at school so that I can get a good job when I leave.”). Sense of purpose is measured by 6 items. The second dimension of sense of self scale is sense of reliance defined as the degree to which a student is self-reliant and confident within academic settings (e.g., “I do not need anyone to tell me to work hard at school”). Eight items measured this dimension. The third dimension is negative self esteem which is the degree to which the student holds negative feelings about his/her general academic ability at school (e.g., “I am always getting into trouble at school.”), composed of 5 items. The last dimension is positive self esteem, the degree to which a student holds positive feelings about his/her general academic ability at school (e.g., “I think I am as good as everybody else at school.”) measured by 5 items.

Inventory of School Motivation: Facilitating Conditions Questionnaire (FCQ). Research on school motivation has assumed that individual students’ personal goals, expectancy and values tend to have considerable influence on their academic behaviors and academic achievement. McInerney (1988; 1989; 1991; 1992), however, proposed that whereas the individual’s internal drive may promote academic behavior and achievement, there are external forces in the school environment that may facilitate or inhibit the translation of these motivational forces into actual behavior. To examine the potential impacts of these external environmental factors, McInerney (1988; 1989; 1991; 1992) designed a Facilitating

Conditions Questionnaire (FCQ) based on Maehr's (1984) hypothesis of action possibilities (also see Maehr & Braskamp, 1986). There are 11 scales and 55 items which composed the FCQ.

University intention is the first dimension of FCQ. It is defined as the degree to which a student perceives that he/she can complete college (e.g., "I am the kind of person who can complete college or university"; Cronbach's $\alpha=.88$). This is measured by 5 items. The second dimension is school valuing which deals with the importance of education to students (e.g., "Education is important to me to get a job; Cronbach's $\alpha=.88$). School valuing is measured by 9 items. In addition, there are 6 items that measured parent support which is defined as the degree of a student's perception of positive parental support (e.g., "My mother helps me with my schoolwork"; Cronbach's $\alpha=.88$). Six items measured teacher support, a student's perception of positive teacher support (e.g., "My teachers help me with my schoolwork"; Cronbach's $\alpha=.77$). In addition to this, there are 5 items for peer help, the extent to which a student perceived that his/her peers help him/her (ex. "Some of my friends help me with my schoolwork"; Cronbach's $\alpha=.70$). Four items measured student's perception of influences on leaving school (e.g., "My mother doesn't mind If I leave school when I want to"; Cronbach's $\alpha=.90$). There are 4 items that measured pride from others, the importance of a student of pride from others in his/her school achievement (e.g., "It's important for my father to be proud of my school work"; Cronbach's $\alpha=.84$). Moreover, there are 5 items measured negative parent influence, a student's perception of negative parent support (e.g. "My father doesn't pay any attention when I bring home report cards"; Cronbach's $\alpha=.90$). There are 3 items measuring affect to school, which describes the inherent value given by a student to school (e.g., "I like studying"; Cronbach's $\alpha=.76$). Additional 4 items measured negative peer influence, a student's perception of negative peer support (e.g., "Some of my friends tell me I should leave school when I can"; Cronbach's $\alpha=.84$); and the remaining 4 items measured positive peer influence, the degree a student perceives their peers value schooling (e.g., "Most students in my class will go on to college or university"; Cronbach's $\alpha=.91$).

Administration

With the intent of standardizing the administration of the questionnaires, teachers and guidance staff who administered the questionnaires had received a copy of the instrument in advance, along with the brief written account of the nature, purpose, and specific instructions of the administration. Those who administered the questionnaires were instructed not to interpret any of the items for the students, instead to advise the students to give their best judgment on the items.

Confirmatory Factor Analysis (CFA)

CFAs assess the extent to which the observed indicators (items) reflect the structure of the underlying constructs. CFAs allow the researcher to specify not only how many factors are measured by a given set of items but, also, which items

function as indicators of which factors. This analysis is conducted to test hypotheses or confirm theories about the factors one expects to find (Vogt, 1999). Relative to the present study involving high school Filipino students, CFA will assess the factor structure of the scale items from 43-item Inventory of School Motivation (ISM); 26-item Inventory of School Motivation: Sense of Self Scales; and 55-item Facilitating Condition Questionnaire (FCQ).

Goodness-of-Fit Indicators

Although a large number of Goodness-of-Fit Indexes (GFIs) are available, the relative performance of them has left considerable room for debate (Huang & Michael, 2000). The multiple indexes recommended by Hu and Bentler (1995) cited in the article by Hoyle and Panter (1995) are reported in this investigation.

Multiple indices provided a comprehensive evaluation of model fit (Hu & Bentler, 1995; 1999). I examined the traditional chi-square fit index. However, given the known dependency of the chi-square statistic on sample size (Bentler & Bonett, 1980; Byrne, 1994), and that the chi-square values are overly stringent in evaluating exact fit, I also examined other fit indices. In the present study, the following goodness-of-fit measures were also used: (a) Comparative Fit Index (CFI), (b) Bentler-Bonett Normed Fit Index (NFI), (c) Bentler-Bonett Nonnormed Fit Index (NNFI; also known as Tucker-Lewis Index), (d) Incremental Fit Index (IFI), and (e) Goodness-of-Fit Index (GFI). Although a value of .90 for CFI, NFI, NNFI, IFI and GFI has served as a rule-of-thumb lower limit cutoff of acceptable fit, a value of .93 is expected of models considered to be well-fitting (Byrne, 1994). RMSEA values of less than .06 indicate a good fit, and values as high as .08 indicate a reasonable fit (Hu & Bentler, 1999).

Confirmatory Factor Analysis Procedure

As reported earlier, the ISM is composed of 43 items with eight to 12 indicators of each factor (see Table 1). Bagozzi and Heatherton (1994) concluded that when there are more than four or five indicators per factor in a large sample, it is quite likely to lead to an unsatisfactory fit in the measurement model. To address this issue, I aggregated the items to form item "parcels" as indicators in the CFA. Parcels are formed by randomly combining two to four items in each scale (Bernardo et al., in press). For the first factor of ISM composed of 11 items, the first item parcel was comprised by the following items: 1, 4, 7, and 10. The second item parcel includes items 2, 5, 8, and 11 and the third parcel was composed of items 3, 6, and 9. There were 11 parcels which were created: 3 parcels from 11 items of the first factor; 3 parcels from 12 items of the second factor; 2 parcels from 8 items of the third factor; and another 3 parcels from 12 items of the fourth factor. This procedure was conducted for the rest of the factors of the ISM. Likewise, in the CFA analyses for ISM: Sense of Self and Facilitating Conditions Questionnaire (FCQ), this procedure was carried out.

In order to test the validity of the ISM scale, a CFA was performed with the following model. Each of the four scales (mastery, performance, social, and extrinsic goals) for the ISM instrument, and another four scales (sense of purpose,

self reliance, negative self esteem, and positive self esteem) for ISM: SOS instrument as latent variables, and the manifest variables are the respective item parcels generated from the array of ISM items linked to the latent variables. Likewise, the four scales (latent variables) were correlated in the model. Structural equation modeling (SEM) was used in this study to test the model in terms of how the hypothesized model fits to the observed data (Tabachnick & Fidell, 2001). The STATISTICA 8 software was used for the entire statistical analysis. As presented earlier, various goodness-of-fit indexes were used in this study. They include: Chi-square (χ^2), NNFI, CFI, GFI, and NFI.

Results

Table 1 presents the various indexes of fit for the three measures considered in this study - the Inventory of School Motivation (ISM), Inventory of School Motivation: Sense of Self, and the Facilitating Conditions Questionnaire (FCQ).

Table 1
Goodness-of-Fit Indices

Model	χ^2	<i>df</i>	NNFI/TLI	CFI	GFI	NFI
ISM	56.086	38	0.988	0.992	0.975	0.976
ISM:SOS	28.113	14	0.974	0.987	0.982	0.974
ISM:FCQ	154.889	85	0.967	0.982	0.956	0.961

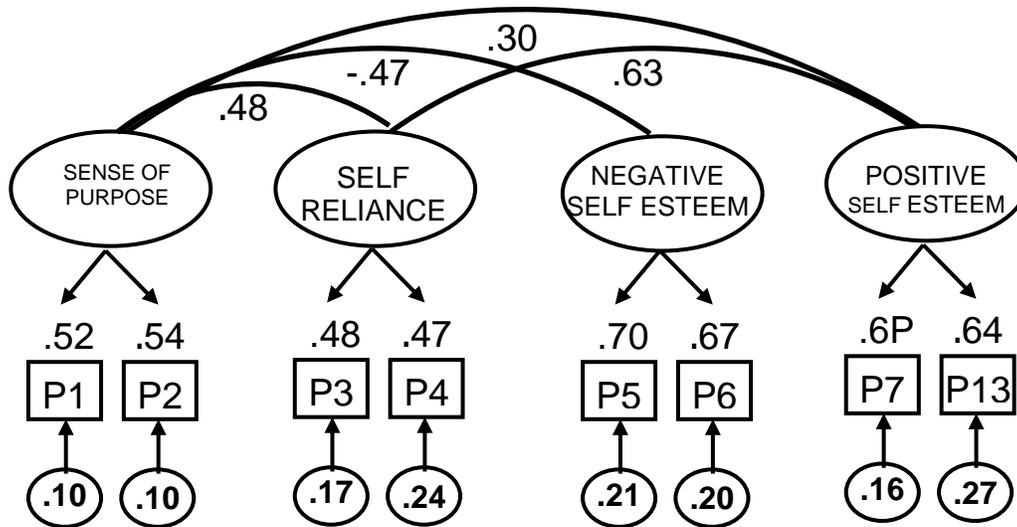
Note: ISM=Inventory of School Motivation; SOS=Sense of Self Scale; FCQ=Facilitating Conditions Questionnaire; NNFI=Non-Normed Fit Index; TLI=Tucker-Lewis Index; CFI=Comparative Fit Index; GFI=Goodness-of-Fit Index; IFI=Incremental Fix Index; NFI= Bentler-Bonett Normed Fit Index.

Confirmatory Factor Analysis (CFA) for Inventory of School Motivation (ISM)

As contained in Table 1, all of the fit indexes met the standard criterion for interpretation: NNFI=.988, CFI=.992, GFI=.0.975, and NFI=.976. These values are considered to be well fitting (Bryne, 1994). The data were a good fit for the four-factor model (please refer to Figure 2). Of all the goodness-of-fit indicators considered in this study, only the chi-square [$\chi^2 (38) = 56.086, p < .0001; \chi^2/df = 1.48$] was not adequate. A significant value of χ^2 suggests that entries for the proposed model deviate from those obtained. On the other hand, a statistically nonsignificant value of χ^2 suggests that a model may be a reasonably satisfactory representation of the data. As discussed by Anderson and Gerbing (1988), and Huang and Michael (2000), however, that the value of the chi-square likelihood ratio statistic is directly dependent on sample size. Because of this, with large sample size, significant values can be obtained even though there are only trivial discrepancies between the model and the data. On the other side, a small sample size would result in a decision to accept a model in spite of the presence of

substantial disagreement between the hypothesized entries for the model and those associated with data obtained.

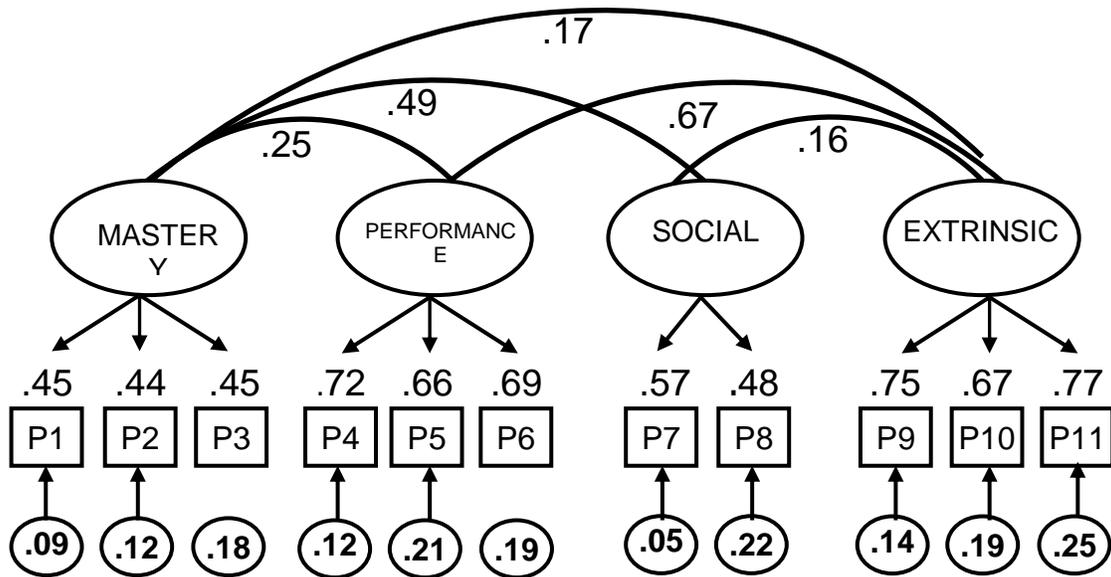
Figure 1
Confirmatory Factor Analysis of 11 Parcels of Inventory of School Motivation (ISM) with 4 Factors



Note. Confirmatory Factor Analysis (CFA) for Inventory of School Motivation: Sense of Self Scale. Inter-Factorial Correlations Indicated are Significantly Different from Zero at $p < .05$

The fit of the Inventory of School Motivation: Sense of Self scale was likewise satisfactory. Although the value of χ^2 was statistically significant, the entries of NNFI, CFI, GFI, and NFI were .97, .99, .98, and .97 respectively and were interpreted as well fitting. The data fits the 4-factor model of the Inventory of School Motivation.

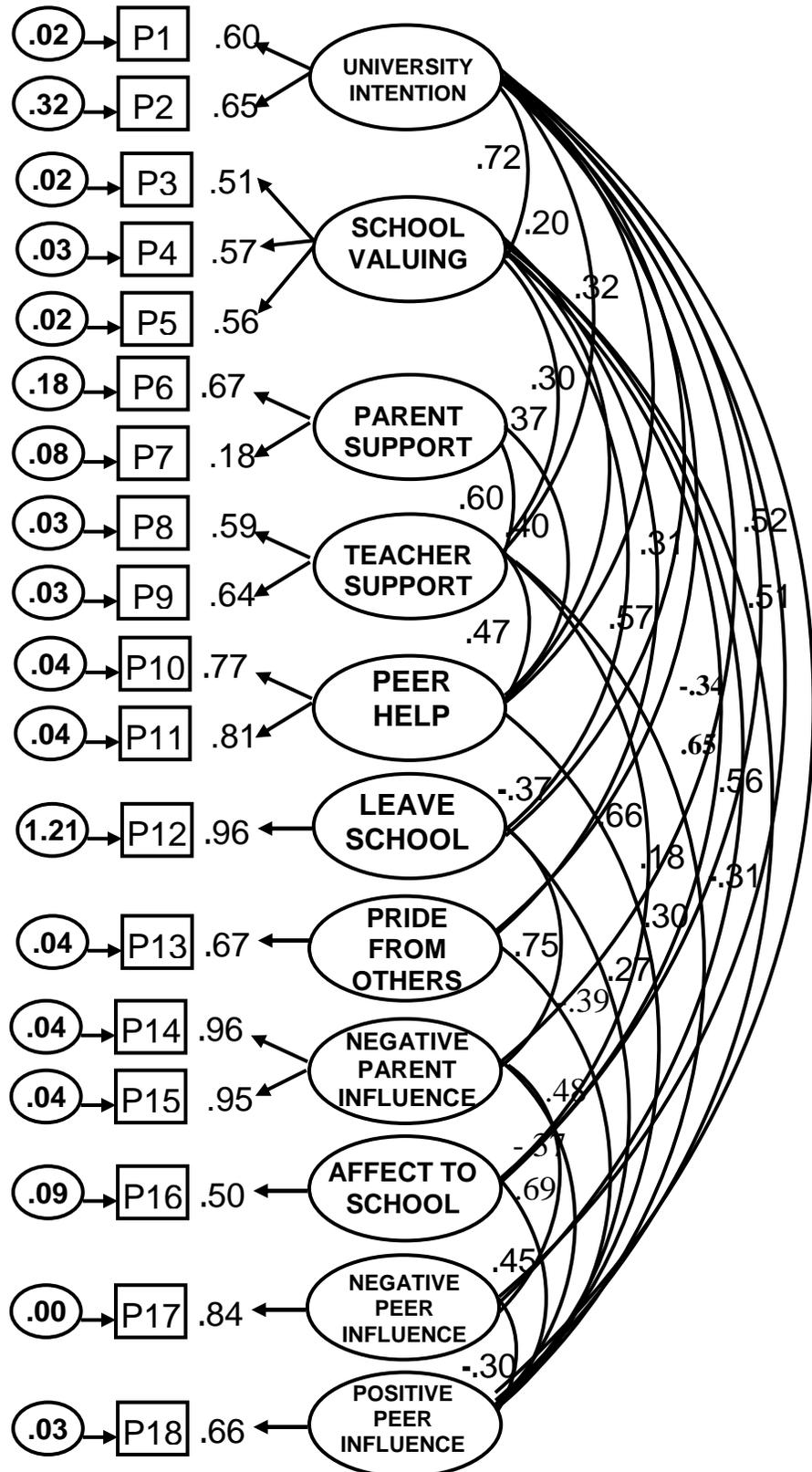
Figure 2
Confirmatory Factor Analysis of 8 Parcels of Inventory of School Motivation: Sense of Self with 4 Factors



Note. Confirmatory Factor Analysis (CFA) for Facilitating Conditions Questionnaire Scale. Inter-Factorial Correlations Indicated are Significantly Different from Zero at $p < .05$

Like the results for the ISM and ISM: Sense of Self Scales, the data on a separate CFA performed on the FCQ reflects that the data were a good fit with the model (see Figure 4). The data fit for the Facilitating Conditions scale was satisfactory. Although the value of χ^2 was statistically significant, the entries of NNFI, CFI, GFI, and NFI were .97, .98, .96, and .96 respectively and were interpreted as well fitting.

Figure 3
 Confirmatory Factor Analysis of 18 Parcels of Facilitating Conditions Questionnaire with 11. Inter-Factorial Correlations Indicated are Significantly Different from Zero at $p < .05$



Correlations among Latent Variables of Inventory of School Motivation (ISM)

Table 2 and Figure 2 reflect the data on the correlations of the 4 factors for the ISM scale. The correlations among the factors of the ISM indicate that mastery goal is positively correlated with performance goal ($r=.25$); social goal ($r=.49$) and extrinsic goal ($r=.17$). The same significant positive correlation happened between performance goal and extrinsic goal ($r=.67$) and between social goal and extrinsic goal ($r=.16$).

Correlations among Latent Variables of Inventory of School Motivation: Sense of Self Scales

The factors of Sense of Self questionnaire were correlated as well. Among these four factors, sense of purpose has significant positive correlation with self reliance ($r=.43$) and positive self esteem ($r=.25$) but has significant negatively correlation with negative self esteem ($r=-.24$). In addition to this, sense of reliance is positively correlated with positive self esteem ($r=.50$).

Correlations among Latent Variables of Facilitating Conditions Questionnaire

The 11 factors of FCQ were also correlated with one another. As expected, a pattern of results on the correlation coefficients can be traced from Table 4. An example of this is between university intention and school valuing ($r=.72$) where it may be inferred that those students who may have clear intention to complete college would tend to value school more than those who do not want to pursue college. Conversely, negative parent influence is negatively correlated with university intention ($r=-.27$). The complete correlation data on FCQ is contained in Table 4.

Table 2
Zero-Order Correlations among Latent Variables of Inventory of School Motivation

<i>Inventory of School Motivation (ISM)</i> <i>Factors</i>	1	2	3	4
1. Mastery Goal	-			
2. Performance Goal	.25*	-		
3. Social Goal	.49*	.08	-	
4. Extrinsic Goal	.17*	.67*	.16*	-

* $p < .05$

Table 3
Zero-Order Correlations among Latent Variables of Inventory of School Motivation (ISM): Sense of Self

<i>Inventory of School Motivation (ISM): Sense of Self Factors</i>	1	2	3	4
1. Sense of Purpose	-			
2. Sense of Reliance	.43*	-		
3. Negative Self Esteem	-.24*	-.04	-	
4. Positive Self Esteem	.25*	.50*	-.02	-

* $p < .05$

Table 4
Zero-Order Correlations Among Scales/Latent Variables of Facilitating Conditions Questionnaire (FCQ)

Facilitating Conditions Questionnaire (FCQ) Factors	1	2	3	4	5	6	7	8	9	10	11
1. University Intention	-										
2. School Valuing	.72*	-									
3. Parent Support	.09	.23*	-								
4. Teacher Support	.20*	.37*	.60*	-							
5. Peer Help	.32*	.30*	.40*	.47*	-						
6. Leave School	-.37*	-.35*	.23*	.16	-.07	-					
7. Pride From Others	.31*	.57*	.43*	.53*	.42*	-.14	-				
8. Negative Parent Influence	-.27*	-.37*	.01	.01	-.09	.75*	-.29*	-			
9. Affect to School	.30*	.65*	.73*	.66*	.20*	-.07	.53*	.05	-		
10. Negative Peer Influence	-.31*	-.34	.07	.14	-.01	.84*	-.08	.69*	-.15	-	
11. Positive Peer Influence	.52*	.56*	.08	.18*	.27*	-.39*	.48*	-.37*	.45*	-	-
											.30*

* $p < .05$

Internal-Consistency Estimates of Reliability

In this research, the internal consistency reliability estimates, the coefficient alphas (α), for responses for ISM specific motivation scales ranged from α .73 (social) to .88 (extrinsic). All these values surpass the generally accepted satisfactory level of .70 (Watkins, McInerney, & Boholst, 2003). Alphas for ISM Sense of Self Scales varied from .76 to .85 with positive self esteem having the lowest alpha of .76 and sense of purpose having the highest alpha of .85. The ISM Facilitating Conditions Questionnaire alike has factors coefficient alpha which surpassed the

satisfactory level of .70 with peer help obtaining the lowest alpha of .70 and positive peer influence having the highest coefficient alpha of .91.

Table 5
Alpha, Mean, and Standard Deviations of the Measures

<i>Inventory of School Motivation (ISM)</i>			
Factors	Alpha (α)	<i>M</i>	<i>SD</i>
1. Mastery Goal (11 items)	.81	1.90	.49
2. Performance Goal (12 items)	.86	3.03	.73
3. Social Goal (8 items)	.73	2.11	.56
4. Extrinsic Goal (12 items)	.88	2.70	.77
<i>Inventory of School Motivation: Sense of Self Scales</i>			
1. Sense of Purpose (6 items)	.85	1.63	.58
2. Self Reliance (8 items)	.84	3.66	.86
3. Negative Self Esteem (7 items)	.79	3.31	.73
4. Positive Self Esteem (5 items)	.76	2.75	.71
<i>Facilitating Conditions Questionnaire (FCQ)</i>			
1. University Intention (5 items)	.88	1.54	.65
2. School Valuing (9 items)	.88	1.60	.58
3. Parent Support (6 items)	.88	2.75	.95
4. Teacher Support (6 items)	.77	2.47	.67
5. Peer Help (5 items)	.70	2.17	.82
6. Leave School (4 items)	.90	4.08	1.05
7. Pride From Others (4 items)	.84	2.13	.82
8. Negative Parent Influence (5 items)	.90	4.09	.97
9. Affect to School (3 items)	.76	2.32	.80
10. Negative Peer Influence (4 items)	.84	1.82	.75
11. Positive Peer Influence (4 items)	.91	3.88	1.04

Discussion

This study examined the factor structures of scores on the 43-item Inventory of School Motivation (ISM); 26-item Sense of Self Scales; and 55-item Facilitation Conditions Questionnaire (FCQ) which were all constructed by Dennis M. McInerney and group. It also examined the estimates of internal consistency and construct validity of the said instruments.

The results, with the 378 sample Filipino high school students, indicate that confirmatory factor analyses conducted for the ISM, Sense of Self, and FCQ scales indicated a good fit between the models and the data using the goodness-of-fit index. Thus, the CFA approach used in this study provided a strong validation, with the Filipino sample, of the Inventory of School Motivation (ISM), Inventory of School Motivation: Sense of Self, and Facilitating Conditions Questionnaires

constructs. Specifically, the data were a good fit for the four-factor model of ISM, four-factor of Sense of Self scale, and eleven-factor Facilitating Conditions Questionnaire.

The internal reliability of the complete ISM, Sense of Self items, and FCQ scales were adequate reaching acceptable levels, with Cronbach alpha values meeting the criterion. These results provided us with confidence that these instruments may yield valid scores in the Philippine setting even if these have been used mostly in Western countries.

In comparison with the previous work of Watkins, McInerney, and Boholst (2003) involving 86 fourth year Filipino students on the validity and reliability of ISM where they found that reliance, self esteem and task not meeting the satisfactory level based on coefficient alpha, the present study revealed otherwise. In this study, the 3 mentioned factors of ISM obtained .84, .79 and .81 alpha coefficients respectively.

Reflecting on the results, the present findings of the CFA conducted on Inventory of School Motivation (ISM) are somehow supported with the studies conducted by Bernardo involving Filipino students (see Bernardo, 2003; 2004; 2005; 2007 for a review). Specifically, in some of his studies in the Philippines, it was revealed that both mastery and performance-approach goals tend to be associated with positive learning variables such as deeper learning strategies (Bernardo, 2004) and higher achievement (Bernardo, 2005); and that interestingly, there is a significant positive association between mastery and performance goals (Bernardo, 2003), which may indicate either that the students are adopting multiple goals, or that the Filipino students are giving different meanings to these two types of goals. Somehow these empirical studies on the learning goals of Filipino students provided theoretical support on the applicability of the ISM constructs to Filipino students.

On the other hand, the result of CFA conducted on FCQ in this study is supported by the previous study of McInerney, Dowson, and Yeung (2005) involving elementary and high school students with Anglo background in the United States. This might connote that somehow the factor structures of the external facilitating conditions and support of students in the Philippines might be somewhat related, if not completely synonymous, with the group of students studied by McInerney, Dowson, and Yeung (2005) in United States particularly in Phoenix, Arizona.

In the Philippines, although there are not many available empirical studies on the impact of external support: social and environmental, provided for the Filipino students vis-à-vis motivation and academic performance, however literature are rich on the effect of significant others - facilitating conditions - on the translation of internal motivations into actual behavior and performance. Some research have demonstrated the relationship between students' perceptions of support and caring from parents, teachers, and peers and aspects of their motivation and academic achievement (Allocca & Muth, 1982; Bempechat, 1990). Research also indicates that the parental involvement in the education of their children is a good predictor of student learning and success (Fantuzzo, Tighe, & Childs, 2000; Hill, 2001; Hill & Craft, 2003). On the part of facilitating conditions provided by peers, the quality of peer interactions can have a substantial bearing on

students' academic performance and social behaviors (Fredricks, Blumenfeld, & Paris, 2004).

With all these, the validation of these scales becomes useful for researchers and even teachers, school psychologists and even counselors. This validation study ascertained the psychometric properties of a foreign-made test when applied to the Philippine setting. Hence, the mentioned professionals can now use the Inventory of School Motivation, Sense of Self Scale and Facilitating Conditions Questionnaire to better understand the uniqueness of Filipino high school students.

The completion of this instrument validation suggests various possible research explorations in the future involving bigger Filipino participants. It will be interesting to explore the variations of students' achievement motivations and the facilitating and/or inhibiting support system accorded to the Filipino students from different groups of divergent cultural backgrounds by their social network-parents, teachers, and peers among others. To find out how these facilitating conditions impact students' school motivation and learning, and to identify if such conditions are translated into academic outcome are objectives worth realizing. Moreover, another research area which can be considered using mixed-methods research design is the investigation of the achievement goals of Filipino students in comparison with Chinese, Chinese-Filipino, and Korean students who are growing in number rather quickly in the landscape of Philippine classrooms. The quantitative phase of possible study might be to explore whether the goals held by these students from diverse cultural backgrounds differ and to determine the relationship of these goals to school motivation and achievement. In order to attain deeper understanding of the quantitative results, the second phase might be the utilization of focus group discussions with students and parents. This can be a way of probing and/or triangulating the data within which to account for the valid explanation of quantitative results. A separate comparative study on the psychometric properties of the possible Filipino version of the ISM and FCQ in order to account for cross-cultural response bias can also be equally interesting. All these may be made possible now that the ISM, Sense of Self and FCQ constructs have been found valid and useful instruments within which to explore goals, sense of self measures and facilitating conditions of students in the Philippine classrooms.

References

- Allocca, R., & Muth, R. (1982). *School culture and academic achievement: A cross-gender, cross-cultural study*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411-423.
- Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modelling, 1*, 25-67.
- Bempechat, J. (1990). *The role of the parent involvement in children's academic achievement: A review of the literature*. Trends and Issues No. 14. Washington, DC: Department of Education, Office of planning, Budget, and Evaluation.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.
- Byrne, B. M. (1994). *Measuring self-concept across life span: Issue and instrumentation*. Washington, DC: American Psychological Association.
- Bernardo, A. B. I. (2003). Do Filipino youth really value education? Exploring Filipino adolescents' beliefs about the abstract and pragmatic value of education and its relationship to achievement goals and learning strategies. *Philippine Journal of Psychology, 36*(2), 49-67.
- Bernardo, A. B. I. (2004). Culturally-rooted beliefs and learning: Exploring the relationships among social axioms, achievement goals, and learning strategies of Filipino college students. *Philippine Journal of Psychology, 37*(2), 79-100.
- Bernardo, A. B. I. (2005). Exploring new areas for helping low-achieving students in the Philippines: Beliefs, goals, and strategies. In A. D. Thomas, N. Dayan, A. B. I. Bernardo, & R. Roth (Eds.), *Helping others grow* (pp. 133-142). Aachen, Germany: Shaker Verlag.
- Bernardo, A. B. I. (2007). *Individual and social dimensions of Filipino students' achievement goals*. Manuscript submitted for publication.
- Fantuzzo, J., Tighe, E., & Childs, S. (2000). Family involvement questionnaire: A multivariate assessment of family participation in early childhood education. *Journal of Educational Psychology, 92*, 367-376.
- Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology, 35*, 263-282.
- Fredricks, J. A., Blumenfeld, P. A., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*, 164-172.

- Hill, N. E. (2001). Parenting and academic socialization as they relate to school reading: The roles of ethnicity and family income. *Journal of Educational Psychology, 93*, 686-697.
- Hill, N. E., & Craft, S. A. (2003). Parent-school involvement and school performance: mediated pathways among socioeconomically comparable African American and Euro-American families. *Journal of Educational Psychology, 95*, 74-83.
- Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*. Thousand Oaks, CA: Sage.
- Hu, L. T., & Bentler, P. M. (1995). Measuring model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*. Thousand Oaks, CA: Sage.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*, 1-55.
- Huang, C., & Michael, W. B. (2000). A confirmatory factor analysis of scores on a Chinese version of an academic self concept scale and its invariance across groups. *Educational and Psychological Measurement, 60*, 772-786.
- Maehr, M. L. (1984). Meaning and motivation: Toward a theory of personal investment. In R. Ames & C. Ames (Eds.), *Research on motivation and education: Student motivation (Vol. 1)*. Orlando, FL: Academic Press.
- Maehr, M. L., & McInerney, D. M. (2004) Motivation as personal investment. In D. M. McInerney & S. Van Etten (Eds.), *Sociocultural influences on motivation and learning: Big theories revisited (Vol. 4)*. Greenwich, CT: Information Age Publishing.
- Maehr, M. L., & Braskamp, L. A. (1986). *The motivation factor: A theory of personal investment*. Lexington, MA: Lexington.
- McInerney, D. M. (1988). *The psychological determinants of motivation of urban and rural traditional and aboriginal students in school settings: A cross-cultural study*. Unpublished doctoral dissertation, University of Sydney, Australia.
- McInerney, D. M. (1989). A cross-cultural analysis of student-motivation. In D. M. Keats, D. Munro, & L. Mann (Eds.), *Heterogeneity in cross-cultural psychology*. Lisse, the Netherlands: Zwets & Zeitlinger.
- McInerney, D. M. (1991). The behavioral intentions questionnaire: An examination of construct and etic validity in an educational setting. *Journal of Cross-Cultural Psychology, 22*, 293-306.
- McInerney, D. M. (1992). Cross-cultural insights into school motivation and decision making. *Journal of Intercultural Studies, 13*, 53-74.
- McInerney, D. M., & Ali, J. (2006). Multidimensional and hierarchical assessment of school motivation: Cross-cultural validation. *Educational Psychology, 26*, 595-612.
- McInerney, D. M., Dowson, M., & Yeung A. S. (2005). Facilitating conditions for school motivation: Construct validity and applicability. *Educational and Psychological Measurement, 65*, 1-21.

- McInerney, D. M., & Sinclair, K. E. (1991). Cross cultural model testing: Inventory of school motivation. *Educational and Psychological Measurement, 51*, 123-133.
- McInerney, D. M., & Sinclair, K. E. (1992). Dimensions of school motivation. A cross-cultural validation study. *Journal of Cross-Cultural Psychology, 23*, 389-406.
- McInerney, D. M., Roche, L. A., McInerney, V., & Marsch, H. W. (1997). Cultural perspectives on school motivation. *American Educational Research Journal, 34*, 207-236.
- Tabachnick, B. G., & Fidell, L. S., (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Van de Vijver, F., & Hambleton, R. K. (1996). Translating test: Some practical guidelines. *European Psychologist, 1*, 89-99.
- Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.
- Vogt, W. P. (1999). *Dictionary of statistics and methodology* (2nd ed.). London: Sage Publications, Inc.
- Watkins D., McInerney, D. M., & Boholst, F. A. (2003). The reliability and validity of the inventory of school motivation: A Philippine Investigation. *The Asia Pacific Education Researcher, 12*, 87-100.



Development of Ethnic Identity Measure (EIM) for Cordillera indigenous people

Cheryl F. Olvida
Apayao State College

Abstract The purpose of the study is to develop then assess the factor structure, reliability and validity of the Ethnic Identity Measure for Cordillera Indigenous People (EIM for CIP). The initial 60-items were developed based on three factors: exploration, commitment and cultural practices. The item structure was reviewed by scale experts while the content was reviewed by a panel of Cordillera indigenous people and experts on Cordillera indigenous people's consciousness. Factor analysis yielded a 27-item EIM for CIP with three factors. The three factors were: Attitude towards own ethnic group, cultural preference, and Cordillera indigenous people's consciousness. The EIM for CIP was found to have extensive internal consistency (.73) and high convergent validity when correlated with self-esteem ($r=.213$; $p=.002$)

Keywords: *Ethnic identity, indigenous people, consciousness*

Introduction

Here in the Philippines, one of the important variables we need to understand is ethnic identity. The Philippines, having 160 ethnolinguistic groups, shows that we are composed of highly diverse population. In Cordillera administrative region alone we have different ethno-linguistic population namely; Isneg, Itneg, Tinguian, Kalinga, Gaddang, Applai, Bontok, Bago, I-lagod, Kankana-ey, Balangao, Ibaloi, Ifugao, Ikalahan, Kalanguya, Ilongot and others. Some of these ethnic groups are considered indigenous people such as Isneg, Kalinga, Ibaloi, Bontoc, Ifugao and Kankana-ey, while others are migrants from nearby provinces such as the Ilongot, Malaweg, and Itawis.

According to Lanson-Billings (2006), it is natural that students and teachers who have the same culture understand each other and thus students perform in ways that the teacher expect and appreciate. This means that understanding the ethnic groups would be the first step in handling culturally diverse classroom in the Philippines. Unfortunately, there is no valid instrument that would help teachers understand ethnic groups in Cordillera specially the indigenous people. Because of the lack of such instrument, teachers most often use their intuition and labeling

student according to their ethnic group, and not according to students' performance, values and view of themselves. Our government also acknowledges the importance of considering ethnic differences in school. However, the governments' focus was on language only and little is known about ethnic identity, its measures and how it affects the students.

Therefore there is a need to develop a valid instrument that will help teachers, parents and the community as well, in understanding the sense of identity of students from different ethnic groups in Cordillera region specially the indigenous people.

The development of EIM for CIP can help teachers in Cordillera (specially non-Cordillerans) to understand students coming from different ethnic groups. It can help teachers see how their students view themselves and thereby help them become more aware of their attitudes, beliefs and self-concept. EIM for CIP can also serve as the guidance counselors starting point of planning how to develop the students' identity, self-esteem and cultural awareness. Furthermore, it can contribute to understanding the dynamics of teaching ethnically diverse classroom. And lastly, the result of the study can be a gauge for parents and the community to know if there is a need to affirm or re-affirm to students their indigenous culture and heritage. However, the developed EIM for CIP is applicable for literate indigenous peoples only.

Ethnic Identity

In educational researches ethnic identity is most often than not treated as a categorical and unidimensional construct. However, recent studies state that ethnic identity should be treated as a multidimensional construct (Cokley, 2007; Helms, 2007; Phinney & Ong, 2007; Quintana, 2007; Trimble, 2007).

Phinney (1990) defined ethnic identity as the extent to which one identifies with a particular ethnic group. He also said that it refers to one's sense of belonging to an ethnic group and the part of one's thinking, perceptions, feelings, and behavior that is due to ethnic group membership. The ethnic group tends to be one in which the individual claims heritage. Phinney further states that ethnic identity is separate from one's personal identity as an individual, although the two may reciprocally influence each other.

Phinney identified four major components of ethnic identity. These are ethnic awareness, ethnic self-identification, ethnic attitudes, and ethnic behaviors. Ethnic awareness is the understanding of one's own and other groups. It is also termed as the extent to which people are consciously aware of their ethnicity.

Ethnic self-identification is the label used for one's own group. At times it is considered as a single linear variable (Phinney, 1990) but research literatures shows that ethnic identification is also complex in nature (Zimmermann, Klaus, & Constant, 2006). Identifying one's self as belonging to an ethnic group is influenced by the family structure, situation and location. Furthermore, ethnic self-identification can be partially imposed on individuals from outside due to observable characteristics like race (e.g., Giménez, 1992; Henry & Bankston, 2001; Phinney, 1990), which forces people into having feelings that they would not have otherwise.

Ethnic attitudes are the feelings about own and other groups. Just like ethnic self-identification, ethnic attitude is also complex. Laishley (1975) sees ethnic attitude as developmental and composed of the following: recognition of differences between people; belief in the importance of these differences; categorization system for differentiating and labeling people; like/dislike, preference, approach/ avoidance; attribution of traits- the stereotyping process and perceptions of social situations, interpretations of the positions of ethnic groups in society.

Ethnic behaviors refer engaging in behaviors and practices that are specific or representative of one's ethnic group such as speaking the language, eating the food, use of clothing and associating with members of one's group.

Ethnic identity is also considered as a self construct similar with self-concept, self-efficacy, self confidence and self-esteem. Just like other self-construct, ethnic identity seems to be positively associated with other desirable qualities, such as better quality of life, higher academic performance, and so on. Several study show that self-esteem is positively correlated with ethnic identity (Lopez, 2008; Umana-Taylor, 2004). According to Umana-Taylor and Shin (2007) the strength of the relation between each ethnic identity component and self-esteem varied based on group membership, suggesting that the variables may be more or less influential on self-esteem depending on one's group membership.

Development of Ethnic Identity

Identities do not come out overnight, a person undergoes the process of socialization, assimilation and acceptance of one's self to be able to form his/her identity. Erikson (1968) theorized identity as a developmental task of youth where their identity is solidified during this time through a process of exploration and a formation of a coherent set of values and beliefs. Marcia (1980) agreed with Erikson and the process of exploration, but saw the youth choosing and committing to an identity after experimenting with available societal options. Marcia identified four categories of individuals making identity commitments, referred to as statuses. These statuses include (a) identity achievement where the individual has made commitments after a period of exploration, (b) foreclosure where people make identity commitments without a period of exploration and experimentation, (c) moratorium where individuals are in a period of crisis or active exploration in an effort to discover values and goals to fit themselves, and (d) identity diffusion where people are without identity commitments and make no effort to explore or create options (Marcia, 1980). Adolescents along with their parents and family members must actively explore and provide options for identity experimentation so youth can commit to a life-course that is right for them. A part of one's overall identity is the exploration and commitment to an ethnic identity.

Measures of Ethnic Identity

The concept of ethnic identity was rooted in Eric Erickson's concept of ego identity and James Marcia's identity formation thus; ethnic identity is viewed as developmental and lies in a continuum. It is also a multifaceted inner structure of the self, developmental and dynamic but measurable.

Several scales have been developed to measure ethnic identity. The most frequently used ethnic identity measure is the Multi-Ethnic Identity Measure (MEIM) by Phinney (Ponterotto & Park-Taylor, 2007). The MEIM has three factors namely (a) affirmation, (b) achievement, and (c) behaviors. Umana-Taylor, Yazedjian, and Bamaca-Go´mez, (2004) also develop an ethnic identity measure and found out that their scale also have three distinct components (a) exploration, (b) resolution, and (c) affirmation.

Phinney and Ong (2007) revised the MEIM and came up with the Revised Multi-Ethnic Identity Measure consisting of two factors, exploration and commitment, which are distinct but closely related processes that make separate contributions to the underlying structure of ethnic identity. The two factors are highly correlated (.74) and with Cronbach's alphas of .76 for exploration and .78 for commitment. The MEIM-R has six items, three items for each factor. The items have been found to have excellent fitting (AGFI=.96, CFI=.98, RMSEA=.04) and can be easily incorporated with other measures. Phinney and Ong (2007) emphasized that the MEIM-R is a generic measure of the development of one's identity and recommended that additional measures should be used in research with particular group to provide understanding of group-specific values, attitudes and behaviors. It is suggested that to improve the ethnic identity measure specific attitudes, values and behavior of the group should be part of the measure. Similarly, Cokley (2007), and Ponterotto and Park (2007) state that cultural practices are very important components that should be included in all ethnic identity measure because the valuing and appreciation of one's culture is one's ultimate expression of the sense of belonging to a group.

Based on the recommendation, an ethnic identity scale should be composed of three factors; exploration, commitment and cultural practices. Exploration pertains to the individual's act of seeking information and experiences relevant to their ethnicity (Phinney & Ong, 2007). Commitment refers to a strong attachment and personal investment the individual gives in a group. Cultural practices refer to specific attitudes, values and behavior of the group.

The Cordillera Indigenous People Sense of Identity and Consciousness

In Cordillera, Victor (2008) identified the development of Cordillera Indigenous people's consciousness based on historical experience. For the past four years she gathered reflections of Indigenous People (IP) participants during trainings, consultations and other gatherings of CBCP Episcopal Commission on Indigenous Peoples (CBCP ECIP). She found out that in Cordillera the sense of self and identity of the indigenous people in the region changes through time. Those who are born in the 1950s and 1960s show sense of self and identity as

“wanting to assimilate”. The urge to be part of the mainstream society is the prevailing norm and the IP way of life is seen as backward and uncivilized.

In the late 60's and 70's, Cordillera indigenous people, particularly students, developed a sense of identity that Victor (2008) called as “covert inferiority.” Students during this time develop a sense of pride as an Igorot based on one's competency, position secured and professional achievement/accomplishments. However, the standard used for measuring achievement and performance is still along the mainstream concepts (roughly equated with the lowlands and the lowlander). She further states that for those studying at that time, academic and professional achievement became a goal not only for its practical benefits but also to prove that Igorots are also as intelligent and as capable as the lowlanders, or even better (Victor, 2008).

Cordillera ethnic identity as seen through the tourist lens emerged strongly in the late 70's to 80's, and still visible today. According to Victor (2008) because of the flourishing tourism in Baguio City, ethnicity is seen as a uniqueness to highlight marketability and showcase value and cultural products as potential commodities to be traded. Furthermore she stated that one's sense of ethnic identity is equated to knowing how to dance, to play instruments, to perform something native during programs and intermission numbers.

The last sense of identity Victor (2008) identified is assertion of identity, heritage and history (emerged significantly in the early 80s, gaining ground until today). Due to social and political issues on ancestral lands and the construction of Chico Dam the IP's struggle results to their reflection and assertion of indigenous history, heritage and identity (Victor, 2008).

The four senses of Cordillera indigenous peoples' identity that Victor identified just show us that identity or our sense of self is clearly influenced by our social and political environment. Thus society's historical development of the time is part of the individuals' ethnic identity development.

Victor (2008) also identified cultural practices that are common among Cordillera indigenous people. She asked questions about these cultural practices to identify the views and consciousness of Cordillerans about their identity and heritage. Some of the questions were: “Can we sell sacred artifacts?” and “Should the elder's opinions matter given that there is an elected barangay council?” Based on the result of the interview, Victor (2008) identified three possible views and consciousness about Cordillera indigenous people ethnic identity and heritage (a) the view that IP identity and heritage is backward, (b) the view that indigenous culture should be showcased, and (c) the view that one's consciousness is influenced by efforts to be rerooted in one's culture, heritage and history. The three possible views and consciousness about identity and heritage of Cordillera indigenous people presented by Victor (2008) can be used as basis for measuring the ethnic identity of Cordillera indigenous people. The magnitude of importance and value given by an indigenous people to their cultural practices will show their level of identification in their group.

The Present Study

The present study developed then assessed the factor structure, internal consistency, and validity of the Ethnic identity measure for Cordillera indigenous people. Specifically it tested whether exploration, commitment and cultural practice are factors that determine ethnic identity for indigenous people in Cordillera.

The study is concerned in the development of Ethnic identity measure for Cordillera indigenous people (EIM for CIP). It does not reflect specific difference among ethnic groups in Cordillera but rather focuses on the common cultural practice of indigenous people in Cordillera. The EIM for CIP seeks to understand the students' sense of self as a group member and sense of belonging in the group. It does not reflect the complexity of the Cordillera culture.

Method

The development and validation of the ethnic identity measure for Cordillera Indigenous people goes through item construction, item review, factor analysis and test for reliability and validity.

Participants

Participants of the study are students from Apayao State College level. Currently there are 1,052 college students in Luna and Conner Campus of Apayao State College. Students from Apayao who belong to the indigenous people served as the respondents. Since the registrar office does not have record on who are the students who belong to the indigenous people of Cordillera, the researcher requested students and faculty to help identify possible respondents for the study. A total of 205 college students who belong to the indigenous people of Cordillera were identified and served as the respondents of the study. The average age of the respondents was 18.57 years old with standard deviation of 2.7. There were more females (71%) than males (29%). Ethnic groups were self-identified, that is, determined on the basis of responses to the open-ended item in the questionnaire. More than half (55%) of the sample was from the Isneg group, followed by the Kalinga ethnic group with 42 (20%), 14 (7%) Kankanaey, 9 (4%) Bago, and 6 (3%) Ibaloi. There are 22 (11%) respondents who are product of amalgamation.

Test Construction and Content

The developed test was primarily based on the Multi-ethnic Identity Measure by Phinney and Ong (2007) and Victor's (2008) result of qualitative study on Cordillera IP consciousness. Three factors are expected to come out of the developed EIM for CIP; exploration, commitment and appreciation of Cordilleran IP culture. The initial test is composed of 60 items: 10 items for exploration; 20 items for commitment and 30 items for cultural practices present in Cordillera and common to all indigenous people in Cordillera. Other items in the questionnaire

are for socio-demographic data. And lastly, three more items were devoted to elicit the respondents' self-ethnic identification and his parents' ethnic affiliation.

Likert scale was used as the form of responses to the items. Instead of the usual 5-point scale the researcher uses a 6-point scale to avoid getting neutral responses. The scale interpretation are as follows; 1- strongly disagree, 2 - disagree, 3 -somewhat disagree, 4 - somewhat agree, 5 agree, and 6 - strongly agree.

Item Review

Before distributing the EIM for CIP, items are subjected to item review by test development experts. Two sets of reviews was done, the first was the item review wherein identified experts on test construction checked the structure of the items. They were also instructed to check how relevant they think each item is to what it is intended to measure. The items clarity and conciseness was also reviewed. Unfortunately the first set of reviewers was not that confident enough to check the content of the cultural practice since they themselves are not aware of the Cordillera indigenous people's culture. Thus a second panel of review was conducted to specifically check whether the content of the cultural practices are true for Cordillera indigenous people.

The second set of review was done with a panel of Cordillera indigenous people. Members of this panel consist of Ibaloi, Kalinga, Bago, Kankanaey, and Ifugao. They are specifically instructed to check whether the items on cultural practices are true for them and suggest on other items that may be true for the Cordillera Indigenous people. Lastly, the reviewed items were further reviewed and commented by experts who are knowledgeable on studies about Cordillera consciousness. Their suggestion and comments are considered in the revision before conducting the survey.

Method of Data analysis

To assess the factor structure, internal consistency, and validity of the Ethnic identity measure for Cordillera indigenous people the following statistical procedures were used.

Sampling adequacy. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was used to measure whether or not we can group the data into a smaller set of underlying factors (Field, 2000). If the value is close to 1.0 it indicates that a factor analysis may be useful for the data. But if the value is less than .50, then factor analysis may not be very useful. To test whether the data was significantly correlated and does not show multicollinearity, Bartlett's test of sphericity was used.

Number of factors. There are several ways we can derive the number of factors, through the use of eigenvalues and variance, scree plot, and communalities (Field, 2000.) Using Guttman-Kaiser rule, all factors with eigenvalues greater than 1 should be retained, however there are instances that there are numerous factors and the clustering of items is not much meaningful. It also suggests that factors which account for 70-80% of the variance should also be kept. Analysis of the scree

plot is another way we determine the number of factors. The rule of thumb in analyzing the scree plot is very simple, factors before the breaking point or elbow should be kept. Furthermore it is also important to check communalities after factor extraction. If the communalities are low, the extracted factors account for only a little part of the variance, and more factors might be retained in order to provide a better account of the variance.

Factor loadings and rotation. Exploratory factor analysis is utilized to determine significant factor loadings ($> .40$). Specifically, varimax rotation was utilized to maximize the factor loadings. Using Thurstone's rule the number of items in each factors were delimited to have significant items retain in the final EIM for CIP scale. Thurstone's rule states that all items with factor loading below $.40$ are dropped as well as those factors that have double loading, in this case those items which have factor loadings of $.25$ and above in two or more factors are also dropped.

Reliability. To test the internal consistency of the test, Cronbach's alpha and item-to-total correlation are computed and analyzed. The following ratings on the strength of Cronbach's alpha levels served as reference in the interpretation of the Cronbach's alpha: (a) $.80$ or better exemplary; (b) $.70$ -. 79 : extensive, (c) $.60$ -. 69 : moderate, and (d) below $.60$ minimal (Floyd & Widaman, 1995).

Validity. Validity of the EIM for CIP and its three factors was conducted in two ways: content and convergent validity. Content validity has been done during the construction of the test. Specifically we looked whether the items in the EIM for CIP are representative collection of items. Convergent validity on the other hand was check using correlation analysis. Self-esteem was identified as the similar construct and Rosenberg's Self-esteem scale (1965) was used.

Results

Sampling Adequacy

Analysis of the data using Kaiser-Meyer-Olkin measure of sampling adequacy yield a value of $.748$. The KMO value is not close to 1.0 but is also not below $.50$ thus it is safe to use factor analysis and suggests that the data may be grouped into a smaller set of underlying factors. The result of Bartlett's test of sphericity (4594.13) was significant at $.00001$ and so there is no multicollinearity among the items thus the data is good for factor analysis.

Factor Loadings

After subjecting into analysis the responses of 205 respondents, using principal component analysis, 19 factors were generated and using the scree plot, three factors can be identified. Since item communalities range from $.56$ to $.77$, the three factors is sufficient enough for the scale. Out of the initial 60-items EIM for CIP 38-items yielded factor loading greater than $.40$. After delimiting the factor a

total of 27 items were left. The mean score of EIM for CIP was 120.46 ($SD=12.10$) with skewness of $-.28$ and kurtosis of $-.14$.

Factor I-III. The first factor, attitude towards own ethnic group includes 13 items and had a mean factor score of 63.05 ($SD=8.11$) and accounted 65% of the variance. The mean factor score has a skew of -1.03 and kurtosis of 2.22 . The items in this factor pertain to developmental aspect of ethnic identity such as recognition and exploration of one's identity and the individuals feelings as a member of an ethnic group (e.g., In order to learn more about my ethnic background, I have often talked to my parents about my ethnic group).

The second factor, cultural preference, had a mean total of 21.45 ($SD=4.97$) and accounting for 24.74% of the variance is composed of six items. The items determine how one clings to his own ethnic group culture (e.g. I prefer Ilokano language than my own ethnic language). The mean factor score has a skew of $-.09$ and kurtosis of $-.72$.

Cordillera indigenous people's consciousness, the third factor, pertains to the feelings, beliefs and opinion of Cordillera indigenous people (e.g. Community justice system is better than the court justice system). It is composed of 8 items with mean total score of 35.90 ($SD=5.63$) and accounted for 32% of the variance. The mean factor score has a skew of $-.36$ and kurtosis of $-.24$.

Reliability

A total of 33 items were dropped from the EIM for CIP and the remaining 27 items were then subjected to reliability test and yielded Cronbach alpha of $.73$. Based on the interpretation of ratings, the result is considered in the "extensive" range. Item-total correlation ranges from $.005$ to $.47$. Although there are items that are below $.20$, these items were still retained and deemed as useful for the overall scale result. Each factor was also subjected to reliability test: Factor I shows exemplary internal consistency with Cronbach Alpha of $.80$; Factor II has minimal internal consistency with Cronbach alpha of $.51$, and Factor III with Cronbach alpha of $.69$ means it has moderate internal consistency.

Validity

Content and convergent validity were explored in this study. Content validity was done during the time the measure was developed. As for convergent validity, self-esteem was correlated with the EIM for CIP total score and the three factors.

Table 1
Correlations between ethnic identity, factors of ethnic identity and self-esteem

	Factor I	Factor II	Factor III	Ethnic identity	Self-esteem
Factor I	---				
Factor II	-.05	---			
Factor III	.37***	-.10	---		
Ethnic identity	.82***	.33***	.67***	---	
Self-esteem	.16*	.14*	.10	.21**	---

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

Correlation between ethnic identity and self-esteem resulted to significant correlation of .21 at $p = .002$. Among the factors of ethnic identity, only Factor III (Cordillera indigenous people's consciousness) was not significantly correlated with self-esteem and Factor II. All factors were found to be significantly correlated with the ethnic identity score.

Discussion

Factor analysis for the original 60 items ethnic identity measure for Cordillera indigenous people resulted to 27 items EIM for CIP with three factors namely: attitude towards own ethnic group, cultural preference and Cordillera indigenous people's consciousness. Factor I: attitude towards own ethnic identity had the most number of items and included items which were developmental aspect of ethnic identity such as awareness and exploration of one's identity (Phinney, 1992) and the individuals feelings as a member of an ethnic group (Laishley, 1975). Factor II: cultural preference includes 6 items which shows the individuals choice of Cordillera material (foods) and non-material (music and language) culture. Ethnic group preference is also evident in the items in Factor II. Although Factor II has the lowest reliability score it is deemed important because the items would give us idea on the individuals' recognition of ethnic differences, cultural differences and individuals choice of friends and peers. Factor III: Cordillera indigenous people's consciousness has 8 items which directly measures attitudes towards specific issues among indigenous people in Cordillera. The mean scores tell us that the respondents agree with the items listed and further more shows there is a growing assertion of one's identity, and heritage (Victor, 2008). Factor I is significantly correlated with Factor III which shows us that one's attitude towards his own ethnic groups is also shape by the social issues surrounding them.

As for its validity, the study shows that there is highly significant correlation between self-esteem and ethnic identity. Among the different components of EIM for CIP only Factor III was not correlated with self-esteem which suggests one's consciousness of their culture and heritage is not related to one's self-esteem or this result could be attributed to the fact that Filipino by nature draws their self esteem from collective group and not from individualistic point of view as measured by Rosenberg's Self esteem scale. Another study using collective self-esteem scale is therefore recommended.

The high interfactor correlations indicated that the 27-items EIM for CIP is highly related to each other and can be used to measure the level of ethnic identity among Cordillera indigenous people. Additionally, each factor can also be used to measure certain aspect of the individual ethnic identity. Factor I can be used to measure the individual attitude toward his own ethnic group. Factor II will give one's cultural preference and Factor III can help us measure the students' disposition on Cordillera indigenous people's issues.

The EIM for CIP's strength is that it is the only ethnic identity measure prepared to measure ethnic identity among Cordillera indigenous people and unlike other measures on ethnic identity (foreign made) EIM for CIP is the only measure that has items which included specific cultural preference and culturally based issues and concerns. This is very important because valuing and appreciation of one's cultural practices according to Cokley (2007), and Ponteretto and Park (2007) is one's expression of the sense of belonging to a group.

The limitation of the EIM for CIP is that it can only be used for literate indigenous people in Cordillera. The EIM for CIP is definitely not intended to measure ethnic identity of other ethnic group in Cordillera, specially the migrants. However, one may try to use Factor I (attitude towards ethnic group) to measure ethnic identity of other Philippine indigenous ethnic group since the statements in Factor I are generally stated and that the concepts of "respect for elders" are common to all ethnic groups in the Philippines specially to indigenous groups. In Factor II, only the first four items may be used to other ethnic group but not the last three items in Factor II since they are specific cultural behaviors in Cordillera. In doing this one should keep in mind to check the reliability, validity and goodness of fit of the data for a more accurate interpretation. Also, one needs to consider translating the item to their native tongue. For Cordillera however, there is no need to translate since Cordillera is the only English speaking region in the north.

References

- Cokley, K. O. (2007). Critical issues in the measurement of ethnic and racial identity: A referendum on the state of the field. *Journal of Counseling Psychology, 54*, 224-239.
- Commission on Higher Education Memorandum 30 Series of 2004.
- Erikson, E. H. (1968). *Identity: Youth and crisis*. New York: Norton.
- Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage Pub.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299.
- Giménez, M. E. (1992). U.S. Ethnic Politics: Implications for Latin Americans, *Latin American Perspectives, 19*(4), 7-17.
- Helms, J. E. (2007). Some better practices for measuring racial and ethnic identity constructs. *Journal of Counseling Psychology, 54*, 235-246.

- Henry, J.M., & Bankston III, C.L. (2001). Ethnic self-identification and symbolic stereotyping: The portrayal of Louisiana Cajuns. *Ethnic and Racial Studies*, 24(6), 1020-1045.
- Laishley J. (1975). Cognitive process in adolescent ethnic attitude". In G.K. Verma & C. Bagley (Eds.), *Race and Education Across culture* (pp. 100-120). London: Heinemann.
- Lanson-Billings, G. (2006). It's not the culture, it's the poverty of culture: The problem with teacher education. *Anthropology and Education Quarterly*, 37, 2.
- López, I. (2008). But you don't look Puerto Rican: The moderating effect of ethnic identity on the relation between skin color and self-esteem among Puerto Rican women. *Cultural Diversity and Ethnic Minority Psychology*, 14(2) 102-108.
- Marcia, J. (1980). Identity in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology* (pp. 159-187). New York: Wiley.
- Phinney, J. S., & Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54, 271-281.
- Phinney, J. S. 1990, Ethnic identity in adolescents and adults: Review of research, *Psychological Bulletin*, 180(3), 499-514.
- Ponterotto, J.G. & Park, T. J. (2007). Special section: Racial and ethnic identity. *Journal of Counseling Psychology*, 54, 282-294.
- Quintana, S. M. (2007). Racial and ethnic identity: Developmental perspectives and research. *Journal of Counseling Psychology*, 54, 259-270.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Trimble, J. E. (2007). Prolegomena for the connotation of construct use in the measurement of ethnic and racial identity. *Journal of Counseling Psychology*, 54, 247-258.
- Umaña-Taylor, A. J. (2004). Ethnic identity and self-esteem: Examining the role of social context. *Journal of Adolescence*, 27, 139-146.
- Umaña-Taylor, A. J., Yazedjian, A., & Ba´maca-Go´mez, M. (2004). Developing the Ethnic Identity Scale using Eriksonian and social identity perspectives. *Identity: An International Journal*, 4, 9-38.
- Umaña-Taylor, A. J., & Shin, N. (2007). An examination of ethnic identity and self-esteem with diverse populations: Exploring variation by ethnicity and geography. *Cultural Diversity and Ethnic Minority Psychology*, 13, 2.
- Victor, L. (2008). *Migration, education and the Cordillera cultural landscape: Memories, reflections and challenges*. Paper presentation during the seminar on "Faces and Challenges of Migration: The Cordillera Experience" held at Hotel Supreme Convention Plaza, Baguio City.
- Zimmermann, B., Klaus L. F., & Constant, A. (2006) *Ethnic self-identification of first-generation immigrants*. Discussion Paper No. 2535 December 2006.



Examining the relationship of personality and burnout in college students: The role of academic motivation

Adonis P. David

Philippine Normal University, Philippines

Abstract This study extends the literature on burnout by examining the relationship of personality, academic motivation, and burnout among college students. Previous literatures had suggested bivariate links between personality and burnout, personality and academic motivation, and academic motivation and burnout. Integrating these variables together, the author hypothesized that personality and burnout are associated and this association is mediated by academic motivation. Results of bivariate correlations using data from 150 college students were consistent with the literature and the author's hypotheses. However, path analyses using the same data indicated that only conscientiousness has a significant link to a burnout dimension and no mediation was found. Furthermore, direct effects from academic motivation to burnout and indirect effects of personality to burnout via academic motivation were observed. The findings were discussed in relation to the self-determination theory (SDT) and its implication to instruction and assessment.

Keywords: *Personality, burnout, motivation*

Introduction

Burnout has originally been conceived as a work-related syndrome experienced by individuals who work in human services domain like health care and social work (Maslach & Jackson, 1981). More recently, it has been shown that burnout is also experienced by school students (e.g., Gan, Shang, & Zhang, 2007; Salmela-Aro, Kiuru, Leskinen, & Nurmi, 2009; Schaufeli, Martinez, Marquez-Pinto, Salanova, & Bakker, 2002). The concept of burnout has been extended to school context because the primary activities of students like attending classes, reviewing lessons, and making assignments can be considered as work. However, although there has been an upward trend in the study of burnout in the school context, much is still to be known about burnout in students. For instance, the extent to which personality relates to student burnout has yet to be fully established and much of the literature on this relationship has been on job burnout. And in

spite of the academic nature of the work of students in school, little has been said on the role of academic motivation in student burnout. The proposed study intends to fill these gaps by examining the relationship between personality, academic motivation, and burnout among Filipino college students.

Academic Burnout

While the development of the term burnout is attributed to Freudenberger (1974), much of what we know about burnout is based on the conceptualizations of Maslach and Jackson (1981) who described burnout as a syndrome of emotional exhaustion, depersonalization, and reduced personal accomplishment caused by chronic stress from continuous involvement with clients in human services occupations. On the other hand, Zhang, Gan, and Cham (2006) described academic burnout as feeling exhausted because of study demands (exhaustion), having a cynical and detached attitude towards one's schoolwork (cynicism), and feeling incompetent as a student (reduced efficacy).

The literature is replete with studies that examine the relationship of student or academic burnout with various individual and situational or contextual variables. Individual variables that have been hypothesized as related to burnout among students are coping flexibility (Zhang, Gan, & Cham, 2007), personality (Jacobs & Dodd, 2003), and sex (Jacobs & Dodd, 2003; Weckwerth & Flynn, 2004). Situational or contextual variables that have been studied in relation to burnout among students include school climate (Salmela-Alo, et al., 2008), social support (Jacobs & Dodd, 2003; Weckwerth & Flynn, 2006), academic major (Bernhard, 2007), academic achievement (Schaufeli, et al., 2002), and coursework (Law, 2007). A review of the studies in school or academic burnout suggest that these studies are focused on situational stressors - burnout as a function of the situation- and less as a function of the individual's characteristics.

Personality and Burnout

Personality can be described in terms of five factors, more popularly known as the Big Five personality factors. According to John and Srivasta (2009), the use of the term Big Five does not imply that personality differences can be reduced to only five traits but these five dimensions represent personality at the broadest level of abstraction. While authors tend to vary with the terms they used, the most common factors described within the Big Five framework are agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience. Examining the relationship of burnout and personality is based on the assumption that individual characteristics like personality (suggestion: add one more individual characteristic) influence the experience of burnout. This means that a person's personality traits will make him/her more or less susceptible to burnout. Also, it has actually been suggested that personality may help individuals be protected against known risks of developing burnout (Bakker, van Derzee, Lewig, & Dollard, 2006). Empirically, several studies examined the relationship of personality and burnout in the workplace. For one, Langelaan, Bakker, van Doornen, and Schaufeli (2006) found that higher level of neuroticism is positively related with higher level of

burnout. This is expected given that neuroticism is linked to a negative general appraisal tendency and maladaptive and ineffective coping (Grant & Langan-Fox, 2007). Mills and Huebner (1998), meanwhile, have found that neuroticism and extraversion are both associated with all three components of burnout; conscientiousness is associated with emotional exhaustion and personal accomplishment; and agreeableness is associated with both emotional exhaustion and depersonalization. On the other hand, Bakker et al. (2006) demonstrated that neuroticism is the sole predictor of emotional exhaustion; neuroticism, extraversion and openness to experience demonstrated significant effects on depersonalization; and both neuroticism and extraversion predicted personal accomplishment. The study of Zellars, Perrewe, and Hochwarter (2000) also found that neuroticism predicted emotional exhaustion, extraversion and agreeableness predicted depersonalization, and openness to experience predicted personal accomplishment.

The relationship of personality and burnout in the workplace has been well studied but this seems not to be the case for student or academic burnout. While there are studies that examined the relationship of personality and burnout in students (e.g., Jacobs & Dodd, 2003; Watson, Deary, Thompson, & Li, 2007), it seems that there are still plenty to explore in terms of the relationship of personality and burnout.

Personality and Academic Motivation

Only a handful of studies have examined the relation of personality traits to aspects of academic motivation (e.g., Busato, Prins, Elshout, & Hamaker, 1999; Kanfer, Ackerman, & Heggestad, 1996, as cited in Komarraju & Karau, 2005). More recently, Komarraju, Karau, & Schmeck (2009) found that conscientiousness is related across all three motivation dimensions, and in particular, it was found to have positive relationship with intrinsic motivation and negative relationship with amotivation- students who are more organized and disciplined are most likely to be motivated and are less likely to be disengaged. A similar positive relationship was found between openness and intrinsic motivation - students who are intellectually curious are more motivated in learning and schoolwork. In addition, agreeableness has been found to be negatively associated with amotivation while extraversion and neuroticism was positively related with extrinsic motivation. The researchers, however, reported that neurotic individuals tend to score higher on amotivation.

The results of the study of Komarraju, Karau, & Schmeck (2009) were somewhat consistent with the earlier findings of Komarraju and Karau (2005) wherein the avoidance aspect of academic motivation was found to be positively related with neuroticism and extraversion and negatively related with conscientiousness and openness. This means that “neurotic students tend to avoid many aspects of academic life” while “conscientious and open students are less likely to be avoidant in their motivation” (Komarraju & Karau, 2005, p. 564). In addition, the same study found out that students with higher level of openness are more likely to engage in educational experiences while students with higher levels of conscientiousness are more achievement oriented.

Academic Motivation and Burnout

Academic motivation is considered as an important factor in school adjustment and is largely anchored in the self-determination theory (SDT) of Ryan and Deci (2000a; 2000b). This theory views motivation as a continuum with intrinsic and amotivation at either end, and extrinsic motivation in the middle (Ryan & Deci, 2000a). Intrinsic motivation refers to engaging in an activity for its own sake and the experience of pleasure and satisfaction derived from participation. Individuals who are intrinsically motivated tend to have an internal locus of control, are driven to accomplish, seek intellectual stimulation, and are enthusiastic about learning new things while extrinsically motivated individuals pursue education to achieve contingent goals, rather than for an intrinsic enjoyment of learning (Komarraju, Karau, & Schmeck, 2009). The third major component of academic motivation, amotivation, refers to the relative absence of motivation (neither intrinsic nor extrinsic). Amotivated individuals experience feelings of incompetence and expectancies of uncontrollability. It occurs when individuals perceive their behaviors not resulting in a certain outcome and instead perceive their behaviors as caused by forces beyond their own control (Cockley, 2000).

College students face many challenges in their pursuit of higher education. The continuing demands of schoolwork and presence of stressors may lead to negative outcomes like burnout. SDT explains that the more intrinsically motivated an individual is, the more that he will persevere in the performance of a task or the achievement of a goal. Because an intrinsically motivated person feels that he has control over the outcomes of his behaviors, such person will not easily be discouraged from the pursuit of his goals and this would lessen the impact of burnout. The SDT also explains why decreased intrinsic motivation likely leads to all three dimensions of burnout (Rubino, Lukyste, Perry, & Volpone, 2009). In one study, Pisarik (2009) found that individuals who experienced greater levels of intrinsic motivation to attend college were more likely to experience lower levels of exhaustion and cynicism, and higher levels of professional efficacy. In addition, he also found out that individuals who experienced greater levels of amotivation were more likely to experience higher levels of exhaustion and cynicism, and lower levels of professional efficacy.

Personality, Academic Motivation and Burnout

In the preceding sections, the literature reviewed shows that (a) personality, in terms of the Big Five dimensions, has been associated with burnout, (b) the personality of individuals has been found to predict their levels of academic motivation, and (c) academic motivation has been linked to burnout. The current study proposes an integration of these three variables wherein academic motivation serves as a linking mechanism between personality and burnout among college students. This hypothesized link is based on the logic that differences in student motivation levels may be related to basic personality differences (Kommaraju & Karau, 2005) and different motivational orientations lead to different student outcomes like academic burnout. This is consistent with the view that more self-determined behaviors (intrinsic motivation) are more associated with positive

psychological states (Pisarik, 2009). In other words, a student's personality predisposes him to a particular motivational orientation (whether he is more or less self-determined) which in turn predisposes him to the experience of burnout. Indeed, the empirical studies reviewed seem to provide some support for this link. However, these studies are rather limited as none examined the possible role of motivation as a linking mechanism between personality and burnout, whether in school or in the workplace.

The Present Study

On the basis of the literature reviewed, the present study addresses some of the limitations in the current literature on the relationship of personality and burnout among college students. First, I examine the relationship between personality and burnout among a sample of Filipino college students. Second, I examine the possible mediating role of academic motivation in the personality-burnout link. For the first purpose, I specifically examine the following hypotheses:

Hypothesis 1a. Higher neuroticism will be positively associated with emotional exhaustion and cynicism while negatively associated or unassociated with efficacy. Since neuroticism is characterized by emotional distress and poor impulse control (Kommaraju & Karau, 2005), I expect that more neurotic students will have a reduced sense of efficacy and greater experience of emotional exhaustion and stronger cynical feelings toward school.

Hypothesis 1b. Higher conscientiousness will be positively related with efficacy while negatively associated or unassociated with emotional exhaustion and cynicism. I expect students who are disciplined, organized and hard working to display higher levels of efficacy while experiencing less emotional exhaustion and cynicism.

Hypothesis 1c. Higher openness to experience will be positively related with efficacy while negatively associated or unassociated with emotional exhaustion and cynicism. I expect students who are intellectually curious to have little or no cynical feelings and emotional exhaustion from school work while having a strong personal efficacy.

Hypothesis 1d. Higher extraversion will be negatively associated or unassociated with cynicism. I find it reasonable to expect that students who are sociable, confident, and assertive will make more use of environmental sources of motivation in their school work which can reduce cynicism.

Hypothesis 1e. Higher agreeableness will be negatively associated with cynicism. I expect students who are friendly and nurturing to have less cynical feelings toward school work.

In terms of the second purpose of the study, I propose that the students' academic motivation may mediate associations between personality and burnout.

Specifically, I expect the association between neuroticism and the three burnout dimensions to be mediated by amotivation. I also expect the association between the three burnout dimensions and both conscientiousness and openness to be mediated by intrinsic motivation. And lastly, I expect that the association between extraversion and cynicism, as well as between agreeableness and cynicism, to be mediated by extrinsic motivation.

Method

Research Design

The current study made use of the quantitative-predictive method using a cross-sectional design. Data were gathered through the use of self-report measures.

Participants

The participants in the present study are 150 college students from a state university in Manila. The participants were selected through convenience sampling (mean age = 18.27; male =25, female =125) and have received course credit for their participation in the study. Self-report measures were given as one set of inventories and participants were given a maximum of 30 minutes to complete the measures.

Measures

Big Five Inventory (BFI). The BFI, developed by John, Donahue, and Kentle (1991), is a 44-item Likert-type measure of extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. In U.S.A. and Canadian samples, the alpha reliabilities of the BFI scales typically range from .75 to .90 and average above .80; three-month test-retest reliabilities range from .80 to .90, with a mean of .85. (John & Srivastava, 1999). In the present study, the instrument's Cronbach's alpha levels are .82 (total scale), .77 (extraversion), .61 (agreeableness); .65 (conscientiousness), .81 (neuroticism), and .73 (openness to experience).

Academic Motivation Scale (AMS). The Academic Motivation Scale (AMS) is a 28-item measure of students' motivation for attending college (Vallerand, et al., 1992). The AMS uses a 7-point Likert-type scale that measures three major motivation dimensions: intrinsic, extrinsic, and amotivation. The intrinsic motivation has three sub-scales (to know, toward accomplishment, to experience stimulation) like the extrinsic motivation (identified, introjected, external regulation). However, for the purpose of the present study, only the three major dimensions were used. The Cronbach's alpha levels for the instrument in the present study are .90 (total), .84 (intrinsic), .78 (extrinsic) and .66 (amotivation).

Maslach Burnout Inventor-Student Scale (MBI-SS). The MBI-SS consists of 16 items that constitute the three scales for exhaustion, cynicism, and efficacy.

All items are scored on a 7-point frequency rating scale ranging from 1 (never) to 7 (always). The authors reported Cronbach's alphas range from .65 to .86 across three nationalities (Schaufeli, et al., 2002). Cronbach's alphas for the instrument in the present study are .86 (total), .75 (emotional exhaustion), .63 (cynicism) and .88 (efficacy).

Data Analysis

Means, standard deviations, and intercorrelations of all the major variables in the study were determined using the software Statistical Package for the Social Sciences (SPSS) Version 15. Path analysis using Amos 7.0 was used to test the hypothesized mediation model. Path analysis was selected because it gives the researcher a clearer picture of the Big Five personality traits as separate manifest variables vis-à-vis the manifest variables of academic motivation and burnout. The goodness of fit of the hypothesized model was tested using various indexes: chi square, chi-square to degrees of freedom, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square of approximation (RMSEA), and Akaike Information Criterion (AIC). To evaluate the model fit, chi-square should not be significant ($p > .05$), chi-square to degrees of freedom should be less than two, CFI and TLI should at least be .90, and RMSEA should be not greater than .08. In terms of the AIC, the model with less value suggests better fit.

Result

Preliminary Analyses

Table 1 summarizes the means, standard deviations, and intercorrelations of all the major variables in the study. The intercorrelations using Pearson r were used to determine the relationships of the participants' personality traits with academic motivation and burnout dimensions. Findings revealed that consistent with previous studies, neuroticism is significantly related to all three burnout dimensions. This is the exact opposite of the extraversion trait which did not correlate to any of the burnout dimension which is surprising given the findings of previous studies suggesting that extraversion is a significant predictor of burnout (e.g., Bakker, et al. 2006; Zellars, Perrewe, & Hochwarte, 2000). Another surprising finding was the significant relationship of agreeableness with all burnout dimensions. Findings reveal that agreeableness was found to have significant negative relationships to both cynicism and exhaustion, and significant positive relationship with efficacy. Furthermore, conscientiousness was found to have significant positive and negative correlation to efficacy and exhaustion respectively, but no association was found with cynicism. On the other hand, openness to experience was found to have a significant positive relationship with efficacy. And lastly, no significant relationship was found between openness to experience and both cynicism and exhaustion.

Table 1
Summary Statistics and Intercorrelations among the Major Variables

	1	2	3	4	5	6	7	8	9	10	11
1 Cynicism	-										
2 Efficacy	-.30*	-									
3 Exhaustion	-.51*	-.25*	-								
4 Intrinsic	-.23*	.33*	-.21*	-							
5 Extrinsic	-.23*	.16	-.10	.62*	-						
6 Amotivation	.48*	-.23*	-.22*	-.16*	.11	-					
7 Extraversion	-.05	.13	-.03	.03	.04	-.04	-				
8 Agreeableness	-.22*	.31*	-.18*	.10	-.03	-.36*	-.00	-			
9 Conscientiousness	-.12	.30*	-.17*	-.04	-.06	-.03	.09	.35*	-		
10 Neuroticism	.25*	-.18*	.23*	-.18*	-.09	.36*	-.20*	-.39*	.07	-	
11 Openness	-.11	.24*	-.06	.29*	.23*	-.21*	.23*	.13	-.05	-.42*	-

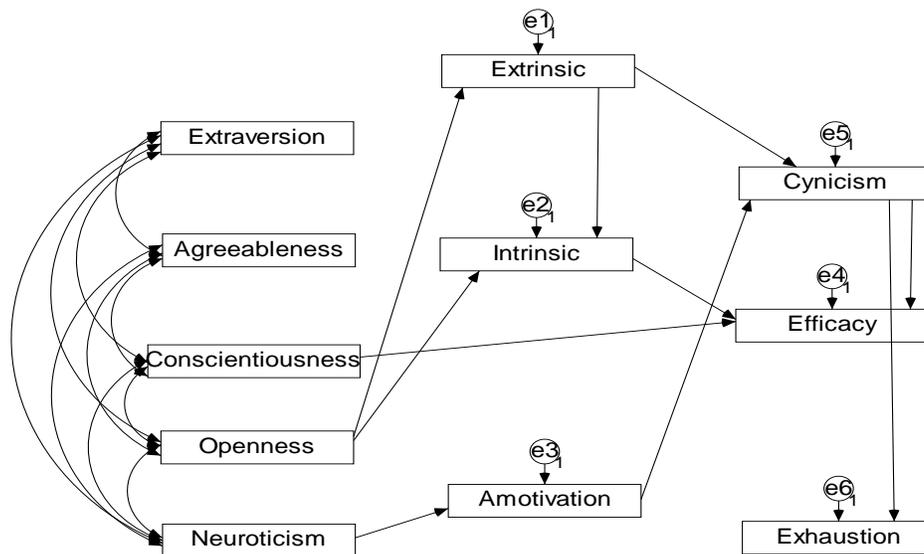
* $p < .05$

Path Analyses

The Empirical Model (Model 1). Using the findings from the intercorrelation of variables, direct and mediated paths from personality to burnout were examined. In this model, paths were added depending on the significance of the relationships between variables from the intercorrelation of variables depicted in Table 1. If an association was found between two variables, a path was added between them, even if they were not part of the initial set of hypotheses. Similarly, if no association was found between two variables, a path was not added, even if a path was initially hypothesized. As a result, Model 1 included direct paths from neuroticism and agreeableness to all three dimensions of burnout; direct paths from conscientiousness to both efficacy and exhaustion; a direct path from openness to efficacy; mediated paths from neuroticism to all three burnout dimensions via amotivation; mediated paths from openness to all three burnout dimensions via intrinsic motivation; and a mediated path from openness to cynicism via extrinsic motivation.

Model 1 was tested for goodness of fit and findings show a rather poor fit across all indexes (chi-square= 91.32, $p=.000$; chi-square over degrees of freedom= 3.38; CFI= .80; TLI=.59; RMSEA=.13). To improve the fit, I made use of the modification indices reported by Amos 7.0 during the data analysis. These indices generate recommendations about loosening constraints on certain model parameters in order to improve the overall model fit. In consonance with the modification indices, I decided to add a path from extrinsic motivation to intrinsic motivation and a path from cynicism to emotional exhaustion and efficacy in the model. These modifications to Model 1 led to the final model (Model 2). Findings revealed a good fit across all indexes for Model 2 (chi-square= 27.57, $p=.38$; chi-square to degrees of freedom= 1.06; CFI= .99; TLI=.99; RMSEA=.02). The AIC of Model 2 was also lower (107.57) as compared to the AIC of Model 1 (169.32), thus providing additional support that Model 2 is a better model than Model 1.

Figure 2
The Final Model Showing the Association between Personality, Academic Motivation, and Burnout (only significant paths are shown)



The Final Model (Model 2). A path analysis of Model 2 provided support to some of the direct paths predicted in the model but not to all. In addition, none of the mediated paths predicted were supported. In the final model, findings revealed that the direct paths from neuroticism to exhaustion (.14), cynicism (.07), and efficacy (.02) were not supported. These findings preclude examining mediation paths from neuroticism across all burnout dimensions. Regardless of this, amotivation was found to have a significant path to cynicism (.42, $p < .001$). This suggests that neuroticism has an indirect effect on cynicism via the path of amotivation because neuroticism has a direct path to amotivation (.36, $p < .001$). Similarly, the direct path from openness to efficacy was also not supported (.15, $p > .05$) but openness has a significant path to both intrinsic (.16, $p < .05$) and extrinsic motivation (.23, $p < .01$). These findings have implications because intrinsic motivation predicts efficacy (.25, $p < .001$) while extrinsic motivation predicts cynicism (-.18, $p < .05$). This suggests that openness to experience does not have a direct or mediated effect to efficacy but it does have an indirect effect toward efficacy through its effect on intrinsic motivation. It also suggests that openness to experience has an indirect effect on cynicism through its effect on extrinsic motivation. On the other hand, the direct path from conscientiousness to efficacy was supported (.25, $p < .05$) while that from conscientiousness to exhaustion was not (.13, $p > .05$). And lastly, the direct path from agreeableness to both cynicism (-.05, $p > .05$) and efficacy (.14, $p > .05$) were not supported. In addition, direct paths from cynicism to both efficacy (-.16, $p < .05$) and exhaustion (.49, $p < .001$) were

supported by the data. Extrinsic motivation was also able to predict intrinsic motivation (.58, $p < .001$).

Discussion

The present study examined the personality-burnout link among college students. Bivariate correlations were used to determine significant associations between the Big Five personality traits and the three burnout dimensions. Findings suggested that four of the Big Five traits (except extraversion) play an important role in the experience of burnout among college students. Most of the findings of the study were consistent with the prediction of the hypotheses outlined. In the path analyses, direct and mediated effect paths were hypothesized and examined to determine whether personality affect burnout directly (independent of academic motivation) or by mediation (through academic motivation). Using an empirical model based on the significant relationships of the personality traits and burnout dimensions of the college students, results indicated (a) substantial link between conscientiousness and efficacy; (b) substantial link between openness and both intrinsic and extrinsic motivation; (c) substantial link between neuroticism and amotivation; (d) substantial link between cynicism and both extrinsic motivation and amotivation; (e) substantial link between intrinsic motivation and efficacy; (f) substantial link between extrinsic motivation and intrinsic motivation; (g) substantial link between cynicism and both efficacy and emotional exhaustion; and (h) absence of mediated effects via academic motivation in the personality-burnout link.

One major finding of the present study is that conscientiousness was the only personality trait that has a significant link to burnout. While the link between conscientiousness and efficacy have not been well-established in previous studies on student burnout, it is reasonable to assume that students who are determined and who are into task and goal-directed behavior may hold high beliefs on their capacity to perform academic tasks. Such personal efficacy may benefit students in terms of experiencing satisfaction and a sense of accomplishment in their school work which may maintain their drive to study and perform well in their academics. This result, however, is in stark contrast with the results of past studies which have shown strong relationship between all Big Five personality traits and the three burnout dimensions. One probable explanation for this divergent finding is that many of these previous studies that were earlier reviewed made use of correlation or regression analyses to examine the personality-burnout link while I made use of path analyses. Indeed, the bivariate correlations of the variables in this study depicted significant associations between personality and burnout consistent with the literature and with my initial hypotheses.

While only one personality trait was able to have a strong association with a burnout dimension, all three academic motivation dimensions were found to be linked with at least one burnout dimension. Previously, it has been shown that intrinsically motivated students tend to have lower levels of burnout while amotivated students were more likely to experience higher levels of burnout (Pisarik, 2009). The results of the present study suggest that more intrinsically motivated students have higher levels of efficacy; more extrinsically motivated students have lower levels of cynicism; and more amotivated students have higher

levels of cynicism. These findings can also be attributed to the role of self-determined motivation in predicting task-oriented coping (Thompson and Gadreau, 2008). It is safe to speculate that more self-determined students experience less burnout because they cope better in stressful situations.

Another important finding is that none of the academic motivation dimensions mediate the link between personality and burnout. This is mainly because conscientiousness was the only personality trait that has a significant direct path to a burnout dimension. In other words, there is little to mediate given the data from the sample of this study. However, the absence of any mediated paths can somewhat be attributed to the indirect association between certain personality traits and burnout dimensions via academic motivation. Indeed, it was found that neuroticism has an indirect effect on cynicism via amotivation; openness to experience has an indirect effect on cynicism via extrinsic motivation; and openness to experience, likewise, has an indirect effect on efficacy via intrinsic motivation. This means that academic motivation influences burnout distinctly and separately from that of personality although its relationship with personality factors lead to personality indirectly influencing burnout. This finding is important given that some authors suggest that significant direct path between a predictor and an outcome variable is not necessary for mediation to occur (see Preacher & Hayes, 2008, p. 880). This means that indirect effects such as the ones found in the present study can be considered as mediated effects in itself. But whether such indirect effects can be construed as mediated effects or not, the more important finding is that academic motivation seems to provide a linking mechanism between certain personality traits and burnout dimensions.

The role of academic motivation in the personality-burnout link is better viewed in terms of the specific variables or dimensions. For instance, a neurotic student is more prone to amotivation, and if he becomes less motivated in his school work, this may lead to feelings of self-doubt and indifference toward academic work and school (cynicism). This can be explained by the fact that being amotivated can already be considered as an emotional response to stressors (e.g., school work, teacher) which may promote disinterest toward school. Similarly, in terms of openness to experience, an intellectually curious and imaginative student who becomes more motivated by intrinsic reasons like goal-achievement may lead to higher confidence and sense of accomplishments in academic tasks (efficacy), while those who are more motivated by environmental reinforcements may tend to explore and enjoy their learning environment which could make them more appreciative of the importance of school and performing well in academic tasks (reduced cynicism). Overall, these associations can be explained by the self-determination theory. No matter what is the predisposition of a student in terms of his personality traits, it is his type or level of motivational orientation (whether he is more or less self-determined), that may dictate whether he will experience burnout or not.

In addition, data from the sample in the present study supported a link between extrinsic and intrinsic motivation, as well as between cynicism and both exhaustion and efficacy. In terms of motivation, this can be explained by the fact that intrinsic and extrinsic motivation under self-determination theory is viewed as a continuum and not as opposite polarities. Hence, being high in extrinsic motivation

may lead to being high in intrinsic motivation. Empirically, I made use of the AMS to measure the students' motivational orientations and the extrinsic motivation dimension of AMS has two sub-scales (introjected and identified) that are designed to capture increasing self-determination among students. For one, Pisaruk (2009) referred to identified regulation as a self-determined motivational orientation like intrinsic motivation. It is not surprising then the data from my sample resulted to a strong positive link between intrinsic motivation and extrinsic motivation. In terms of the three burnout dimensions, there is a dearth of empirical data on the interrelationship of the three burnout dimensions beyond correlations. However, there have been efforts in the past to view burnout as a process wherein one dimension appears first before the other dimensions and this includes a conceptualization where depersonalization (the equivalent of cynicism) appears first prior to the experience of diminished feelings of personal accomplishment and emotional exhaustion (for a brief review, see Cordes & Dougherty, 1993).

Turning now to implications for practice, one recommendation that can be made is for teachers to differentiate their instruction, taking into account the personality and motivational orientation of their students. Because intrinsic motivation results in high-quality learning and creativity (Ryan & Deci, 2000b), it would also be important to provide a classroom and school environment where self-determination or autonomy in learning is facilitated and not constricted. This, in turn, could prevent the experience of burnout that could lead to serious academic problems for students. Clearly, teachers must become aware of their students' motivational orientations (intrinsic or extrinsic) and levels of burnout so differentiated instruction and other forms of intervention from teachers or school counselors can be made. Henceforth, it is important that teachers or counselors must have an assessment of their students' academic motivation and burnout, on top of the usual assessment of standard variables like intelligence and personality. Data from such assessment may go a long way in assisting both teachers and counselors prevent future incidence of burnout and other maladjustments in relation to school work. In addition, it seems logical that when assessing students' motivation, measures that are capable of determining dimensions of motivation (ex. intrinsic, extrinsic) should be used. While the results of this study showed a strong positive association between intrinsic motivation and extrinsic motivation, these two motivational orientations are distinct and separate. As also shown by the path analysis, intrinsic motivation is linked with efficacy while extrinsic motivation is linked with cynicism. It is then imperative that students' intrinsic motivation and extrinsic motivation in studying or attending school are both assessed and used to guide instruction and counseling. Furthermore, it may also be important that psychometricians or educators develop indigenous measures of school or academic burnout and academic motivation for students. This could guarantee the use of culturally and socially appropriate measures of academic motivation and burnout.

In summary, the present study provided preliminary evidence on the possible role of academic motivation in the relationship between personality and burnout among college students. However, there is still a need to examine alternative models of mediation as well as other possible associations like moderation to create a clearer picture on the role of personality and academic motivation to the experience of burnout among college students. Obviously, one

limitation of his study is its cross-sectional nature and its use of self-report scales. Future studies may want to examine personality, academic motivation, and burnout using a longitudinal design and may use high school students in lieu of college students to determine if the same findings will be found to other educational levels. Another limitation was that extrinsic motivation was examined as a singular dimension. Future studies may want to examine the sub-scales of AMS, particularly as Ryan and Deci (2000b) emphasize the multidimensional nature of extrinsic motivation. In spite of these limitations, examining student burnout and motivation seems to be a worthy endeavor that merit further attention. It is my hope that this study is able to provide a springboard for other researchers to investigate the personality-burnout link among college students and the possible role that academic motivation plays in this link.

References

- Bakker, A., Van Der Zee, K., Lewig, K., & Dollard, M. (2006). The relationship between the big five personality factors and burnout: A study among volunteer counselors. *The Journal of Social Psychology, 146*, 31-50.
- Bernhard, C. (2007). A survey of burnout among college music majors. *College Student Journal, 41*, 392-401.
- Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1999). The relation between learning styles, the big five personality traits, and achievement motivation in higher education. *Personality and Individual Differences, 26*, 129-140.
- Cokley, K. (2000). Examining the validity of the academic motivation scale by comparing scale construction to self-determination theory. *Psychological Reports, 86*, 560-564.
- Cordez, C., & Dougherty, T. (1993). A review and an integration of research on job burnout. *The Academy of Management Review, 18*(4), 621-656.
- Freudenberger, H. (1974). Staff burn-out. *Journal of Social Issues, 30*(1), 159-165.
- Gan, Y., Shang, J., & Zhang, Y. (2007). Coping flexibility and locus of control as predictors of burnout among Chinese college students. *Social Behavior and Personality, 35*(8), 1087-1098.
- Grant, S., & Langan-Fox, J. (2007). Personality and the occupational stressor-strain relationship: The role of the Big Five. *Journal of Occupational Health, 21*(1), 20-33.
- Jacobs, S., & Dodd, D. (2003). Student burnout as a function of personality, social support, and workload. *Journal of College Student Development, 44*(3), 291-303.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 5a*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York: Guilford Press.

- Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills and self-regulation for learning: A trait perspective. *Learning and Individual Differences, 8*, 185-204.
- Komarraju, M., Karau, S., & Schmeck, R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences, 19*, 47-52.
- Komarraju, M., & Karau, S. (2005). The relationship between big five personality traits and academic motivation. *Personality and Individual Differences, 39*, 557-567.
- Law, D. (2007). Exhaustion in university students and the effect of coursework involvement. *Journal of American College Health, 55* (4), 239-245.
- Langelaan, S., Bakker, A., van Doornen, L., & Schaufeli, W. (2006). Burnout and work engagement: Do individual differences make a difference? *Personality and Individual Differences, 40*, 521-532.
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior, 2*, 99-113.
- Mills, L., & Huebner, E. (1998). A prospective study of personality characteristics, occupational stressors, and burnout among school psychology practitioners. *Journal of School Psychology, 36*, 103-120.
- Pisarik, C. (2009). Motivational orientation and burnout among undergraduate college students. *College Student Journal, 43*(4), 1238-1252.
- Preacher, K. & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891.
- Rubino, C., Luksyte, A., Perry, S. J., & Volpone, S. (2009). How do stressors lead to burnout? The mediating role of motivation. *Journal of Occupational Health Psychology, 14*, 289-304.
- Ryan, R. M., & Deci, E. L. (2000a). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68-78.
- Ryan, R. M., & Deci, E. L. (2000b). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67.
- Salmela-Alo, K., Kiuru, N., Leskinen, E., & Nurmi, J. (2009). School burnout inventory (SBI). Reliability and validity. *European Journal of Psychological Assessment, 25*, 48-57.
- Salmela-Alo, K., Kiuru, N., Pietikainen, M., & Jokela, J. (2008). Does school matter? The role of school context in adolescents' school-related burnout. *European Psychologist, 13*, 12-23.
- Schaufeli, W., Martinez, I., Marquez-Pinto, A., Salanova, M., & Bakker, A. (2002). Burnout and engagement in university students- a cross national study. *Journal of Cross-Cultural Psychology, 33*, 464-481.
- Weckwerth, A., & Flynn, D. (2004). Effect of sex on perceived support and burnout in university students. *College Student Journal, 40*(2), 237-249.
- Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of

- intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, *52*, 1003-1017.
- Watson R., Deary, I., Thompson, D., & Li, G. (2008). A study of stress and burnout in nursing students in Hong Kong: a questionnaire survey. *International Journal of Nursing Studies*, *45*, 1534-1542.
- Zellars, K. L., Perrewe, P. L., & Hochwarter, W. A. (2000). Burnout in health care: the role of the five factors of personality. *Journal of Applied Social Psychology*, *30*, 1570-1598.
- Zhang, Y., Gan, Y., & Cham, H. (2007). Perfectionism, academic burnout and engagement among Chinese college students: A structural equation modeling analysis. *Personality and Individual Differences*, *43*, 1529-1540.



A short form of the subjective well-being scale for Filipinos

Royce Hernandez

De La Salle-College of Saint Benilde, Philippines

Abstract Using past data (sample of $N=774$), a Short Form of the Subjective Well-Being Scale for Filipinos (SWBSF) (Hernandez, 2006) was developed. Of the original 56 items, only 20% (11 items) comprise the Short Form. This was done using regression, principal component analysis, and confirmatory factor analysis. The Short Form demonstrates good item characteristics, reliability, and validity. Also, convergent validity through correlation with the original SWBSF and Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) shows that the Short Form can measure both global and multidimensional aspects of Subjective Well-being.

Keywords: *Subjective well-being, short form*

Introduction

In the recent years, more and more studies are emerging on positive psychology. Researcher sought to develop new measures as well as models that would best encapsulate the question, how does one measure happiness. Diener and Fujita's (1994) conceptualized happiness as Subjective Well-Being. It represents the emotion as well as the state of happiness. Locally, Hernandez (2006) developed an instrument that would be able to measure SWB in the Filipino context. His findings reveal that the Subjective Well-being Scale for Filipinos (SWBSF) is both reliable and valid. The SWBSF can be considered a multi-dimensional measure of SWB (Diener, Scollon, & Lucas, 2003). Unlike other global measures of SWB, the SWBSF looks into the underlying components of SWB - both cognitive and affective (Diener, 1984).

Unfortunately, global approaches to measuring SWB have often been used and preferred by researchers for quick measures of SWB. An example of this is the widely used Satisfaction with Life Scale (SWLS) which has 5 items only (Diener, Emmons, Larsen, & Griffin, 1985). Compared to the SWBSF with 59 items, the

brevity of SWLS not only gives a holistic and quick measure of SWB but also avoids test-taker fatigue.

With a Short Form of the SWBSF, it is hoped that a global measure of SWB can also be achieved similar to the SWLS. The reduction of the items is meant to reduce the time to answer the test but not to sacrifice the reliability and validity of the instrument. Also, the Short Form is expected to act as a rapid screener whose results can be further complemented by the use of the original SWBSF.

Method

Participants

Data were pooled from two previous studies carried out by the author in the development of the SWBSF. The first set of data ($N=504$) is from the actual construction, development, and standardization of the instrument (Hernandez, 2006). The second set of data ($N=270$) came from the re-validation of the instrument (Balingit & Hernandez, 2008). The use of previously gather data was also done by the group of Smith in the Rasch Analysis of the Hospital and Depression Scale (Smith, et al., 2005). Furthermore, the reported demographics show that: (a) the average age of the participants is 21.67 ($SD=4.83$); (b) females comprise 62% of the sample; and (c) 87.8% are single.

Materials

The sources of data were from two previously done studies. These two studies used two tests that measure SWB:

(1) The Subjective Well-Being Scale for Filipinos (SWBSF) by Hernandez (2006) measures the SWB levels of young adults with a suggested age range of 18 to 35 across 14 scales/dimensions (i.e. positive affect, negative affect, past, future, desire to change, education, work/career, family, leisure/social, health, financial/material, self/physical appearance, relationships, and spiritual). The instrument contains 56 items plus 3 validity check items which can be answered using a 4 point scale. Internal consistency of the 14 scales ranges from .47 to .86 and .94 for the over-all score. Correlations with three standardized tests confirmed the validity of the test construct.

(2) The Satisfaction with Life Scale (SWLS) by Diener et al. (1985) is a 5-item instrument designed to measure global cognitive judgments of the rater's lives. Each item is rated using a 7 point scale. The test has a high temporal stability ($r=.92$) and high internal consistency ($r=.87$). It is reported to have moderate to high correlations with other SWB measures and is suitable for different age groups.

Procedure

The researcher made use of available data from two previously conducted studies using SWBSF and SWLS. The data were then analyzed using SPSS and AMOS software.

Data Analysis

In the course of development of the short form of the SWBSF, the researcher made use of the following steps:

Step 1: Using the SWLS as a global measure of SWB, the items of the SWBSF were correlated with it to determine which of the items are significantly correlated. Items with less than .30 correlation coefficient were discarded and not incorporated in further analysis.

Step 2: Using linear regression, significantly correlated items were used as predictors of the SWLS. The researcher made use of the three methods (i.e. stepwise backward, and forward regression methods) to check which of the SWBSF items can be considered good and acceptable predictors.

Step 3: Good or acceptable items were then subjected to Principal Component Analysis to determine if there were underlying factors in the new Short Form of the SWBSF.

Step 4: To establish the validity of the test, confirmatory factory analysis (CFA) through Structural Equation Modeling (SEM) was done. The items of the short form of SWBSF were considered as manifest variables of the SWBSF-SF (Short Form). The SWLS was used as a latent exogenous variable on the SWBSF which is considered as the endogenous variable. Model fit summaries of goodness of fit were then used to validate the SEM results.

Step 5: To establish the reliability of the test, the Cronbach Alpha was computed.

Results

Item Correlations

Of the 56 items found in the original SWBSF, only 30 items are qualified. The rest of the items have correlation coefficients that are less than .3 as set by the researcher. The qualified 30 items are presented in Table 1.

Table 1
Qualified Correlation coefficients of SWBSF items

SWBSF items	Correlation Coefficient with SWLS
2	.380**
3	.543**
5	.379**
7	.365**
9	.308**
12	.322**
14	.383**
24	.459**
25	.308**
26	.308**
28	.406**
29	.182**
30	.313**
31	.349**
33	.587**
34	.378**
35	.398**
36	.480**
38	.422**
40	.343**
42	.461**
44	.323**
45	.308**
47	.312**
48	.388**
49	.336**
50	.149**
51	.140**
53	.394**
56	.340**

** $p < .01$

Linear Regression

Using the 30 SWBSF items that are highly correlated with the SWLS as predictors of SWLS, the results reveal that the Adjusted R^2 for all of the methods are equal to .549 value. Of the 30 items, only 11 items are significant predictors of SWLS. Shown in Table 2 are the standardized and unstandardized coefficients and significance of the 11 items.

Table 2
Linear Regression of Qualified Items

Model	Unstandardized Coefficients		Standardized Coefficients		
	<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>t</i>	<i>p</i>
(Constant)	-2.130	1.253		-1.699	.090
Item_2	.724	.298	.080	2.433	.015
Item_3	1.333	.268	.183	4.982	.000
Item_5	.552	.238	.077	2.323	.021
Item_24	.856	.263	.111	3.249	.001
Item_33	1.738	.312	.218	5.577	.000
Item_35	.616	.234	.090	2.630	.009
Item_36	.955	.239	.138	3.996	.000
Item_45	.739	.215	.106	3.437	.001
Item_47	.420	.171	.078	2.456	.014
Item_49	.775	.286	.089	2.712	.007
Item_56	.421	.195	.069	2.164	.031

Principal Component Analysis

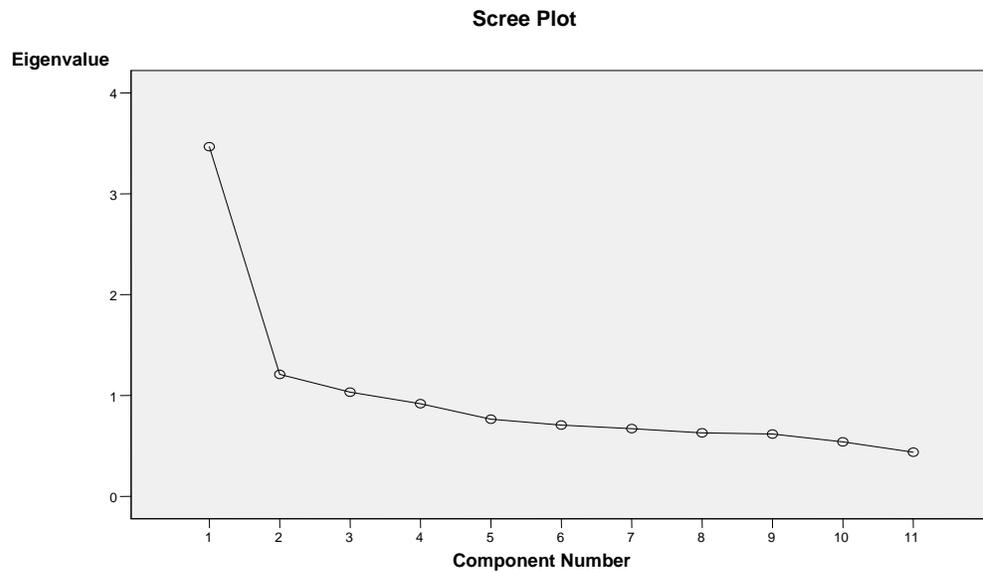
Using the same set of 11 items, these were subjected to principal component analysis with varimax rotation to determine if there exists underlying components. Results indicate that there are three possible components which accounts for 51.909 cumulative variance as shown in Table 3. However, Scree plot (Figure 1) reveals that there is only one factor. Given this, the researcher opted to vary further the three models in the succeeding analysis.

Table 3
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.47	31.515	31.515	3.47	31.52	31.52	2.68	24.348	24.348
2	1.21	11.000	42.515	1.21	11.00	42.52	1.67	15.209	39.557
3	1.03	9.395	51.909	1.03	9.39	51.91	1.36	12.352	51.909

Note. Extraction Method: Principal Component Analysis.

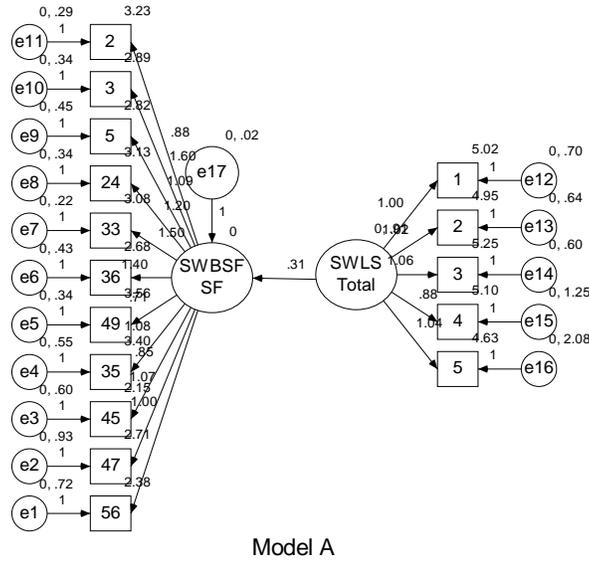
Figure 1
Scree Plot of the 11 items



Validity

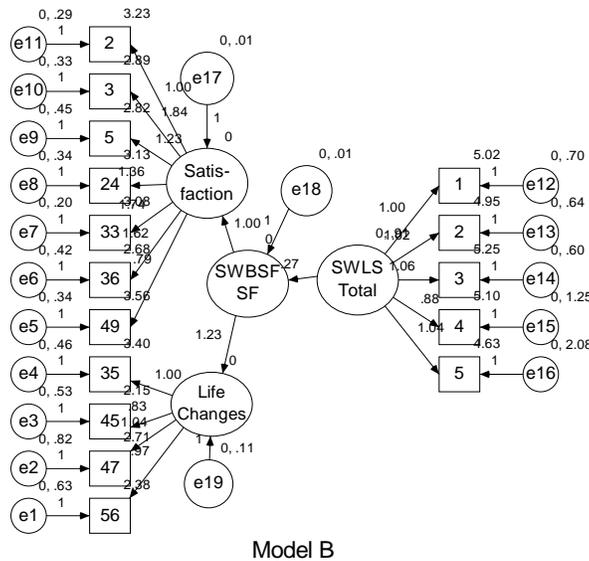
In the preceding factor analysis, there were three suggested model structures of the Short Form instrument. Given this, the researcher ran the data using AMOS software to come up with three models (see Figures 2, 3, and 4). In Model A, all of the items were considered as just belonging to one component which is the Short Form Total of the SWBSF. The SWLS score was then used to establish convergent validity and act as a latent exogenous variable. In Model B, two factors were created - satisfaction and life changes. In Model C, only the first component was included (satisfaction).

Figure 2
Model A SEM



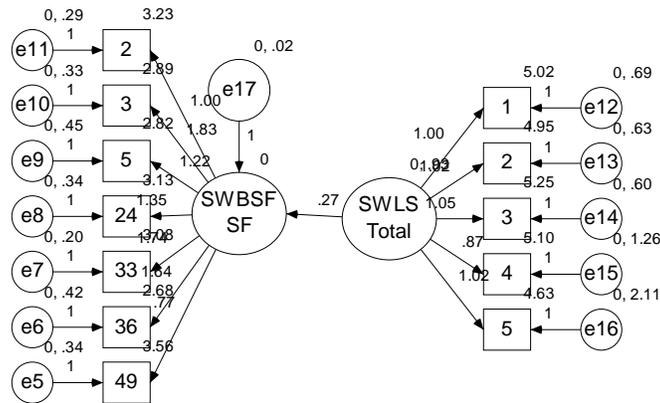
RMSEA = .056 AIC = 454.805 BCC = 457.006

Figure 3
Model B SEM



RMSEA = .046 AIC = 366.940 BCC = 369.230

Figure 4
Model C SEM



Model C

RMSEA = .049 AIC = 224.468 BCC = 225.732

The Goodness of Fit as indicated by the RMSEA value reveal that in all of three models, the data fits the model. As a rule of thumb, the RMSEA should be equal to .05 or less in order to say that there is goodness of fit. However, a value of 0.08 or less would also be considered reasonable.

Table 4
Summary of SEM results

Model	Number of		Goodness of Fit Indicators		
	Items	Component/s	RMSEA	AIC	BCC
A	11	1	.056	454.805	457.006
B	11	2	.046	366.940	369.230
C	7	1	.049	224.468	225.732

To determine which of the three models is acceptable, the Akaike Information Criterion (AIC) and Brown-Cudeck Criterion (BCC) were used. Generally, of the different models presented, the one with the smallest value is considered the acceptable model (see table 4). In this case, it would be Model C (AIC=224.468; BCC = 225.732). However, considering the RMSEA, Model B can be considered the best of the three models (RMSEA = .046).

In addition, to establish convergent validity the significant and very high correlations of the Short Form with the SWBSF ($r=.827, p < .05$) and SWLS ($r=.722, p < .05$) indicates that it is comparable to these two measures.

Reliability

The computed Cronbach Alpha Coefficient is .76 which is within the minimum acceptable value (George & Mallery, 2003). As such, it can be said that the Short Form of the SWBSF is a reliable test comparable to the original SWBSF. Shown in table 5 are the item-total statistics of the 11 items.

Table 5
Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Cronbach's Alpha if Item Deleted
Item 2	28.84	19.891	.401	.751
Item 3	29.17	18.276	.533	.734
Item 5	29.24	19.077	.422	.747
Item 24	28.93	19.041	.476	.742
Item 33	28.98	18.597	.582	.731
Item 35	28.67	18.684	.437	.745
Item 36	29.39	18.782	.437	.745
Item 45	29.91	19.385	.331	.759
Item 47	29.35	18.307	.355	.761
Item 49	28.51	20.274	.320	.758
Item 56	29.69	18.766	.362	.757

Discussion

The purpose of this paper was to develop a short form of the SWBSF using previously available data. First, item correlations with the SWLS were obtained to determine which of the original items can globally measure SWB. Of the 56 items original items in the SWBSF (the additional 3 items are validity checks), only 30 were found to be highly correlated ($r=.30$ and above, $p < .05$). This indicates that the rest of the items that were discarded are not global measures of SWB. To verify this, linear regression was done as the next step. In all of the three methods employed (stepwise, forward, and backward approaches), the adjusted R^2 is .549. This tells us that the 11 items can significantly predict roughly 55% of the variance of SWLS.

Afterwards, the structure of the surviving items were checked for underlying factors using principal component analysis. Initial results inferred the possibility of two to three possible models which were then verified using CFA and SEM. Of the three models tested, the second model showed that there are two underlying components of the Short Form of the SWBSF which were named Satisfaction and Life Changes. This was established using the goodness of fit measures (RMSEA) and model comparison indices (AIC and BCC). Further correlation of the Short Form with the original SWBSF and the SWLS indicate that the new shortened instrument can measure both global and multi-dimensional SWB. As such, the

Short Form can account for both and be utilized as a rapid screener or quick measure. Also, even though the instrument was reduced to about 1/5 or 20% of the original SWBSF, the reliability Cronbach Alpha coefficient is still acceptable. This signifies that the Short Form is reliable.

Lastly, given that there were limitations in the early development stages of the instrument (i.e. the original intended and theorized factor structure did not match the principal component analysis), the current research looked into available data and was able to verify the potential structure of SWB. However, future research can further examine its structure stability, reliability, as well as validity. The possibility of coming up with norms can also be looked into. Nonetheless, the instrument seems to have merit as a research tool that requires fast, dependable, and justifiable results to researchers.

References

- Balingit, F., & Hernandez, R. (2008). *Re-validation of the constructs of the Subjective Well-Being Scale for Filipinos*. Unpublished manuscript, De La Salle University.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*, 542-575.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71-75.
- Diener, E., Scollon, C.N., & Lucas, R.E. (2003). The evolving concept of subjective well-being: The multifaceted nature of happiness. *Advances in Cell Aging and Gerontology*, *15*, 187-219
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Hernandez, R (2006). *The construction, validation, and standardization of the subjective well-being scale for young adults*. Unpublished masters thesis, University of Santo Tomas.
- Smith, A. B., Wright, E. P., Rush, R., Stark, D. P., Velikova, G., & Selby, P. J. (2005). *Rasch analysis of the dimensional structure of the hospital anxiety and depression scale*, *Psycho-Oncology*. Wiley InterScience (www.interscience.wiley.com).

Appendix
The Short Form of the SWBSF

New Item #	Original Item #	Item Stem
1	2	I feel good about myself.
2	3	I am contented with my life right now.
3	5	I am satisfied with the decisions I made in the past.
4	24	I am contented with my social life.
5	33	I can say that I generally lived a satisfying life.
6	35	I sometimes wish that I was born to another family.
7	36	I believe that I have everything I need in life.
8	45	I believe that I'm better off if my life changed for the better.
9	47	Given the chance, I would like to forget my past.
10	49	I look forward to spending time with my family.
11	56	There are times when I wish that I have another work/degree/career other than what I have right now.
Validity Item 12	58	I answered truthfully and honestly all of the items.



Determining the degree of inter-rater consistency in a high stakes, developing country, examination

Jack Holbrook

Visiting Professor in Science Education, University of Tartu, Estonia

Abstract This study examines the reliability of scoring (awarding marks) for essay-type questions in a high stakes, external examinations where marking schemes are not well developed. The study compares the inter-rater reliability between the initial marks awarded in an actual examination with a first remarking where no additional guidance was given and with a second remarking following a two day training programme. As with studies in other countries, the findings show poor correlations between marks awarded and substantial changing of grades on remarking. The additional guidance given in the workshop was found to increase reliability as shown by the correlation with a purposive remarked sample by an expert from the examination board, but the potential to use double marking as a mean of increasing reliability was shown not to be appropriate. Suggestions are made to develop question setter-produced marking schemes and the use of sample marking to give greater guidance to markers to raise the reliability of results.

Keywords: *Inter-rater consistency, high stakes, examinations, reliability, Rasch analysis*

Introduction

The setting of examinations in various subjects by an examination board external to the school is common around the world. The examinations are taken at the end of a course, or school curriculum, and determine success or failure for future advancement and are often referred to as once only, high stakes, examinations. These can involve objective testing, often multiple choice (MCQ) items, and/or more subjective questions where students construct the responses themselves. Where MCQ items are involved, the scoring is dichotomous, undertaken by machine and reliability is less of a concern than validity of the items chosen. As it is generally accepted that valid instruments are important, essay-type questions still remain an obvious choice for evaluation of knowledge (Verma,

Chhatwal, & Singh, 1997). Unfortunately allocating marks for essay-type, student free-response questions can be very unreliable, unless marking systems are well developed (Baird, Greatorex, & Bell, 2004). In fact, poor intra- and inter-rater reliability of rater evaluations in free-response assessments have long been recognized (Van der Vleuten, 1996). With this in mind, it has been suggested that awarding bodies take clear steps to ensure examinations are marked reliably and procedures are detailed in a code of practice (Baird, Greatorex, & Bell, 2004; Newton, 1996).

Reliability is usually taken to mean the reproducibility of scores on another occasion. Reliability has been defined (Berkowitz, Wolkowitz, Firch, & Kopriva, 2000) as the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker. An index of reliability of a test for a given population can be taken to be the ratio of true score variance to observed score variance (Dimitrov, 2002). True scores cannot be directly determined and hence the reliability is an estimate. And of course, this assumes that the scores were obtained from a test that was seen as sufficiently valid for the reliability of the results to have any meaning at all. A very reliable mathematics test would be totally out-of-place as a reliably instrument for the determination of history ability! But if validity is at an acceptable and interpretable level, then reliability is an important consideration facing all examination boards (Brooks, 2004).

Reliability of test scores can be influenced by the type of performance data and the metric in which the scores are express (Yen & Candell, 1991). Unreliability in the assessment system can be traced to a number of factors, such as:

- (a) Use of short tests
- (b) Use of essay scripts where the marking is subjective
- (c) Lack of rater guidelines
- (d) Rater variability
- (e) Bias in the distribution of examination scripts
- (f) Lack of structuring of essay-type questions

(a) Use of Short Tests

In such tests there are insufficient items to sample the domain of learning and hence leads to unreliable measures. Steps to overcome this can, of course, be to increase the length of the examination, but then there is the question of examinee fatigue and if the test is over lengthy, the score again becomes unreliable. Another approach is to make use of MCQ items which are answered quickly, marked reliably and allow coverage of the curriculum content. The MCQ items can then be coupled with a shorter array of other type of questions which sample the testing domain.

(b) Use of Essay Scripts Where the Marking is Subjective

Impression marking has been shown to be unreliable without carefully developed guidelines (Baird, Greatorex, & Bell, 2004). Verma, Chhatwal, and Singh (1997) claimed that the 'structuring' of essay questions provides a tool for improving

rater reliability. In their study, they concluded that structuring an essay question helps improving rater reliability as well as the internal consistency of the test, without any additional input. This can be an important consideration for testing subjects where creativity, or the testing of free writing ability, is not an important issue. It can be seriously considered for science and social science subjects.

(c) Lack of Rater Guidelines

Reliability can be expected to be problematic if there are no marking guidelines, or the guidelines are inadequate. A range of strategies has been adopted to increase the reliability of marking. Among these are better specification of scoring criteria, including in some cases the use of analytic (awarding marks and even half marks, for each specific component) rather than holistic rating scales (indicating a mark range and expecting the rater to indicate whether the script being marked is high medium or low within the range) (Baird, Greatorex, & Bell, 2004; Elder, Knoch, Barkhuizen, & Von Randow, 2005; Moon & Hughes, 2002; Van der Vleuten, 1996;). The goal is to move away from impression marking.

(d) Rater Variability

Variability in rater behaviour may produce invalid, or unfair results for candidates whose absolute scores or relative positions may differ depending on who assesses their performance. To address this, rater training has been suggested (Baird, Greatorex, & Bell, 2004; Elder, et al., 2005; Munro, et al., 2005; Weigle, 1998). The training usually consists of a number of raters being (re)introduced to the scoring criteria and then asked to rate a number of scripts as a group (sample marking). Within the training, ratings are carried out individually, then discussed by the whole group and reasons for discrepancy clarified. The discussion rarely covers the awarding of full marks as this is usually agreed, but the focus of the discussion relates to the manner in which partial marks are awarded and the degree to which explanations given are to be considered appropriate. Where a candidate has written a weak response, clearly far less than adequate, much discussion often takes places as to whether this should be awarded any marks at all.

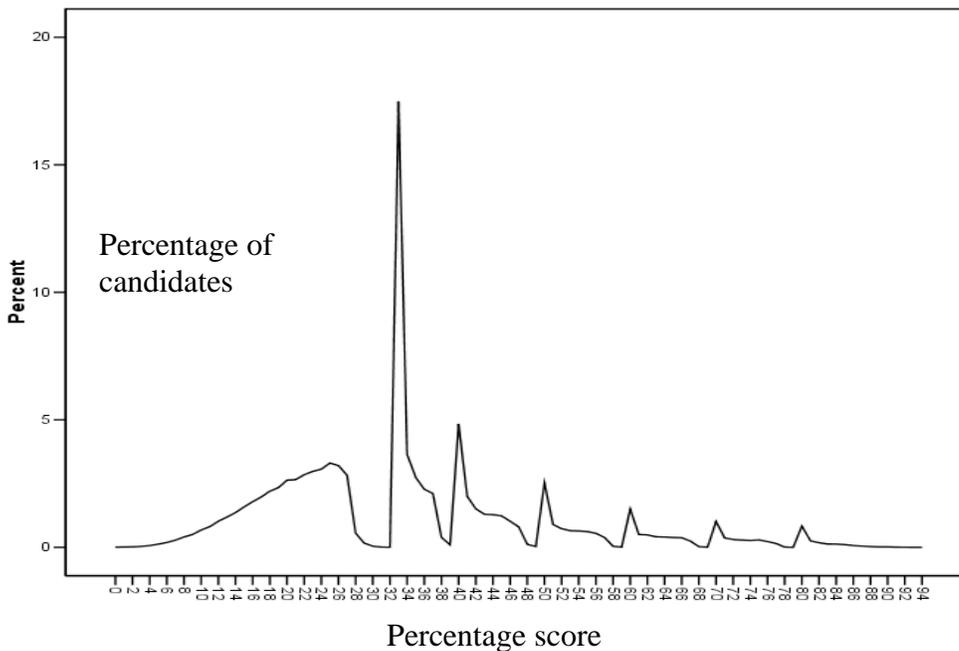
Some of these scripts used for training purposes can be specifically chosen because they correspond to coverage of different score levels on the rating scale, or they might exemplify certain problematic, or unusual issues arising in written assessment. Rater training has been shown to increase the self-consistency of individual raters by reducing random error, to reduce extreme differences between raters, to clarify understanding of the rating criteria, and to modify rater expectations in terms of both the characteristics of the writers and the demands of the writing tasks (Weigle, 1998), although the effectiveness of such training does not make raters 'duplicates of each other' and increases rater consistency rather than rater severity of marking. Unfortunately Lunz, Wright, and Linacre,(1990) note that raters (or judges) often employ unique perceptions which are not easily altered by training. Such perceptions could be to finish the marking as quickly as possible, or to ensure that candidates near the borderline are not failed on account of any rater.

(e) **Bias in the Distribution of Examination Scripts**

Bias can occur if examination scripts are not randomised and one rater gets a collection of scripts which are all worthy of high marks, while another rater obtains low scoring scripts only. This can lead to a tendency to be lenient with the low scoring candidates and more harsh with the high scoring candidates. Randomisation can be expected to lead to a range of scores for the scripts marked and hence allow the marking to have a more full perception of the marking range.

This stems from situations where little control is exercised over the marking process and raters are aware that the marks they award will affect pass rates. It can lead to situations where raters award ‘grace’ marks (extra marks to enable the candidate to reach the pass mark or to reach a mark that would enable a change from one grade score to another). Figure 1 gives a typical outcome taken from a SSC examination in Bangladesh. The pass mark in this examination is 33% and grade awarded to candidates change at 40%, 50%, 60%, 70%, and 80%.

Figure 1
Percentage Distribution for Scores in One Examination by one Examination Board in Bangladesh



(f) Lack of Structuring of Essay-Type Questions

A study, undertaken in India (Verma, Chhatwal, & Singh, 1997), compared the effect of structured essay type questions on rater reliability from a large-scale public examination in history. The study clearly brought out the low reliability of traditional essay questions (TEQ). Further objective evidence of this fact was provided by analysis of the variance of mean marks awarded by seven examiners to the entire group. The internal consistency of the TEQ paper was found to be poor, although a structured essay paper (SEQ) gave a significant internal consistency. Unfortunately they suggested that providing a model answer or check lists to the examiners, are not feasible, at least in their circumstances, because the person setting the paper is usually different from the person evaluating it and thus the chances of agreeing to a common check list are lower. Furthermore, with large numbers of candidates marking TEQs, the time-consuming marking process does not accommodate marking by more than one rater.

Verma, Chhatwal, & Singh (1997) thus claimed that the 'structuring' of essay questions provides an important tool for improving rater reliability. In fact, they concluded that structuring essay questions helped in improving rater reliability as well as the internal consistency of the test, without any additional input. Making the marking of questions less dependent on the rater is certainly an important consideration in subjects where free writing is not considered essential for construct validity.

Reliability of Raters

It has long been recognized that the scoring of essay type questions is a complex process, with many variables, in multiple combinations, influencing the reader differentially (Chase, 1986). Single rater reliabilities (correlation between responses by two raters) can vary between .3 and .8 depending on 'the length and topic of the essay, the amount of freedom students have in selecting and responding to the essay, the experience of the raters, the extent of training given the raters and the control exercised in the marking environment' (Marsh & Ireland, 1987).

Inter-rater reliability refers to reliability of marking between raters. Unfortunately raters vary in their degree of leniency or harshness and the average proportion of variance due to rater effects can be as high as 49% (Eckes, 2005). To address this a number of researchers have suggested rater training (Baird, Greatorex, & Bell, 2004; Weigle, 1998). A fairly simple training approach is to introduce sample marking in which each rater marks the same scripts and then their individual marks are compared, discussed and adjustments made to the marking scheme so as to further clarify the expected marking procedure as put forward. If this is carried out with scripts across the mark range so that raters can agree on the meaning of full marks and partial marks at a number of levels, then the inter-rater marking can be increased (Bédard, Martin, Kruger, & Brazil, 2000). Retraining has been the dominant method to induce judges to make similar assessments of candidate performances (Lunz, Wright, & Linacre, 1990). However its effectiveness is questioned and correcting measure are felt necessary to ensure judge severity is considered more manageable and more fair.

A further important consideration which gained much attention in the mid-20th century is double marking. This involves two raters independently marking each script and the marks compared. It is still a common University practice (Dahl, 2006) and in a country like Bangladesh forms the major attempt at reliability at this level. The public Universities in Bangladesh make use of double marking by sending candidate scripts to two raters and the average of the two marks is taken as the true mark. If the difference between the two percentage marks is greater than 20, then the script is sent to a third rater and the third mark is considered as representative. Johnson, et al. (2005) refer to this as the expert judgement model as the mark replaces both marks for the other examiners. An alternative is to follow what Johnson, et al. (2005) refer to as the tertium quid model in which the mark is averaged with that of the rater closest to the mark given by the third rater and hence essentially eliminating the rater furthest from the average. Unfortunately the use of double marking is time-consuming and costly and has lost favour at secondary levels where examination boards are under pressure to complete their marking as quickly as possible and also to maintain reasonable costings which do not cover the notion of double marking for large numbers of candidates (Newton, 1996). This study examines the effect of double marking and its effectiveness at secondary level, noting the time consuming factor (Verma, Chhatwal, & Singh, 1997).

Measurement of Rater Reliability

It is often believed that both inter- and intra-rater reliability must be documented to support the adequacy of an instrument. However, inter-rater reliability is unlikely to be stronger than intra-rater reliability, because measurement error is more likely to occur with different raters than with the same rater. Thus, one can be convinced of the instrument's intra-rater reliability if inter-rater reliability is adequate, whereas the opposite is not necessarily true. Nevertheless, it is noted that with fewer raters, it is easier to control the reliability factor (Campbell, 2005).

For instruments using continuous scales, reliability is generally measured with correlation coefficients, or paired *t* tests (Bédard, et al., 2000). Researchers often use them interchangeably in the belief that they produce similar results when applied to assess reliability. Bédard, et al., (2000) recognise this assumption is not correct and gives examples to illustrate the point. The Pearson product moment correlation assesses linear relationships and is not affected by systematic differences between sets of scores, as long as the linear relationship between scores remains the same. Specifically, if one rater consistently rates participants higher or lower than another rater, the resulting correlation coefficient will be unaffected by this discrepancy. Against this, the *t* test detects systematic differences between scores, but not random differences above and below the mean, because it is devised to compare means (Altman & Bland, 1983). Newton (1996) examined the reliability using two indices – the coefficient of correlation and the mean mark differences between the initial prime mark and the re-mark. This gives an indication of the change of rater severity in awarding marks. He found extremely high Pearson product-moment correlations, especially for mathematics, and utilised ANOVA to determine the significance in differences of mean scores. In this study correlation

are employed to show relationships between grades for remarking. The t-test is used to show significance of changes of grades on remarking.

Measurement of internal consistency using Cronbach's alpha, whilst not perhaps as authentic as other methods, is realistic. Although a high coefficient alpha will result when a single general common factor runs through items of a test, high values can be found when item variance is determined by several common factors. In their analysis of GCE examinations in the UK, Nuttall and Willmott (1972) observed that most values lay between .8 and .9 and described this as a 'credible achievement'. The selected nature of such a student population makes it unlikely that coefficients above .9 could be obtained. In this study, it is preferred to examine inter-rater consistency using Rasch analysis.

Based on analysis of variance, generalizability analysis or theory (Crossley, Davies, Humphris, & Jolly, 2002) accounts for differing methods of reliability estimation as well as different kinds of reliability—between-rater, within-rater, within-candidate, within-teacher. By placing emphasis on estimating variance components rather than effect, the contributions of each source of difference between the observed score and the true scores can be calculated (Brennan & Pediger, 1981). It may then be possible to forecast the number of measurements required to reach a reasonable estimate of the true score. Once key sources of error have been identified, prediction of reliability levels can be made when alternative assessment arrangements are envisaged, e.g., decreasing the number of papers or increasing the number of raters. As the number of papers, marks per paper and the number of scripts marked by a rater are controlled by examination boards, factors associated with changes in these areas were not explored.

Additional Perspectives on Reliability

In previous research, interrater reliability has been studied using Rasch measurement models (e.g., Engelhard, 1994; Eckes, 2005; Elder, et al., 2005; Lunz, et al., 1990; Weigle, 1998). Rasch modelling allows rater severity to be considered alongside the candidate ability and level of difficulty of the essay-type paper. In this study, Rasch analysis is used to examine rater severity and the effects of training on this.

The Bangladesh Practice

How reliable is the marking of essay-type questions in a country such as Bangladesh, where the external examination is of vital importance for a candidate's future education? Little, if any, training is offered to raters and no answers are supplied by the examination board to the questions set. In Bangladesh, at SSC (grade 10) level, an examiner, utilising guidelines from a scheme which gives only a distribution of marks and, prepared by head examiners, plus verbal directives from the controller of examinations, marks a set of candidate answer scripts. These guidelines really amount to little more than the number of marks to be awarded for each question and the degree of leniency to use. As a result, there is a widespread belief that unreliability prevails in marking the SSC scripts i.e. there is a strong

possibility that marking the same scripts by a different examiner would produce very different results.

As no previous study had been conducted to highlight this issue in the context of Bangladesh, a research study was carried out to determine whether the remarking of SSC scripts would show any significant difference in the marks awarded. As the actual marks are not disclosed to candidates, but a grade used instead, the study considered the change of grade as well as total marks that were obtained on remarking each script. This was possible as no standardisation process is used to match grades to criteria with a view to increasing the fairness of marks awarded.

The study described here was conducted to identify the degree of rater consistency and possible steps that could be taken to improve inter-rater reliability. The paper focusses on inter-rater reliability, in terms of consistency and accuracy of marks given to candidates' written performance on the non-MCQ component of the grade 10 external examinations counting for 50% of total marks in each subject examination.

The study also examined the effect of additional guidance given to raters in terms of training related to a marking scheme per question. This did not involve detailed answers as the questions were sufficiently open-ended to preclude this, but the guidelines were put forward for banding marks, based on descriptors.

In this study the following hypothesis were tested: (1) there is no significant difference in grades, derived from total marks assigned to candidates, on the remarking of scripts; (2) averaging marks from double marking produces a more consistent mark for each candidate; (3) providing limited in-service guidance to examiners produces more consistent marks as shown by a similar distribution in grades per script by raters.

Methodology

The initial marks were taken from those awarded to 4030 candidates in an actual SSC examination. The 1st remarking was undertaken with a mixed group of 11 experienced (9) and new raters (2), marking on a separate occasion, but under similar guidance to that given to the raters during the marking of the original scripts. Each rater marked a separate set of approximately 400 answer scripts. A further remarking (2nd remarking) was carried out after providing the 10 raters with better marking range descriptors and guiding the raters on its use during a 2-day workshop. Each rater marked approximately the same number of randomly assigned answer scripts, a few of which they had marked previously for the 1st remarking. The marks from the second remarking were compared with previous marks.

To carry out the study, an arbitrary script code was recorded on the remaining portion of an OMR (optical machine reading) attachment to the scripts as taken from the earlier SSC examination, as well as to the top of each original answer script. After assigning scripts code, the OMR attachments were removed from the scripts so that total marks from the actual examination were no longer indicated. Any previous rater marks, recorded on the various pages of each answer script, were covered by coloured paper, the scripts randomised and then allocated to a group of

raters comprising both experienced and fresh raters. The raters were not given any guidelines, apart from being supplied with the actual questions used and the number of marks to be assigned for each question; and they were advised to mark the scripts on the basis of their previous experience in internal and external examinations avoiding emotional or sympathetic attitudes i.e. to follow a professional approach. The marks obtained in this way were recorded as the 1st remarking of answer scripts.

After completing the remarking study, the rater were given a short in-service course of two days and guided in the use of a marking approach, developed for this specific purpose by one of the authors. Each rater was guided in the use of this marking scheme by being asked first to mark a sample script. In this, each rater marked the scripts according to the marking guidelines provided and then discussed their marks with the other raters so as to arrive at a consensus on how the marking system is to be interpreted.

Following the short in-service training course, each rater was allotted another set of scripts to mark. No attempt was made to give raters the same scripts as before, although an attempt was made to randomise the scripts so that any one rater did not get all scripts having higher marks, or all the low scoring scripts.

All marks obtained were entered into the computer against the rater number. Marks were totalled and the final mark was record out of 50. As corresponding MCQ marks (which made up the remaining 50% of examination marks) were not available, the marks obtained were converted to grades based on the standard system in use (>40 - A+; 35-39 - A; 30-34 - A-; 25-29 - B; 20-24 - C; 15-19 - D; <15 - F).

To ascertain the expected mark, where marks given to a sample set of 58 scripts differed from each other on the initial, first and second remarking, these scripts were re-marked by a member of the Examination Board familiar with the examination and the subject matter. These marks were compared with the other marks and the correlations determined.

Rasch analysis was conducted using Conquest (Wu, Adams, & Wilson, 1998). The grades given by the raters who participated in the 2-day training programme against the first and second remarking grades were compared to determine inter-rater severity in allocating marks, following a similar study by Weigle (1998).

Findings

The grades obtained from the initial examination, as well as the 1st and 2nd remarking are given in Tables 1, 3, and 5 respectively. These Tables provide a comparison of the grades obtained from the different marking exercises. A correlation of the change in grades between the first set of grades and the grades obtained from the 1st and 2nd remarking is indicated in Table 7, whereas a similar correlation, from the purposive sample additional remarked by an expert, is shown Table 8.

Tables 2, 4, and 6 show the degree of change of grades on remarking from the first, initial marking to the two subsequent sets of remarking. The small type indicates actual changes in numbers of candidates which occur.

Figures 1-4 show the output from the Rasch analysis based on, respectively, remarked scripts allocated to the raters for the 1st remarking, the same scripts remarked for the second time (largely by different raters), allocated scripts marked by raters for the 2nd remarking, and also the analysis for the scripts allocated to the rater for the 2nd remarking, but referring to the 1st remarking. Figure 5 shows the difference when raters marking the initial, examination scripts are also included.

Table 1
Comparison of Grades Obtained from the Initial Marking and the 1st Remarking

Grade Initial marking	Grades from 1st remarking														Total	
	A+		A		A-		B		C		D		F		N	%
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
A+			1	33.3	1	33.3	1	33.3							3	0.1
A	2	4.9	10	24.4	17	41.5	9	22.0	3	7.3					41	1.0
A-	2	1.1	17	9.4	41	22.8	71	39.4	33	18.3	12	6.7	4	2.2	180	4.5
B			5	1.4	42	11.7	114	31.8	134	37.6	49	13.6	15	4.2	359	8.9
C			2	0.2	16	1.9	113	13.3	298	35.0	274	32.2	149	17.5	852	21.1
D					2	0.1	30	2.1	169	11.7	510	34.2	739	51.9	1450	36.0
F			-				1	0.1	5	0.4	59	5.2	1080	94.3	1145	28.4
Total	4	0.1	35	0.9	119	3.0	339	8.4	642	15.9	904	22.4	1987	49.3	4030	100

Note. N= number of candidates

Table 2
Change of Grades from the Initial Marking to the 1st Remarking

Grade Initial marking	Grades from 1st remarking														Total		
	A+	A	A-	B	C	D	F	Total									
	N	N	N	N	N	N	N	N	N								
A+	0	-1	1	-2	1	-3	1										
A	+1	2	0	10	-1	17	-2	9	-3	3							
A-	+2	2	+1	17	0	41	-1	71	-2	33	12	-4	4				
B			+2	5	+1	42	0	114	-1	134	-2	49	-3	15	-4	4	0.10
C			+3	2	+2	16	+1	113	0	298	-1	274	-2	149	-3	31	0.77
D					+3	2	+2	30	+1	169	0	510	-1	739	-2	241	5.98
F			-				+3	1	+2	5	+1	59	0	1080	-1	1236	30.67
Total									+3	5	+2	58	+1	402	0	2053	50.94
%										0.12		1.44		9.98		50.94	-

Note. Subscript numbers = number of grade changes from the initial marking to the 1st remarking

Table 3
Comparison of the Grades obtained from the Initial Marking and 2nd Remarking

Grade Initial marking	Grades from 2 nd remarking														Total	
	A+		A		A-		B		C		D		F		N	%
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
A+			1	33.3	1	33.3	1	33.3							3	.1
A	1	2.4	10	24.4	16	39.0	12	29.3	1	2.4	1	2.4			41	1.0
A-			14	7.8	49	22.7	67	37.2	30	16.7	15	8.3	5	2.8	180	4.5
B	1	0.3	3	0.8	45	12.5	113	31.5	121	33.7	63	17.5	13	3.6	359	8.9
C			2	0.2	25	2.9	136	16.0	292	34.3	272	31.9	125	14.7	852	21.1
D					6	0.4	49	3.4	237	16.3	511	35.2	647	44.6	1450	36.0
F			-				1	0.1	17	1.5	95	8.3	1032	90.1	1145	28.4
Total	2	0.05	30	0.7	142	3.5	379	9.4	698	17.3	957	23.7	1822	45.2	4030	100

Table 4
Change of Grades from Initial Marking to the 2nd Remarking

Grade Initial marking	Grades from 2nd remarking												Total	%			
	A+ N	A N	A- N	B N	C N	D N	F N										
A+	0	-1	1	-2	1	-3	1										
A	+1	0	10	-1	16	-2	12	-3	1	-4	1						
A-	1	+2	+1	14	0	49	-1	67	-2	30	-3	15	-4	5			
B	+3	+2	3	+1	45	0	113	-1		-2	63	-3	13	-4	6	0.15	
C	1		+3	2	+2	25	+1	136	0	292	-1	272	-2	125	-3	30	0.74
D				+3	6	+2	49	+1	121	0	511	-1	647	-2	231	5.73	
F					+3	1		237	+2	17	+1	95	0	1032	-1	1124	27.89
Total								+3	10	+2	94	+1	528	0	49.80	49.80	
%									0.25		2.33		13.10		49.80	-	

Table 5
Comparison of Grades obtained from the 1st and 2nd Remarking

Grade 1st re-marking	Grades from 2 nd remarking														Total N	%
	A+ N	A+ %	A N	A %	A- N	A- %	B N	B %	C N	C %	D N	D %	F N	F %		
A+					3	75	1	25							4	0.1
A			11	31.4	16	45.7	5	14.3	3	8.6					35	0.9
A-	1	0.8	8	6.7	39	32.8	50	42.0	16	13.4	5	4.2			119	3.0
B	1	0.3	7	5.9	62	18.3	129	38.1	97	28.6	34	10.0	9	2.7	339	8.4
C			4	0.6	18	2.8	134	20.9	246	38.3	187	29.1	53	8.3	642	15.9
D					3	0.3	49	5.4	239	26.4	382	42.3	231	25.6	904	22.4
F					1	0.1	11	0.6	97	4.9	349	17.6	1529	77.0	1987	49.3
Total	2	0.0	30	0.7	142	3.5	379	9.4	698	17.3	957	23.7	1822	45.2	4030	100

Table 6
Change of Grades from 1st to 2nd Remarking

Grade 1 st re-marking	Grades from 2nd remarking												Total N	%			
	A+ N	A N	A- N	B N	C N	D N	F N										
A+			-2	3	-3	1											
A		0	11	-1	16	-2	5	-3	3								
A-	+2	1	+1	8	0	39	-1	50	16	-3	5						
B	+3	1	+2	7	+1	62	0	129	-1	97	-2	34	-3	9			
C			+3	4	+2	18	+1	134	0	246	-1	187	-2	53	-3	18	0.45
D				+3	3	+2	49	+1	239	0	382	-1	231	-2	111	2.75	
F				-	+4	1	+3	11	+2	97	+1	349	0	1529	-1	581	14.42
Total						+4	1	+3	19	+2	172	+1	792	0	2336	57.97	
%							0.02		0.47		4.27		19.65		57.97	-	

Table 7
Correlations between Marks and Grades for the Different Markings ($N= 4030$)

		Initial marking		1st remarking		2nd remarking	
		Marks	Grades	Marks	Grades	Marks	Grades
Initial marking	Pearson Correlation	1	1	.812	.744	.783	.713
1st remarking	Pearson Correlation	.812	.744	1	1	.829	.753
2nd remarking	Pearson Correlation	.783	.713	.829	.753	1	1

Table 8
Correlations of Marks from the Purposive Sample ($N= 58$)

		Initial marking	1st remarking	2nd remarking	Expert marks
Initial marking	Pearson Correlation	1	.662	.325	.484
1st remarking	Pearson Correlation	.662	1	.344	.562
2nd remarking	Pearson Correlation	.325	.344	1	.833
Expert marks	Pearson Correlation	.484	.562	.833	1

Interpretation of the Findings

From the 1st remarking (different between the initial grades and those obtained from the first remarking study), Table 1 shows a substantial change in grades awarded. In total, 1512 (37.5%) candidates obtained reduced grades and 465 candidates (11.5%) were upgraded. This means that a staggering 49% of candidates received a different grade on remarking. Also, the failure rate was increased from 28.4% to 49.3% illustrating the large degree of 'over-marking' in the actual examination to ensure candidates did not fail if their marks were close to the borderline mark.

On remarking the second time (the difference between the initial grades and those obtained on 2nd remarking), Table 3 shows that, in total, 1391 (34.5%) candidates obtained reduced grades. However, 632 (15.7%) candidates were upgraded meaning 50.2% candidates received a different grade on remarking. Also, the failure rate increased from 28.4% to 45.2%.

The difference between the two sets of remarking is also compared and this is given in Table 5. Table 5 indicates that from the 1st remarking to the 2nd remarking, 710 (17.6%) candidates obtained reduced grades. However, 984 (24.4%) candidates were upgraded. In total, 42% candidates received a different grade on remarking. Also, the failure rate decreased from 49.3% to 45.2%.

The Pearson product moment correlation between marks and grades obtained in the initial marking, 1st remarking and 2nd remarking are given in Table 7. These correlations are not particularly high, the highest being that between marks for the 1st and 2nd remarking (.829).

The correlations for a purposive sample of 58 answer scripts additionally marked by an expert for the examination board are given in Table 8. The data shows the highest

correlation with the sample from the 2nd remarking, illustrating that the raters in the 2nd remarking were closer to the intended marking direction.

Tables 2, 4, and 6 indicate actual grade changes from the initial to 1st remarking, initial to 2nd remarking and 1st to 2nd remarking respectively. The Tables also indicate the magnitude of the grade changes. Specific grade change magnitudes peak at -4 in Tables 2 and 4 and -3 in Table 6, supporting the notion that the 1st and 2nd remarking are closer than the grades given in the initial marking.

A paired comparison of the means, based on the grade obtained in the original marking, was undertaken to determine whether the difference in the marking, from the initial marking, the 1st remarking and the 2nd remarking, were significant. The results are shown in Table 9. The analysis illustrates that the change of grades on remarking between the initial grades awarded in the actual examination and the grades obtained on 1st remarking were significant ($p < 0.001$) for candidates in all grades, except for the few candidates who obtained a grade of A+ on the initial marking. To obtain this data, mean candidate grades on the 1st and 2nd remarking were compared with original grades in the initial examination on a grade by grade basis.

The Rasch analysis Figures (1-5) show the range of severity among the raters. Figure 1 derives from scripts re-marked randomly by 11 raters. It shows that candidates (indicated by x's and with examination score totals remarked by the 11 raters) are of relatively low ability, lower than the arbitrary 0 on the logit scale. In the remarking, raters 5 and 7 were the most severe and rater 9, the most lenient. The spread of severity of inter-rater marking is quite large and thus the unity of inter-rater marking relatively poor. Figure 2 replicates Figure 1, but applies to the inserting of marks of the same scripts in a 2nd remarking, with the 11 raters marking similar numbers of scripts, but no longer marking, for the most part, the same scripts as in Figure 1. The range in severity of the raters is shown to be much less, although rater 5 is still the most severe and rater 9, one of the most lenient. Figure 3 derives from the same random set of scripts, but marked by the 11 raters in a 2nd remarking after receiving guidance training. Unfortunately the diversity in severity of marker is now much greater with rater 7 becoming much more severe and rater 1, much more lenient. Figure 4 replicates Figure 3, but relates to the insertion of marks on the same scripts from the 1st remarking, but by the 11 raters now marking, for the most part, different scripts. Here the range of rater severity is shown to be much less and indicates that the removing the influence of the training effect leads to more interrater agreement. Figure 5 shows the 3 marking outcomes put together as undertaken by the 11 raters, showing overall the range of severity is reduced from rater allocations on the first marking and that raters are moving towards some commonality. Unfortunately the differences are not diminished following the 2-day training programme; in fact if anything it has increased the differences among the raters! This finding is in agreement with that found by Stahl and Lunz (1991) and Weigle (1998). It seems that such training, not linked to a purposive activity in the eyes of the raters, is not particularly useful. The separation reliability is very high $>.98$, indicating that the separation of the rates into different levels is very reliable. And with significant values for chi-squared, the ordering of candidates by raters is not constant with the estimated ability measure of candidates.

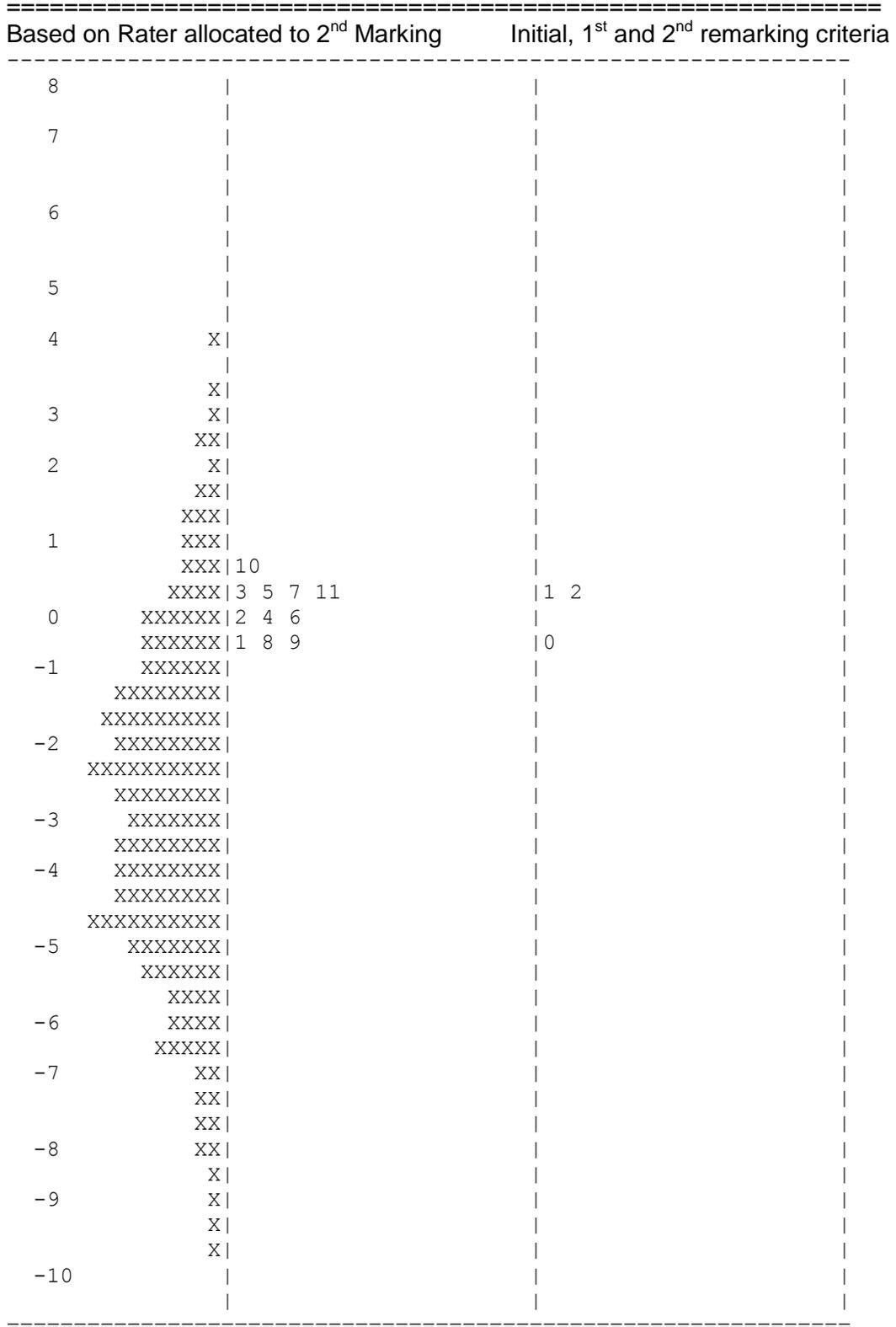
The examination also clearly shows that the questions, as a whole, were not answered well and presumably these questions were far different from those expected by the candidates. This clearly suggests a construct validity issue. If the

questions have really been developed at the appropriate level, then the teaching in schools is very much out of step. Further, recognising that the question setters are teachers from another district of the country, there is the suggestion that expectations differ from examination board to examination board within the country. And if this is the case, then examination boards would seem to have little worth.

Table 9
Paired Samples Test of Significant Differences between Means Marks based on Initial Grades

Comparisons	Paired Differences						<i>t</i>	<i>df</i>	<i>P</i>
	<i>M</i>	<i>SD</i>	<i>SE</i>	95% Confidence Interval of the Difference					
				Lower	Upper				
1st remarking mean marks compared with initial A+ grade marks	8.00	4.36	2.52	-2.83	18.83	3.18	2	.086	
2nd remarking mean marks compared with initial A+ grade marks	7.00	3.61	2.08	-1.96	15.96	3.36	2	.078	
1st remarking mean marks compared with initial A grade marks	3.93	4.94	0.77	2.37	5.49	5.09	40	.000	
2nd remarking mean marks compared with initial A grade marks	4.49	4.75	0.74	2.99	5.99	6.05	40	.000	
1st remarking mean marks compared with initial A- grade marks	4.01	5.56	0.41	3.19	4.82	9.67	179	.000	
2nd remarking mean marks compared with initial A- grade marks	4.49	5.89	0.44	3.63	5.36	10.24	179	.000	
1st remarking mean marks compared with initial B grade marks	2.73	4.82	0.25	2.23	3.23	10.73	358	.000	
2nd remarking mean marks compared with initial B grade marks	2.77	4.90	0.26	2.26	3.28	10.71	358	.000	
1st remarking mean marks compared with initial C grade marks	2.39	5.05	0.17	2.05	2.73	13.83	851	.000	
2nd remarking mean marks compared with initial C grade marks	1.85	5.17	0.18	1.50	2.19	10.43	851	.000	
1st remarking mean marks compared with initial D grade marks	2.29	4.92	0.13	2.04	2.55	17.77	1449	.000	
2nd remarking mean marks compared with initial D grade marks	1.45	5.33	0.14	1.18	1.73	10.40	1449	.000	
1st remarking mean marks compared with initial F grade marks	-0.32	3.88	0.11	-0.55	-0.10	-2.87	1144	.004	
2nd remarking mean marks compared with initial F grade marks	-1.06	4.35	0.13	-1.32	-0.81	-8.28	1144	.000	

Figure 5
Rater Effect on 3 Levels of Marking
Map of Latent Distributions and Response Model Parameter Estimates



Each 'x' represents 24.7 candidates

Discussion

The remarking gave significantly different total marks and different grades at all levels except for those few candidate who obtained a grade of A+. It is thus clear that the hypothesis put forward stipulating there is no significant change in grades on remarking is found to be incorrect. There are change of grades for a significant number of candidates (except A+ grades). This of course is very undesirably, as it undermines the validity of the whole examination. The legitimisation of setting up and running expensive examination boards, to run high stakes examinations, is under threat. It is important that steps be put in place to redress this concern.

Clearly the marking of these answer scripts is very unstable and there is little reliability in the marks obtained. It is staggering that remarking can change the grade awarded by as much as 4 grade points in some cases (see Tables 1 and 2 where (i) the grades for 4/5 candidates were changed on remarking from A- to F in each case and (ii) 1 candidate, on remarking, changed grade from F to B). Furthermore this highlights the chance of personal error/bias in marking and the minimal student-teacher relationship regarding the question content. It is thus clear that the study points to the need for a more reliable marking system (if candidate total marks as well as grades, are to be taken as standard and there is a need to minimise personal bias /error and encourage better teaching-learning situations in institutions). It is suggested that detailed marking guidelines (and even full answers where questions tend towards a more structured format) are produced by question setters, not by head examiners. In this way, the construct validity of the questions is enhanced, because the marks are more likely related to the intentions of the questions. Such detailed mark schemes will also assist question paper moderators better understand the questions set and again aid the validity of the question paper.

The need to develop a more reliable marking system suggests that better guidance to raters is required, illustrating how to mark candidate answer scripts more effectively. It is suggested that more reliable marking can be obtained by:

- (a) presenting each examiner with a detailed marking scheme with actual answers for marking the script so that it becomes easier to award marks to common criteria;
- (b) setting questions that better lend themselves to more objective and hence more specific marking schemes. This means, for subjects where creativity and presentation are less important than the expression of conceptual ideas, moving away from essay-type questions with its subjective marking procedures and towards questions which have a set structure and which can be used to award marks in a more objective manner. Such questions can be structured questions which are broken down into sub-parts and each part is marked separately based on a set of specific criteria (Verna, Chhatal, & Singh, 1997);
- (c) requiring all examiners to mark initially the same set of answer scripts (sample marking using 6-9 answers scripts across the mark range) and then discussing the marks awarded to determine a common set of detailed marking guidelines for all raters.

One approach suggested earlier is to consider double marking by averaging marks from the initial and first remarking. Table 8 compares the grades awarded by

the averaging with those from a third marking (the 2nd remarking). As the 2nd remarking was carried out after a 2 day in-service course, these marks can be expected to be more authentic. This is supported by a correlation of marks (Table 9) from a purposive sample of 58 scripts in which the the grades awarded by all 3 forms of marking were compared with the marks awarded by an 'expert' familiar with the questions and the answers expected. The finding suggest, however, that double marking is not likely to be as effective. Add to this the time and cost factor (Newton, 1996; Verma, Chhatwal, & Singh, 1997) and clearly double marking is not really as appropriate as taking steps to marking the questions more objectively and mandating a detailed marking scheme.

Conclusion

The examination questions are poorly marked, with far too great a leniency for candidates with lower marks. The marking improved on the remarking, but the guidance given before raters attempted the second remarking did not show substantial gains. A comparison of the rater leniency and harshness changed little. The inter-rater reliability of the examination under scrutiny was found to be low. This is alarming given that the examination is high stakes and the grades awarded to candidates means much in determining their future. Clearly the examiners were given insufficient guidance as shown by the greater correlation with an expert rater when marking after a 2 day in-service course. But the lack of agreement in the marking of all questions suggests that examiners do need a detailed answer script, besides practice in its use by marking of sample scripts.

Noting that most candidates obtained grades in the middle of the range and that even in this range there were substantial discrepancies between grades awarded on remarking, inter-rater consistency must be a great cause for concern. Examination boards clearly need to give much more attention to using raters who are prepared for the task and who recognise the importance of reliable marking. The need for examination boards is being undermined by the lack of inter-rater consistency.

This was an initial study into reliability of the marking and shows that all 3 hypotheses need to be rejected, although there is some evidence that limited in-service training did lead to more appropriate marking. Further studies can be undertaken such as the effects of motivation and rewards given to the raters and the careful scrutinising of marks after being recorded on the candidate answer scripts. These are all matters, suggested by raters, to be of concern.

References

- Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *The Statistician*, *32*(3), 307-317.
- Baird, J-A., Greatorex, J., & Bell, J. P. (2004). What makes marking reliable? Experiments with UK Examinations. *Assessment in Education*, *11*(3), 331-348.

- Bédard, M., Martin, N., Kruger, P., & Brazil, B. (2000). Assessing reproducibility of data obtained with instruments based on continuous measurements. *Experimental Aging Research*, *26*, 353-365.
- Berkowitz, D., Wolkowitz, B., Firsch, R., & Kopriva, R. (2000). *The use of tests as part of high-stakes decisions for students: A resource guide for educators and policy-makers*. Washington DC: US Dept. of Education.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, *20*, 37-46.
- Brooks, V. (2004). Double marking revisited. *British Journal of Educational Studies*, *52*(1), 29-46.
- Campbell, A. (2005). Application of ICT and rubrics to the assessment process where professional judgement is involved: the features of an e-marking tool. *Assessment and Evaluation in Higher Education*, *30*(5), 529-537.
- Chase, C. I. (1986). Essay test scoring: Interaction of relevant variables. *Journal of Educational Measurement*, *23*, 33-42.
- Crossley, J., Davies, H., Humphris, G., & Jolly, B. (2002). Generalisability: a key to unlock professional assessment, *Medical Education*, *36*, 972-978.
- Dahl, T. I. (2006). When precedence sets a bad example for reform: conceptions and reliability of a questionable high stakes assessment practice in Norwegian universities. *Assessment in Education*, *13*(1), 5-27.
- Dimitrov, D. M. (2002). Reliability: Arguments for multiple perspectives and potential problems with generalization across studies. *Educational Psychological Measurement*, *62*(5), 783-801.
- Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessment: A many facet Rasch analysis. *Language Assessment Quarterly*, *2*(3), 197-221.
- Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training. Does it work? *Language Assessment Quarterly*, *2*(3), 175-196.
- Engelhard, G. Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, *31*(2), 93-112.
- Lunz, M., Wright, B., & Linacre, J. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, *3*(4), 331-345.
- Marsh, H. W., & Ireland, R. (1987). The assessment of writing effectiveness: a multidimensional perspective. *Australian Journal of Psychology*, *39*, 353-367.
- Moon, T. R., & Hughes, K. R. (2002). Training and scoring issues involved in large-scale writing assessments. *Educational Measurement: Issues and Practice*, *21*(2), 15-19.
- Munro, N., Denney, M. L., Nughani, A., Foulkes, J., Wilson, A., & Tate, P. (2003). Ensuring reliability in UK written tests of general practice: The MRCGP examination 1998-2003. *Medical Teacher*, *27*(1), 2005, 37-45.
- Newton, P. E. (1996). The reliability of marking of general certificate of secondary education scripts: Mathematics and English. *British Educational Research Journal*, *22*(4), 405-420.
- Nuttall, D. L., & Willmott, A.S. (1972). *British examinations: Techniques of analysis P12*. Slough: NFER Publishing.

- Van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, resources and practical implications. *Advances in Health Sciences Education, 1*, 41-67.
- Verma, M., Chhatwal, J., & Singh T. (1997). Reliability of essay type questions-- Effect of structuring, *Assessment in Education, 4*(2), 265-71.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing, 15*, 263-287.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). ACER ConQuest. Melbourne, ACER.
- Yen, W. M. & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education, 4*(3), 209-228.

Jack Holbrook is a visiting professor in science education at the University of Tartu, Estonia and an independent consultant, specialising in assessment, curriculum and teacher education at the primary and secondary level. He has a PhD in Chemistry from the University of London and was trained as a secondary school teacher before spending most of his career in teacher education - in the UK, Africa and Asia. His main area of research is in the philosophy of science education and relating this to the teaching of science subjects. He is the current President of the International Council of Associations for Science Education (ICASE).



A brief history of educational assessment in the Philippines

Carlo Magno

De La Salle University, Manila, Philippines

Abstract The report presents a primer on the history of educational assessment in the Philippines. The history of educational assessment is described by the different pillars that contributed to its development. These factors include timelines of government mandates, studies done in the national level, universities that shape experts in the field, professional association, and pioneered researches. The history of educational assessment is divided by the early years, contemporary period, and future direction. The early years include the Monroe survey, Research, Evaluation and Guidance Division of the Bureau of Public Schools, Economic Survey Committee, Prosser Survey, and UNESCO survey. The contemporary period is marked by the EDCOM report of 1991, Philippine Education Sector Study, the role of the Fund for Assistance to Private Organization (FAPE) and the creation of the Center for Educational Measurement (CEM) and the Asian Psychological Services and Assessment Corporation (APSA). The article described future directions of educational assessment about the expanding role of educational assessment specialists in school and creating niches for future studies.

Keywords: *Educational assessment, history of educational assessment*

Introduction

The growth and development of a scientific discipline depends on the contribution of experts and events that led to the construction of knowledge. The history of educational assessment in the Philippines is likewise shaped by the contribution of experts in line with research, government policies on educational development, universities that produced competent experts and professional associations that provide a venue for experts to promote knowledge. I shall briefly present the history of educational assessment, measurement, and evaluation in the Philippines based on these four pillars that influence the development of the field.

It is important to put into account the history of educational assessment in the Philippines to reorient students and the primary practitioners of assessment in educational contexts about its rich origins. Students taking courses in educational measurement and evaluation and similar courses (e.g., psychological testing, psychometrics) are usually oriented with the western development and history of

educational assessment. As a consequence, they are more familiar with the development of educational assessment outside of their country. Students are familiar with the Binet and Weschler scales but not with local intelligence and personality measures in the Philippines. This scenario is considered limiting because in order for one to contribute to the development of educational assessment, one must know where to begin the development and which specific area one can contribute.

A timeline of events that shaped educational assessment allows contemporary experts and practitioners to further develop and advance the use of methods, techniques, theory, and approaches in the field (Magno & Ouano, 2009). The first part of the article enumerates the pioneering studies in educational assessment in the Philippines (Monroe Survey, economic survey committee, Prosser survey, UNESCO, EDCOM, PESS) and government policies that were made (Research, Evaluation, and Guidance Division of the Bureau of Public Schools, Joint Congressional Committee on Education). The second part presents the universities that offer programs on educational measurement and evaluation, professional associations, as well as educational testing and assessment institutions.

Assessment in the Early Years

Monroe Survey (1925). Formal Assessment in the Philippines started as a mandate from the government to look into the educational status of the country (Elevazo, 1968). The first assessment was conducted through a survey authorized by the Philippine legislature in 1925. The legislature created the Board of Educational Survey headed by Paul Monroe, and later, this board appointed an Educational Survey Commission who was also headed by Paul Monroe. This commission visited different schools in the Philippines. The commission observed different activities conducted in schools around the Philippines. The results of the survey reported the following:

1. The public school system that is highly centralized in administration needs to be humanized and made less mechanical.
2. Textbook and materials need to be adapted to Philippine life.
3. The secondary education did not prepare for life and recommended training in agriculture, commerce, and industry.
4. The standards of the University of the Philippines were high and should be maintained by freeing the university from political interference.
5. Higher education be concentrated in Manila.
6. English as medium of instruction was best. The use of local dialect in teaching character education was suggested.
7. Almost all teachers (95%) were not professionally trained for teaching.
8. Private schools except under the religious groups were found to be unsatisfactory.

Research, Evaluation, and Guidance Division of the Bureau of Public Schools. This division started as the Measurement and Research Division in 1924

which was an off-shoot of the Monroe Survey. It was intended to be the major agent of research in the Philippines. Its functions were to:

1. coordinate the work of teachers and supervisors in carrying out testing and research programs
2. conduct educational surveys
3. construct and standardize achievement tests

Economic Survey Committee. In a legislative mandate in 1927, the director of education created the Economic Survey Committee headed by Gilbert Perez of the Bureau of Education. The survey studied the economic condition of the Philippines. They made recommendations as to the best means by which graduates of the public school could be absorbed to the economic life of the country. The results of the survey pertaining to education include the following:

1. Vocational education is relevant to the economic and social status of the people.
2. It was recommended that the work of the schools should not be to develop a peasantry class but to train intelligent, civic-minded homemakers, skilled workers, and artisans.
3. Devote secondary education to agriculture, trades, industry, commerce, and home economics.

The Prosser Survey. In 1930, C. A. Prosser made a follow-up study on vocational education in the Philippines. He observed various types of schools and schoolwork. He interviewed school officials and businessmen. He recommended in the survey to improve various phases of the vocational educational such as 7th grade shopwork, provincial trade schools, practical arts training in the regular high schools, home economics, placement work, gardening, and agricultural education.

Other Government-Commissioned Surveys. After the Prosser survey, there were several surveys conducted to determine mostly the quality of schools in the country after the 1930s. All of these surveys were government-commissioned like the Quezon Educational Survey in 1935 headed by Dr. Jorge C. Bacobo. Another study was made in 1939, which is a sequel to the Quezon Educational Surveys which made a thorough study of existing educational methods, curricula, and facilities and recommended changes on financing public education in the country. This was followed by another congressional survey in 1948 by the Joint Congressional Committee on Education to look into the independence of the Philippines from America. This study employed several methodologies.

UNESCO Survey (1949). The UNESCO undertook a survey on Philippine Education from March 30 to April 16, 1948 headed by Mary Trevelyan. The objective of the survey was to look at the educational situation of the Philippines to guide planners of subsequent educational missions to the Philippines. The report of the surveys was gathered from a conference with educators and layman from private and public school all over the country. The following were the results:

1. There is a language problem.

2. There is a need to for more effective elementary education.
3. Lengthening of the elementary-secondary program from 10 to 12 years.
4. Need to give attention to adult education.
5. Greater emphasis on community school
6. Conduct thorough surveys to serve as basis for long-range planning
7. Further strengthening of the teacher education program
8. Teachers income have not kept pace with the national income or cost of living
9. Delegation of administrative authority to provinces and chartered cities
10. Decrease of national expenditures on education
11. Advocated more financial support to schools from various sources

After the UNESCO study, it was followed by further government studies. In 1951, the Senate Special Committee on Educational Standards of Private schools undertook a study about private schools. The study was headed by Antonio Isidro where he investigated the standards of instruction in private institutions of learning, and provide certificates of recognition in accordance with their regulations.

In 1967, the Magsaysay Committee on General Education was created which was financed by the University of the East Alumni Association. In 1960, the National Economic Council and the International Cooperation Administration surveyed public schools. The survey was headed by Vitaliano Bernardino, Pedro Guiang, and J. Chester Swanson. Three recommendations were provided to public schools: (1) To improve the quality of educational services, (2) To expand the educational services, and (3) To provide better financing for the schools.

The assessment conducted in the early years were mostly mandated and/or commissioned by the government. The private sectors were not included in the studies as proponents. Most of these studies are usually headed by foreign counterparts such as the UNESCO, Monroe, and Swanson survey. The focus of the assessments was on the implementation of education in the country. These national researches were conducted with the need of the government to determine the status of the education in the country.

Assessment in the Contemporary Period and Future Directions

EDCOM Report (1991). The EDCOM report in 1991 indicated that high dropout rates especially in the rural areas were significantly marked. The learning outcomes as shown by achievement levels show mastery of the students in important competencies. There were high levels of simple literacy among both 15-24 year olds and 15+ year olds. Repetition in Grade 1 was the highest among the six grades of primary education which reflects the inadequacy of preparation among the young children. The children with which the formal education system had to work with at the beginning of EFA were generally handicapped by serious deficiencies in their personal constitution and in the skills they needed to successfully go through the absorption of learning.

Philippine Education Sector Study (PESS-1999). Jointly conducted by the World Bank and Asian Development Bank, the PESS recommended the following:

1. A moratorium on the establishment of state colleges and universities;
 2. Tertiary education institutions be weaned from public funding sources;
- and
3. A more targeted program of college and university scholarships.

Aside from the government initiatives in funding and conducting surveys that applied assessment methodologies and processes, the government was also involved in the practice of testing where they screened government employees in 1924. Grade four to fourth year high school students were tested in the national level in 1960 to 1961. Private organizations also spearheaded the enrichment of assessment practices in the Philippines. These private institutions are the Center for Educational Measurement (CEM) and the Asian Psychological Services and Assessment Corporation (APSA).

Fund for Assistance to Private Education (FAPE). FAPE started with testing programs such as the guidance and testing program in 1969. They started with the College Entrance Test (CET) which was first administered in 1971 and again in 1972. The consultants who worked with the project were Dr. Richard Pearson from the Educational Testing Service (ETS), Dr. Angelina Ramirez, and Dr. Absraham Felipe. FAPE then worked with the Department of Education, Culture, and Sports (DECS) to design the first National College Entrance Exam (NCEE) that served to screen fourth year high school students who were eligible to take a formal four-year course. There was a need to administer a national test then because most universities and colleges do not have an entrance exam to screen admission of students. Later, the NCEE was completely endorsed by FAPE to the National Educational Testing Center of the DECS.

The testing program of FAPE continued where they developed a package of four tests which are the Philippine Aptitude Classification Test (PACT), the Survey/Diagnostic Test (S/DT), the College Scholarship Qualifying Test (CSQT), and the College Scholastic Aptitude Test (CSAT). In 1978, FAPE institutionalized an independent agency called the Center for Educational Measurement (CEM) that undertook the testing and other measurement services since then.

Center for Educational Measurement (CEM). CEM started as an initiative of the Fund for Assistance to Private Education (FAPE). The CEM was then headed by Dr. Leticia M. Asuzano who was the executive vice-president. Since then, several private schools have been members in the CEM network. Since 1960, the CEM developed over 60 tests focused on education like the National Medical Admissions Test (NMAT). The main advocacy of the CEM is to improve the quality of formal education through its continuing advocacy and supporting systematic research. The CEM promotes the role of educational testing and assessment to improve the quality of formal education at the institutional and

systems levels. Through testing, the CEM helps to improve effectiveness for teaching and student guidance.

Asian Psychological Services and Assessment Corporation (APSA). In 1982, there was a growing demand for testing not only in the educational setting but in the industrial setting. Dr. Genevive Tan who was then a consultant in various industries felt the need to measure the Filipino ‘psyche’ in a valid way because most industries use foreign tests. The Asian Psychological Services and Assessment Corporation (APSA) was then created to fulfill this need. In 2001, headed by Dr. Leticia Asuzano, the former Executive Vice President of CEM, the APSA extended its services for testing in the academic setting because of the felt need for quality educational testing in the private sector.

The mission of APSA includes a commitment to deliver excellent and focused assessment technologies and competence-development programs to the academe and the industry to ensure the highest standards of scholastic achievement and work performance, and to ensure stakeholders' satisfaction in accordance with company goals and objectives. The APSA envisions itself as the lead organization in assessment and a committed partner in the development of quality programs, competencies, and skills for the academe and industry.

The APSA has numerous tests that measure mental ability, clerical aptitude, work habits, and supervisory attitudinal survey. For the academic side, they have tests for basic education, Assessment of College Potential, Assessment of Nursing Potential, and the Assessment of Engineering Potential. In the future, the first Assessment for Accounting Potential and Assessment of Teachers and Maritime Potential will be available for use in higher education.

The APSA pioneered on the use of new mathematical approaches (IRT Rasch Model) in developing tests which went beyond the norm-reference approach. In 2002, they launched the standards-based instruments in the Philippines that serve as benchmarks in the local and international schools. Standards-based assessment (1) provides an objective and relevant feedback to the school in terms of its quality and effectiveness of instruction measured against national norms and international standards; (2) identifies the areas of strengths and the developmental areas of the institution's curriculum; (3) pinpoints competencies of students and learning gaps which serve as basis for learning reinforcement or remediation; and (4) provides good feedback to the student on how well he has learned and his readiness to move to a higher educational level.

Building Future Leaders and Scientific Experts in Assessment and Evaluation in the Philippines

There are only few universities in the Philippines that offer graduate training on Measurement and Evaluation. The University of the Philippines offer a master's program in education specialized in measurement and evaluation and doctor of philosophy in research and evaluation. The De La Salle University, Manila has a Master of Science in Psychological Measurement offered by the Psychology Department. Their college of education, which is a center for excellence, also offers the graduate program Master of Arts in Educational Measurement and Evaluation,

and Doctor of Philosophy in Educational Psychology major in Research, Measurement, and Evaluation. The Philippine Normal University also offers a degree in Master of Arts in Education specialized in measurement and evaluation. The graduate school of Mirriam College also offer the Masters of Arts in Education major in Measurement and Evaluation that addresses the need to fill gaps in the application of statistics to measurement and evaluation.

Some practitioners in line with measurement and evaluation were trained in other countries such as in the United States and Europe. There is a greater call for educators and those in the industry involved in assessment to be trained to produce more experts in the field.

Professional Organization on Educational Assessment

Aside from the government and educational institutions, the Philippine Educational Measurement and Evaluation (PEMEA) is a professional organization geared in promoting the culture of assessment in the country. The organization was initiated during the National Conference on Educational Measurement and Evaluation (NCEME) in 2008. It was headed by Dr. Rose Marie Salazar-Clemeña who was then the dean of the College of Education in De La Salle University-Manila (formerly EVP of De La Salle-College of Saint Benilde) together with the De La Salle-College of Saint Benilde's Center for Learning and Performance Assessment. It was attended by participants all around the Philippines. The theme of the conference was "Developing a Culture of Assessment in Learning Organizations." The conference aimed to provide a venue for assessment practitioners and professionals to discuss the latest trends, practices, and technologies in educational measurement and evaluation in the Philippines. In the said conference, the PEMEA was formed with the following purposes:

1. To promote standards in various areas of education through appropriate and proper assessment.
2. To provide technical assistance to educational institutions in the area of instrumentation, assessment practices, benchmarking, and process of attaining standards.
3. To enhance and maintain the proper practice of measurement and evaluation in both local and international level.
4. To enrich the theory, practice, and research in evaluation and measurement in the Philippines.

The first batch of board of directors elected for the PEMEA in 2008 are Dr. Richard DLC Gonzales as President (University of Santo Tomas Graduate School), Neil O. Pariñas as Vice President (De La Salle-College of Saint Benilde), Dr. Lina A. Micalat as Secretary (De La Salle-College of Saint Benilde), Marife M. Mamauag as Treasurer (De La Salle-College of Saint Benilde), Belen M. Chu as PRO (Philippine Academy of Sakya). The board members are Dr. Carlo Magno (De La Salle University, Manila), Dennis Alonzo (University of Southeastern Philippines, Davao City), Paz H. Diaz (Miriam Collage), Ma. Lourdes M. Franco

(Center for Educational Measurement), Jimelo S. Tipay (De La Salle–College of Saint Benilde), and Evelyn Y. Sillorequez (Western Visayas State University).

Aside from the universities and professional organization that provide training on measurement and evaluation, the field is growing in the Philippines because of the periodicals that specialize in the field. The CEM has its “Philippine Journal of Educational Measurement.” The APSA is continuing to publish its “APSA Journal of SBA Research.” And the PEMEA launched the “Assessment Handbook” and “Educational Measurement and Evaluation Review.” Aside from these journals, there are Filipino experts from different institutions who published their works in international and abstracted journals.

The Psychology Act

In March 2010, the president of the Philippines Gloria Macapagal-Arroyo signed the Republic Act No. 10029 which is called the ‘Philippine Psychology Act of 2009.’ This Act further maintains the development of assessment which is part of the services offered by psychologists. The practice of assessment by licensed psychologists can range in the school setting. In the said Act, psychological assessment which can be applied in diverse settings include (Article III, Definition of Terms):

gathering and integration of psychology-related data for the purpose of making a psychological evaluation, accomplished through a variety of tools, including individual tests, projective tests, clinical interviews, and other psychological assessment tools, for the purpose of assessing diverse psychological functions including cognitive abilities, aptitudes, personality characteristics, attitudes, values, interests, emotions and motivations, among others, in support of psychological counseling, psychotherapy and other psychological interventions

The scope of psychological assessment also covers implications in the educational context especially in the assessment cognitive abilities, aptitudes, attitudes, values, interests, emotions, and motivations which also covers the scope of educational assessment. This defined role implies a collaborative work between assessment specialists and psychologists.

In terms of defining the person involved in measurement and evaluation, the Philippine Psychology Act defined it within the work of a “Psychometrician.” A psychometrician is involved in “(1) Administering and scoring of objective personality tests, structured personality tests, pen-and-paper intelligence, achievement, and interest tests, excluding projective tests and other higher level forms of psychological tests; (2) Interpreting results of the same and preparing a written report on these results; (3) Conducting preparatory intake interviews of clients for psychological intervention sessions” (Philippine Psychology Act, Article III, Definition of Terms).

Given the delineated role of a psychometrician, educational institutions are recommended to have a person who is qualified in the administration, scoring, and interpretation of test results. Test results are useful information in the educational

context and a psychometrician is defined to have the role over the use of psychological tests.

Future Directions

The field of educational assessment, measurement, and evaluation is shaped by numerous forces due to its range of applications. The article presented that the production of research and studies make way in the development of national policies and the creation of associations. Given the advent of the Philippine Psychology Act, the educational assessment is carefully delineated with that of psychologists specifically in assessment in schools. In terms of the studies produced so far, the future of the educational assessment, measurement, and evaluation is stretching itself to produce more the use of mathematical models in item analysis. A group of Filipino authors produced a special issue in the *International Journal of Educational and Psychological Assessment* in 2009 about comparison of classical test theory and item response theory. The Filipino authors made use of a variety of model parameters using a modern test theory approach (Hernandez, 2009; Magno, 2009; Morales, 2009; Tiratira, 2009). In other social science journals, Amora and Bernardo (2009) made use of a one-parameter Rasch model for items in a vocabulary learning strategies. Moreover, Pedrajita (2009) used differential item functioning to detect item bias in a chemistry achievement test. Amarnani (2009) made an introductory essay explaining the conceptual focus of item response theory as alternative to classical test theory. I am hoping that more Filipinos would create a niche in the use of mathematical models in item analysis.

In terms of the practice of educational assessment in the Philippines, the role of assessment specialists is increasing widely due to the demand for quality assurance in schools that especially in teaching and implementation of programs. In this aspect, more collaboration is expected between teachers and psychometricians to improve and understand better the learners.

References

- Amarnani, R. (2009). Two theories, one theta: A gentle introduction to item response theory as an alternative to classical test theory. *The International Journal of Educational and Psychological Assessment*, 3, 104-109.
- Amora, J. T., & Bernardo, A. S. (2009). Testing and Reducing L2 Vocabulary Learning Strategies Inventory Using Rasch Model. *Philippine ESL Journal*, 3, 38-73.
- Asian Psychological Services & Assessment Corporation*. (n. d.) Retrieved May 1, 2010 from <http://apsa.chlopez.net/mission.htm>
- Center for Educational Measurement, Inc.* (n. d.). Retrieved May 1, 2010 from http://www.cem-inc.org.ph/about/about_CEM.htm
- Elevazo, A. O. (1968). *Educational research and national development*. Manila: National Science Development Board.
- Hernandez, A. (1999). PCER examines the education system anew: Releases first draft recommendations. *Presidential Commission on Educational Reform*, 1, 1-5.

- Hernandez, R. (2009). Comparison of the item discrimination and item difficulty of the quick-mental aptitude test using CTT and IRT methods. *The International Journal of Educational and Psychological Assessment, 1*, 12-18.
- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment, 1*, 1-11.
- Magno, C., & Ouano, J. (2009). *Designing written forms of assessment*. QC, Manila: Phoenix Pub.
- Miriam College, Manila, Philippines. (n. d.). Graduate school academic programs Retrieved May 1, 2010 from <http://www.mc.edu.ph/academicunits/graduateschool/academicprogram.html>
- Morales, R. (2009). Evaluation of mathematics achievement test: A comparison between CTT and IRT. *The International Journal of Educational and Psychological Assessment, 1*, 31-38.
- Pedrajita, J. Q. (2009). Using logistic regression to detect biased test items. *The International Journal of Educational and Psychological Assessment, 2*, 54-73.
- Philippine Educational Measurement and Evaluation Association (2009, January 1). *National conference on educational measurement and evaluation association*. Retrieved May 1, 2010, from <http://pemea.club.officelive.com/default.aspx>
- Philippine Educational Measurement and Evaluation Association (2009, January 1). *Educational measurement and evaluation review*. Retrieved May 1, 2010, from <http://pemea.club.officelive.com/default.aspx>
- Philippine Educational Measurement and Evaluation Association (2009, January 1). *Continuing education program, The assessment handbook*. Retrieved May 1, 2010, from <http://pemea.club.officelive.com/default.aspx>
- Philippine Normal University, Manila, Philippines. (n. d.). *Academic programs* Retrieved May 1, 2010 from http://www.pnu.edu.ph/UNIVERSITY_ACADEMIC_PROGRAMS.php
- Psychological Association of the Philippines. (2010, March 16). *An act to regulate the practice of psychology*. Retrieved May 1, 2010, from <http://papnews.wordpress.com/2009/06/21/psychology-bill-update-2/>
- Salamanca, B. S. (1981). *FAPE the first decade*. Manila: FAPE.
- Tiratira, N. (2009). Cutoff scores: The basic Angoff method and the item response theory method. *The International Journal of Educational and Psychological Assessment, 1*, 39-47.



An evaluation of the teaching-learning process: A transformative learning set-up

Violeta C. Valladolid

De La Salle University, Manila, Philippines

Abstract A new teaching-learning paradigm, known as Transformative Learning (TL) framework was formulated as a result of the development of the new General Education Curriculum (GEC). Using Huitt's Transactional Model (2003), various evaluation studies were lined up to determine whether the TL framework has been implemented as planned and has achieved its objectives. Five research initiatives have been carried out: (1) evaluation forms for faculty handling TL classes were developed and validated; (2) an implementation evaluation was conducted to identify problems in the early stage of program implementation; (3) a descriptive study of the TL faculty was conducted; (4) rubrics to assess the teaching-learning process were formulated; and (5) a Thinking Skills Test was developed to determine significant changes in the students' level of critical thinking and problem solving skills due to their exposure to the TL set-up.

Keywords: *Evaluation, transformative learning teaching-learning process*

Introduction

In SY 2006-2007, the De La Salle University officially implemented a new General Education Curriculum (GEC), known as the Lasallian Pedagogical Framework of Transformative Learning (TL). The key part of the new GEC is the development of, knowledge and skills of students to engage in more specialized study in various disciplines, capacity to use knowledge and skills for varied developmental pursuits, and foundation for lifelong learning. It envisions that all DLSU students become good at a wide variety of modes of learning, and at engaging in a variety of modes of thinking, such as historical inquiry, scientific and quantitative inquiry, humanistic, interpretative, and aesthetic inquiry, and ethical, value-based, or faith-based inquiry (DLSU Committee on Lasallian Pedagogical Framework, 2004).

In this new framework of instruction, the student is the primary agent of learning. The framework includes the following areas: the students' learning process, the teacher's work, the students' learning environment, and assessment of student learning.

The transition from the old and traditional system of transmission to transformative learning was expected to bring about changes in curriculum design, syllabus construction, lesson planning, use of class time, grading system, faculty evaluation. Thus, various programs were initiated by the University to meet the demands of the TL set-up. Among these initiatives were faculty and trainers' training, curriculum review and development, development of modules, conduct of seminar-workshops on contemporary pedagogy, and design of technology-supported learning environments, technology-based materials, and on-line courses.

One important initiative of the University is the evaluation of the teaching learning process in the transformative learning set-up. To this date, five research initiatives have been carried out: (1) development of faculty evaluation forms for TL classes, (2) evaluation of the implementation of a TL program, (3) a descriptive study of TL faculty, (4) development of rubrics for TL classes, and (5) development and validation of the DLSU Thinking Skills Test.

Theoretical Perspectives on Transformative Learning

Transformative learning (or transformational learning) is a process of getting beyond gaining factual knowledge alone to instead become changed by what one learns in some meaningful way. It involves questioning assumptions, beliefs and values, and considering multiple points of view (Mezirow, 2000). There are different theories or perspectives on transformative learning.

The Transformational Learning Theory of Mezirow (1978) is "constructivist, an orientation which holds that the way learners interpret and reinterpret their sense experience is central to making meaning and hence learning". It came from his earlier theory of *perspective transformation*. According to him, perspective transformation is "the process of becoming critically aware of how and why our assumptions have come to constrain the way we perceive, understand, and feel about our world; changing these structures of habitual expectation to make possible a more inclusive, discriminating, and integrating perspective; and finally, making choices or otherwise acting upon these new understandings."

The theory posits that for learners to change their *meaning structures* that is, beliefs, attitudes, and emotional reactions - they must engage in critical reflection on their experiences, which in turn leads to a transformation of perspective. These meaning structures are divided into two categories namely, (1) meaning perspectives, and (2) meaning schemes. The meaning perspectives are defined as "broad sets of predispositions resulting from psycho-cultural assumptions which determine the horizons of our expectations" (Mezirow, 1997). A transformation in meaning perspective can happen only through perspective taking, assimilating the perspectives of others; it is not role taking. Perspective taking implies a conscious recognition of the difference between one's old perception and the new one and a desire to appropriate the newer perspective because it is of more value (Mezirow, 1978).

Meaning schemes are the more specific dimensions of one's personal frame of reference or meaning perspective. They contain the specific beliefs, knowledge, feelings, and value judgments that become articulated in an interpretation. They

may change as a person adds to or integrates ideas within an existing scheme and, in fact, this transformation of meaning schemes occurs routinely through learning.

Taylor (1998) and Imel (1998), in their review of literature, cited Boyd's model of transformation. Transformation for Boyd is defined as a "fundamental change in one's personality involving [together] the resolution of a personal dilemma and the expansion of consciousness resulting in greater personality integration" (Taylor, 1998, p.20). The process of discernment is central to transformative education. Discernment calls upon such extra-rational sources as symbols, images, and archetypes to assist in creating a personal vision or meaning of what it means to be human. It includes manifestation of receptivity, recognition, and grieving. First, an individual must be receptive or open to receiving alternative expressions of meaning. Grieving, considered to be the most critical phase of the discernment process, takes place when an individual realizes that old patterns or ways of perceiving are no longer relevant, moves to adopt or establish new ways, and finally, integrates old and new patterns (Imel, 1998).

Perhaps one of the best definitions of transformative learning was put forward by O'Sullivan (2003). He said that transformative learning involves experiencing a deep, structural shift in the basic premises of thought, feelings, and actions. It is a shift of consciousness that dramatically and irreversibly alters one's way of being in the world. Such a shift involves understanding of one's self and self-locations; one's relationships with other humans and with the natural world; one's understanding of relations of power in interlocking structures of class, race and gender; one's body awareness and visions of alternative approaches to living; and one's sense of possibilities for social justice and peace and personal joy (O'Sullivan, 2003, p. 327).

Framework for the Evaluation of the Teaching Learning Process in the TL Set-up

The evaluation of the teaching learning in the Lasallian Pedagogical Framework of Transformative Learning is guided by Huit's Transactional Model of Teaching-Learning Process (2003). This model was formulated based on the perspective of the systems theory. The model was developed to categorize the variables essential in answering the question: "Why do some students learn more than other students do in classroom and school settings?"

According to the model, the reasons can be classified into four categories: (1) context, (2) input, (3) classroom processes, and (4) output. Context includes all factors outside of the classroom that might influence teaching and learning; inputs are qualities or characteristics which teachers and students bring with them to the classroom experience; classroom processes are teacher and student behaviors in the classroom as well as other variables such as classroom climate and teacher-student relationships; and output includes measures of student learning apart from the normal processes.

The output is the most important of these categories because the variables in the rest of the categories are used to predict or relate to the variables measured in this output. The second most important category is the classroom processes category, which includes all the variables that would occur in the classroom. There are three subcategories on the classroom processes category: Teacher behavior,

student behavior, others/miscellaneous. Teacher behavior consists of all the actions a teacher makes in the classroom. The third major category of variables, input, refers to descriptions of teachers and students prior to their coming into the classroom. Two important subcategories of variables are: teacher characteristics and student characteristics. Among the most important teacher characteristics are teacher's values and beliefs, efficacy, and teacher's knowledge. Among the student characteristics which have been related to classroom behavior and student achievement are student's prior knowledge, intelligence or academic ability, and study habits. The context includes all variables outside of the classroom that have an impact on teacher and student characteristics, classroom processes, and output. Subcategories of context variables include school characteristics and school processes.

The evaluation of the different dimensions of the teaching learning process in the TL set-up is important in order to ascertain whether or not the core curriculum has been implemented as planned and has achieved its goals and objectives. Various evaluation studies have been outlined by the University to be conducted, as indicated below.

Areas	Indicators
Context:	Support needed by faculty and students and extended by school Needs assessment on curriculum development and implementation Quality and use of modules Reliability and validity of departmental tests in GE/major courses School policies affecting TL implementation
Inputs:	Teacher characteristics, such as gender, status, department, terms taught TL subjects, framework acceptance, training/seminar workshops attended Student characteristics {i.e., critical thinking and problem solving skills (pretest)}
Classroom Processes:	<i>Teacher Behavior</i> Faculty Evaluation Ratings Every Term Comparison of the evaluation ratings obtained by the TL teachers based on class type, gender, and employment Effects of the teachers' training, teaching experience, and attitude on their teaching effectiveness in TL classes <i>Student Learning Process</i>
Outputs:	Student Achievement (Critical thinking and problem solving skills after 1,2 3 years and upon graduation) Program Effectiveness (program benefits to school, students and faculty)

Three dimensions of the teaching learning process in the TL set-up have so far been subjected to evaluation: (1) the input, specifically the teacher and student characteristics; (2) the classroom processes, particularly the faculty effectiveness and student learning processes, and (3) the context, particularly the implementation of the program. For input evaluation, the following studies have been conducted:

- (1) On teacher characteristics:
 - 1.1 Profile of faculty handling TL classes
- (2) On students characteristics:
 - 2.1 Development and validation of DLSU Thinking Skills Test
 - 2.2 Profiling of all freshmen of their critical thinking and problem solving skills
 - 2.3 Profiling of COS freshmen scientific inquiry skills.

For process evaluation, the following studies have been undertaken:

- (1) On teacher and student behavior:
 - 1.1 Development and validation of faculty evaluation forms for TL classes
 - 1.2 Evaluation of rubrics to assess the teaching-learning process in the TL classes
 - 1.3 Evaluation of all TL teachers' effectiveness in handling TL classes
 - 1.4 Evaluation of student learning process in the TL set-up.

Lastly, for the context evaluation, the implementation evaluation of the TL program was undertaken.

This paper will present five of the initiatives of the University in the measurement of the teaching-learning process in TL set-up. It will present the research findings of the five separate research studies conducted on the:

- (1) development and validation of evaluation forms for faculty handling TL classes
- (2) implementation evaluation of the TL program
- (3) profile of TL faculty members
- (4) development of rubrics to assess the teaching-learning process in TL classes
- (5) development and validation of a students' thinking skills test

Development of the Faculty Evaluation Forms for TL Classes

Since the beginning of the conceptualization of the new framework, it was known that the scope of this new paradigm is very much different from the traditional framework of transmission, from which the current faculty evaluation was based. Thus, it was important that a new faculty evaluation form be developed and validated.

Thus, during the Third Term of AY 2005-2006, the preparation for the development of a faculty evaluation form that will be used for classes using the said framework was started. There were six phases in the development of the evaluation form: (1) Conceptualization, (2) development of items, (3) try-out of the prototype instrument, (4) item analysis of the proposed evaluation form, (5) establishing the reliability and validity of the evaluation forms, and (6) finalization of the evaluation forms.

The four areas of learning in the Transformative Learning framework, namely student learning process, teacher's work, students' learning environment, and assessment of student learning were adopted to be the factors or areas of assessment. A total of 46 items were formulated for the four areas. These items were then submitted for content validation and administered to a total of 235

students enrolled in eleven (11) classes. The data were then subjected to two item analysis procedures: item-total correlation and coefficient alpha deleted. The item-total correlation and the alpha coefficients for each item were then ranked from highest to lowest. The first seven items for each area were included in the proposed evaluation form. The proposed form was then submitted for review to the Committee Chair, VP-AR as well as to the chairpersons of Biology, Chemistry and Physics (for laboratory classes), and Physical Education (for PE classes). Four types of evaluation forms were formulated: (1) for lecture classes, with two more items added to the pre-final form; (2) for PE classes, with four more items added; (3) for laboratory classes, with five more items added, and (4) for Filipino classes, with items translated to Filipino.

The four sets of evaluation forms were then administered to the students. Data from these samples were used for item analysis, and for examining the reliability and validity of the proposed evaluation forms. Results indicate that the items within the area related well with each other and with the composite score. The average inter-item correlation coefficients range from .42 (for Filipino Form) to .57 (for Lecture Form). High item-total correlation coefficients (.49 to .80) were also computed for each of the four forms. Similarly, high Cronbach Alpha coefficients were computed (.95 to .97). Thus, it was decided that no items will be dropped from the final form since it will not result to a dramatic increase in the Cronbach Alpha coefficients. The reliability of the different forms will not at all be affected.

The reliability of the evaluation forms was tested using the split-half technique. Two types of split-half reliability methods were employed: (1) Dividing the evaluation forms in some random manner into two halves, through STATISTICA, and (2) odd-even reliability. The STATISTICA produced two types of split-half reliability coefficients: the split-half and the Guttman split-half reliability. Results indicate very high correlation coefficients between the two halves of the evaluation form. The correlation coefficients computed exceeded the .90 mark for the four types. Per area, the odd-even reliability coefficients ranged from .78 to .98 for the four types. This indicates that the evaluation forms are highly reliable.

Three types of validation procedures were employed to ensure that the evaluation forms measure what they are supposed to measure: (1) content validity, (2) construct validity, and 3) concurrent validity. The content validity of the four forms was indicated by the method by which the items were selected. In the development of the four forms, the items were based on the four learning areas identified by the Lasallian Core Curriculum Committee. The items also underwent content validation to ascertain suitability to the learning areas. Results of the factor analysis yielded the same hypothesized number of factors for the different forms: four factors for the lecture, Filipino and PE forms, and five factors for laboratory form. The items also loaded highly (if not the highest) in their original factors. Thus, the names of the factors and the item memberships were retained.

For the concurrent validity, the new evaluation form (for lecture) was correlated with the existing indigenous faculty evaluation to determine its equivalence with this standardized form. The evaluation ratings received by the faculty in their two classes (regular and TL classes) were subjected to correlation

analysis. The correlation coefficient obtained was .38, which was only moderate. This indicates that while the two forms should be measuring the same general area of behavior (i.e., teaching effectiveness), they however, contain different dimensions of teaching effectiveness. This is a positive result since it proves that the new evaluation form is not equivalent to the existing evaluation – the old one measuring teaching effectiveness in a transmission system of education and the new form, measuring teaching effectiveness in a transformative learning set-up.

To further determine the non-equivalence of the two forms, the mean ratings of the 58 faculty members in the two forms were compared using the t-test for dependent samples. A t-value of 3.87 was obtained. This implies that there is a significant difference between the evaluation ratings obtained by the two faculty members in the two evaluation forms. The forms are, therefore, not equivalent and are measuring different dimensions of teaching effectiveness.

Comments and suggestions about the evaluation forms were solicited from the students. These comments and suggestions were taken into consideration for the improvement of the evaluation forms. Based on the students' feedback and the results of the item-analysis, reliability and validity of the evaluation instruments, it was concluded that the items proved to be good items and are suitable to be included in the final forms. All items in the proposed evaluation forms were retained and were made part of the final forms to be used for classes using the Transformative Learning Framework.

Implementation Evaluation of the Transformative Learning

Since the TL Framework is a new program, there is a need to conduct an implementation evaluation of the program to monitor program activities in order to identify problems in program implementation. According to Love (2004), implementation is an integral part of the program cycle. It refers to all of the activities that focus on the actual operation of a program once it moves from the drawing board and into action. The evaluation of implementation is important since it provides feedback about what worked and what did not produce the intended outcomes.

An implementation evaluation study of the TL program was conducted during the Second Term of SY 2007-2008. The study adopted a management-oriented evaluation using the framework of program theory evaluation, as described by Chen (2005). It included as respondents students, faculty members, and chairpersons. The following are the salient findings of the implementation evaluation of the TL program.

The student respondents generally rated the faculty members' teaching performance as "good" while the faculty and chair respondents gave them "very good" evaluation rating. Specifically, the TL teachers received "good" to "very good" evaluation ratings in the following areas: (1) Use of teaching skills and strategies appropriate for TL classes, (2) creation of conducive learning environment, (3) employment appropriate assessment techniques, and (4) processing of students' learning.

The faculty and chair respondents likewise generally rated the modules used for TL classes as “very satisfactory”. They also indicated that these modules were used to a great extent.

Too much or heavy workload is the problem most students encountered in the TL classes. For the faculty and chairs, on the other hand, lack of time is the major obstacle in handling TL classes.

The student respondents were given opportunities to learn on their own and to become independent and self-reliant. These learner traits are among the strengths of the TL classes. According to the faculty respondents, the TL classes enabled the students to discover and appreciate their own learning (i.e., active, participative, independent learning) and to develop their critical and analytical skills. The chairpersons cited promotion of students’ independent learning as the strength of the TL classes. However, both groups mentioned time constraint as the biggest drawback of the set-up.

The student respondents rated the TL Program “good” while both the faculty and chair respondents rated it “very good”.

The Profile of Faculty Members Handling Classes Using the Transformative Learning (TL) Framework

This is a descriptive study of the faculty members who have handled TL classes. Specifically, it sought to: (1) describe the TL faculty members’ demographic profile (i.e., gender, and employment status), exposure to and training in TL classes (i.e., number of terms of teaching TL classes and attendance in seminar-workshops on TL), teaching performance in TL classes, and attitude towards program; (2) compare the evaluation ratings obtained by the TL teachers based on class type, gender, and employment status; (3) assess the effects of the teachers’ training, teaching experience, and attitude in their teaching effectiveness; and (4) determine the students’ feedback regarding the TL program and the faculty members.

The study included all the 349 faculty members who had handled one or more TL classes from the Second Term of SY 2006-1007 to the Second Term of SY 2007-2008. The data came from the evaluation database of the Institutional Testing and Evaluation Office. Descriptive statistics, t-tests, and correlation were employed for data analysis. Content analysis was done for open-ended data.

Results indicated that majority of the faculty members handling TL classes are generally from the College of Liberal Arts, teach part-time, have taught TL subjects only for a term, have attended at least one seminar-workshop, have received “satisfactory” to “very satisfactory” evaluation ratings, and are very satisfied with the TL program as a whole. There was an equal number of male and female teachers who had handled TL classes. Comparisons of the evaluation ratings of the TL faculty members also indicated that teachers obtained significantly higher evaluation ratings in their regular classes than in their TL classes; male tutors obtained higher ratings than their female counterpart, although, the difference was not significant. There was no significant difference in the overall evaluation ratings between part-timers and full-timers; both were generally rated “very satisfactory” by the students. Attendance in seminar-workshops and attitude towards the TL

program, on the whole, did not make a difference in the evaluation ratings of the faculty members. Faculty members who have participated in training programs and those who have not, both obtained satisfactory to very satisfactory ratings. Similarly, faculty members obtained “satisfactory” to “very satisfactory” evaluation ratings, irrespective of how they viewed the TL program (i.e., whether excellent, very good, good, or needs improvement).

However, teaching experience in TL classes seems to have some bearing on the teachers’ teaching effectiveness, since the evaluation ratings were observed to have increased with the increase in the number of terms teaching TL classes. The students enrolled in TL classes were satisfied with the TL program and their faculty members, as shown by more students citing more program strengths than weaknesses.

Development of Rubrics to Assess the Teaching-Learning Process in Transformative Learning (TL) Classes

One of the unique features of the TL model is the use of authentic assessment tools, specifically performance assessment approaches such as rubrics (DLSU Committee on Lasallian Pedagogical Framework, 2004). Rubrics are scoring guides that evaluate the quality of work or output against a set of criteria. They are a set of ordered categories to which a given piece of work can be compared. They specify the qualities or processes that must be exhibited in order for a performance to be assigned a particular evaluative rating (Mamouth University, 2005).

One of the criticisms of using absolute standards in assessing student performance is that they encourage rote instead of meaningful learning. According to the DLSU GEC Committee (DLSU Committee on Lasallian Pedagogical Framework, 2004), in a system of transmission, tests are expected since instruction consists mainly of providing factual information and prescribing procedures. But within the perspective of transformative learning, results from such tests yield little information about the kind of cognitive growth that has taken place in students, the changes that have occurred in their conceptual representations, or their ability to solve problems in the field. One solution that has been proposed to solve the inadequacy of traditional testing formats is to apply the concept of authentic assessment to classroom testing.

While rubrics are basically employed to assess student performance and output, they can also be applied to various components of the teaching and learning processes. Particularly, in the transformative learning set-up, rubrics can be used to assess the effectiveness of the faculty; the student outcomes or learning process; the students’ performance in classroom activities, such as class participation or oral presentations, the students’ outputs (such as projects, portfolio and term papers) the quality of instructional materials used, particularly the modules; the learning environment set-up; the relevance of the assessment tools; and a lot more.

Because of the applicability and relevance of rubrics in the transformative learning classes, rubrics are deemed needing to be developed that can be used for assessing the teaching and learning processes in the classroom.

The action plan primarily aimed to present a comprehensive blueprint for the development and application of performance assessments in classes using the transformative learning paradigm. Furthermore, it presented samples of rubrics for assessing the teaching and learning processes in the classroom, particularly:

- (1) the teaching performance of faculty members in the different learning areas of the TL paradigm and in using inquiry-based teaching in the classroom,
- (2) the students' performance, particularly in terms of applying inquiry-based learning in class and in terms of collaboration with other students,
- (3) students' products that represent major instructional aspects of the course, such as portfolio, and
- (4) students' learning outcomes, particularly in terms of their learning process.

Development and Validation of the DLSU Thinking Skills Test

During the Second Term of School Year 2007-2008, the Institutional Testing and Evaluation Office (ITEO) started the work on the development and validation of an instrument that will measure students' thinking and inquiry skills. Based on the models/frameworks and tests developed in other universities abroad, two measures/areas of students' thinking skills were identified: (1) Critical thinking and (2) problem solving skills. There are seven kinds of questions adopted for assessing of critical thinking skills: (1) Summarizing the main conclusion, (2) drawing a conclusion, (3) identifying an assumption, (4) assessing the impact of additional evidence, (5) detecting reasoning errors, (6) matching arguments, and (7) applying principles.

On the other hand, problem solving skill involves reasoning using numerical and spatial skills. There are three kinds of questions for the assessing problem solving skills: (1) selecting relevant data/information, (2) finding procedures, and (3) identifying similarities.

On November 2007, recommended faculty members from the Departments of Mathematics, English and Applied Linguistics, Physics, and Philosophy were tapped as items writers for the two areas. The items submitted were pre-tested on students and the data gathered were subjected to item analyses to determine the discrimination and difficulty indices of each item. Items that met the criteria and were considered as good items were included in the final form of the test. A total of 50 items comprised the final form of the DLSU Thinking Skills Test.

During the first few weeks of the First Term of SY 2008-2009, ITEO administered the DLSU Thinking Skills to all freshmen to come up with baseline data on the incoming students' critical thinking skills and inquiry. Two other standardized tests developed by universities abroad were likewise administered to the students to measure their thinking and scientific inquiry skills.

The same process will be done to the same cohort during the Third Term, starting March 2009 until the end of the students' third or fourth year (i.e., upon

graduation). The data that will be gathered will be used to determine significant changes, if any, in the students' level of thinking skills as a result of their exposure to the transformative learning set-up.

References

- Boyd, R. D., & Myers, J. G. (1988). Transformative education. *International Journal of Lifelong Education*, 7(4), 261-284.
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks: SAGE Publications.
- DLSU Committee on Lasallian Pedagogical Framework. (2004). *Towards a Lasallian pedagogical framework of transformative learning*. Retrieved from De La Salle University Web site: http://www.dlsu.edu.ph/offices/iaa/downloads/Lasallian_Pedagogical_Framework.doc
- Huitt, W. (2003). A transactional model of the teaching/learning process. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved May 15, 2008, from <http://chiron.valdosta.edu/whuitt/materials/tchlrnmd.html>
- Imel, S. (1998). *Transformative learning in adulthood*. Columbus: ERIC Clearinghouse on Adult Career and Vocational Education . (ERIC No. ED423426)
- Love, A. (2004). Implementation evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.
- Mamouth University. (2005). *Rubrics*. Retrieved March 8, 2007 from <http://its.monmouth.edu/FacultyResourceCenter/rubrics.htm>
- Mezirow, J. (1978). Perspective transformation. *Adult Education*, 28, 100-110.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco, CA: Jossey-Bass.
- O'Sullivan, E. (2003). Bringing a perspective of transformative learning to globalized consumption. *International Journal of Consumer Studies*, 27 (4), 326-330.
- Taylor, E. W. (1998). *The theory and practice of transformative learning: A critical review*. Information Series No. 374. Columbus: ERIC Clearinghouse On Adult, Career, And Vocational Education, Center On Education and Training for Employment. (ERIC No. 423422)



Organizing and reporting assessment results

Jennifer Ann L. Lajom

De La Salle University, Manila, Philippines

Abstract This article discusses the scope of organizing and reporting assessment results, which include the processes of collecting, managing, analysing, and reporting assessment information. Specific sources of data, as well as strategies in obtaining and interpreting the information are listed. A synthesis on how the entire process of organizing and reporting assessment result tied with teacher role and responsibility is provided.

Keywords: *Assessment, organizing and reporting assessment*

Introduction and Basic Considerations

What does it take for the organization and report writing of assessment information to be precise and efficient? Does precision result from the type of assessment tests? Is efficiency brought about by the number or even types of strategies employed during assessment? Indeed these components of assessment contribute to the effectiveness of the assessment process. However, it is imperative that the types of measures or strategies used are all encompassed by the underlying objectives and expected outcomes as agreed upon prior to the execution of assessment. As such, the preparation behind an assessment program, along with the planning as to which measures and strategies are to be employed can fully contribute to the precision and efficiency of assessment and evaluation.

Planning is at the core of organizing and reporting assessment results. As earlier mentioned, effective assessment practice is born when assessment data reflect the learning objectives and learning outcomes of a lesson. By means of conceptualizing an assessment plan, teachers are able to narrow down the area to assess and maintain a clearer perspective of what students should learn. For instance, Helman (2006) explains that classroom literacy practices should be included in the assessment plan. By means of monitoring student progress and identifying who needs extra guidance, one is able to make the most of assessment data. Likewise, it is also ideal that the assessment data will be examined by all

stakeholders that hold various perspectives in the realm of learning. Reporting assessment data, thus, should be treated as important as test construction, where what needs to be reported should be determined beforehand.

The aforementioned rationale cites specific reasons regarding the significance of reporting assessment results. Navarrete and Gutske (1996) presented three main issues in reporting assessment data, which include the purpose of assessment, strengths and limitations of assessment, and methods of presenting results. A clear purpose facilitates clarity in communicating assessment results, which in turn, satisfies the informational need of the student. Reporting thus, allows for the availability of feedback about student performance for both student and parents. A clear purpose also promotes an understanding of its implications which can be reflected in the assessment design, data collection, instrumentation and analysis. Such understanding can serve as basis whether instruction is meeting the established literacy goals or there is a need to modify such goals. Over time, the degree of their progress may result to improved teaching and learning, and perhaps to funding, or continued funding, for improvement of instruction.

In order to achieve the proper assessment measures that appropriately reflects the objectives and outcomes of a learning material, it is imperative that the teacher has a clear idea on what type of data are necessary to assess the learning process, as well as knowing how to acquire such data. As the learning experience vary due to differences in context, the sources of information can be varied as well. Given that there are different sources of information, there are also various ways of obtaining them. The subsequent sections discuss collection of assessment data, with descriptions of the kinds of data that can be used for assessment as well as the strategies on how to collect them.

Assessment Data Collection

Organizing and reporting assessment begins with collection and storage of data. Since there is a vast range of assessment data that can be collected, stored, and analyzed, it is important to be selective in order to prevent being overwhelmed with data. It also allows one to be focused on what the target of assessment really is, which can have precise implication on specific learning targets. An advisable practice with regards to obtaining assessment data is to establish a range of mechanisms for data collection and to interpret and analyze as soon as possible. It is preferred that the gap between data collection and analysis will not be prolonged, but rather, is acted immediately as soon as data is available. The rationale behind this lies in the way the teacher can treat the data in an authentic manner, where it is still contextualized in the scenario where the data was acquired.

There are two ways in which information can be collected, and that is either it is obtained formally or informally. Formal data collection includes the use of research instruments, clinical techniques, and standardized tests. Informal data collection, on the other hand, pertains to observation, collection of classroom products or information sourced from parents or other individuals. In collecting information for assessment, it is important to acknowledge the importance of the

context where learning occurred. McAfee and Leong (2007) coined this as “multiple windows” from which data are obtained from. Multiple windows pertain to the varied measures of data that can provide a more valid estimate about the student’s achievement. In this sense, not only do more windows depict relatively complete information about the student, but it also increases the reliability and representativeness of the student’s learning accomplishments and capabilities. Multiple windows also liberate teachers from the constraint of one approach, which in turn, decreases the possibilities of errors. Three aspects of classroom management, namely, (a) source of information, (b) method of obtaining information and (c) context, setting or situation of the appraisal, can be varied to provide multiple windows or measures.

Sources of Assessment Data

Source of information can be primarily acquired through the student who will be assessed. As the primary source of data, talking and listening with the student during class can be a convenient way to acquire data. This shows how students themselves volunteer information about their own learning and difficulties in the natural classroom setting. In a more formal set up of class discussion, student participation, perhaps in response to questions posed by the teacher, is an indispensable information source.

Parents, specialists, other adults, peers or classmates, and records about the student can also serve as sources of assessment data. Parents and other adults can provide yet another varied perspective about the student. Being the primary caregivers of students, parents know their children more than anyone else and generally have a closer experience of seeing how their children behave when they are not in school. In this sense, parents are more knowledgeable about the situations at home that might affect the performance of their children in school. Differences in home and school culture that can play a role in classroom dynamics of the student can be provided by parents and these inputs are useful in a school that caters to diversified student learning environments.

Other teachers who have handled the student in different contexts (e.g., school subject) can also provide how the child becomes ‘different’ based on the activities they perform in their respective classes. A student may behave in various ways in different classes and the perspective various teachers can provide can help in depicting a holistic picture of the student. Finally, records that depict attendance, intake, health and school history, progress reports, report cards, test results or portfolios can be rich sources of assessment data and can be used prior and after assessment of the student. As time, money and energy were utilized to arrive at these documentations, it is but right to put such materials to good use.

Peers and classmates can also serve as informant on how a particular student progresses in the classroom. Apart from the student to be assessed, the classmates also experience the same standards of instruction which can serve as comparison with regards to pace of learning. Classroom activities that require students to work together as a group can also serve as gauge as to the performance of each member with regards to a particular task by means of feedback or peer

evaluation. By means of having various sources on how learning is achieved can lead to a broader perspective with regards to the learning processes of a student.

Method of obtaining information can be by means of systematic observation, eliciting responses from students and, eliciting information from parents and other adults. In systematic observation, attention is directed towards a particular target such as the pattern of the child's behaviour, a situation or problem that needs resolution, or the child's progress towards an identified goal. When executing systematic observation, one should do so unobtrusively, ideally in a place with a full view of the target situation and behaviour but without disrupting the natural progression of the learning environment. During observation, verbal and non-verbal behaviour should be noted. For instance, the actual words uttered, its intonation, enunciation, and pronunciation, as well as body stance, movement, motor responses, gesticulations, and facial expressions should be recognized and recorded.

Strategies in Obtaining Assessment Data

As previously mentioned, the students themselves are ideal sources of their own learning status, which is why it is imperative that these insights are elicited in the proper manner. There are various ways in attaining this such as initiating instructional conversations or dialogues during class, as well as incorporating elicitation of feedback in daily classroom activities. This becomes evident during class discussions or structured learning exercises, where students are required to participate and produce a certain degree of quality output.

Alternative assessments, such as performance assessment, dynamic assessment, interviews, conferences, and discussions, are also other options by which teachers can secure student responses. Performance assessments allow pupils to demonstrate what they know and can do in real life situations. This also reflects a level of authentic assessment, and although indirectly, provides insights on the pedagogy of the instructor. On the other hand, dynamic assessments probe skills that are on the verge of emergence. The potentiality of these skills are recognized and acted up, that is, the potential is tapped and further honed for improvement. Indicators such as hints, prompts, cues or questions that aid in identifying the strategies and processes that enable the students to learn more effectively in the future are also noted. As the teacher and student continue to interact with each other in class, the skills are further developed until it reaches maturity and the student can perform optimally.

Interviews are also efficient means to get student information. The content of the interview should comprise of well thought of questions that reflect what teachers expect students to do relative to the objectives of the assessment. By means of this method, it allows the teacher to probe deeper into the experiences of the student during the learning process. It is important to note however, that there is equal importance placed in both initial and follow up questions since it allows the teacher to create a holistic portrait of the student's learning experience.

Another way to get information is through holding student-teacher conferences. The process entails the two parties to sit together and discuss what the student has done and what the teacher thinks of it. However, it is not just the teacher who gets to provide inputs during conferences, but the student is also given the opportunity to explain his or her work relative to a particular class standard. Thus, conference discourse results to mutual exchange of insights and suggestions.

Periodic sampling of classroom work products aid in assessment as this shows what the student has learned over a period of time. Within the collection of work products, breakthrough samples that show a particular growth and development toward a certain learning standard also provide valuable information about the student's learning process. Finally, by means of holding informal conversations, conferences, and administering questionnaires to parents also achieve the objective of acquiring ample information about the student's learning progress.

Given the various sources and strategies that reflect multiple windows approach to assessment, it is noteworthy to mention several guidelines in choosing the appropriate assessment window. First is to identify the behaviour to be assessed and find out the contexts that would best demonstrate the behaviour. Second is to consider whether the source, method, and context will produce an authentic assessment. Finally, the teacher should maximize the chances of seeing the behaviour within a context. This is important because certain contexts restrict and limit the occurrence of certain types of behaviour, while certain activities can be more engaging and elicit better, more reliable sources of behaviour. Recognizing the extent to which a particular context can provide the most useful data is most useful in the assessment process.

Management of Assessment Data

While assessment information will more or less depict the status of a learning individual, properly handling such data is also a relevant issue. Knowing the source and the process of how to get information is one aspect of collecting assessment data. However, having the proper collection tools will enable the teacher to efficiently proceed with the assessment process.

Some assessment tools include paper and pencil collection of simple numerical and textual data. These can be incorporated through the use of devices such as word processors and spreadsheets with appropriate an analysis-ready software. Aside from the traditional tools, it is advisable to make use of a range of technologies that can help collect, store and manage assessment data.

Among these technologies include digital cameras, audio recorders and video cameras that are relatively affordable and easy to use with minimal skill requirements needed to operate. One can easily access and manage the digital data by means of, for instance, the 'drag and drop' method for transferring files to the computer and specific software are available that can help manage data files. Likewise, today's mobile phones have features that make its functions highly flexible, such as recording audio and video. The flexibility of the mobile phone, matched with a regulated policy on the extent of mobile phone utility in class, can

pave way to the involvement of the students themselves in gathering assessment data using the mobile phone as means for documentation.

After collection, there are strategies that can be utilized in order to facilitate compilation, storage and presentation of assessment data. Portfolios, individual and group profiles and technology-aided storage are some examples of managing information used for assessment. Using portfolios can determine the student's status and progress within a learning context. It can also be a source of instruction and provides information that is useful for reporting assessment results. Portfolios also allow for the preliminary identification of students who need special attention.

Individual profiles document a student's capabilities and behaviour patterns, including the extremes and uniqueness, and stylistic approaches to learning that are not easily seen in just one assessment. Group profiles, on the other hand, show class performance on one or more items. They focus on a range of class behaviour and identify subgroups or clusters of students with similar strengths and needs. The behaviours can be presented qualitatively, such as in terms of the degree of detail, and quantitatively, such as the frequency of the behaviour's occurrence. Given the profiles, varied instructions can be planned in response to the student's needs.

Technological advancement has its fair share in shaping the strategies for storing and managing assessment data. As mentioned earlier, word processors and spreadsheets, as well as software that can create data bases of information can help a teacher in assessment data management. Other software that come in generic computer programs such as Media Player, Movie Player, Picasa or iTunes can also serve as storage and presentation tools for assessment data. In using these devices, however, it is advisable to use specific file names for each data file for easy reference, as compared to the automatic abstract or numerical file name generated by the digital device. Online resources such as Google Docs also allow one to solicit responses for data by means of online instrument administration which will automatically be encoded in a spreadsheet file.

The emergence of technological utility for recording assessment data shows that while it is standard to produce a written report as an assessment output, there are options as to how to most effectively convey the results and implications of an assessment. Shepard (1980) cites how other media can be used such as film strips or slide show presentations as this can be effective in catching and holding the attention of an audience, perhaps in longer duration, as opposed to words and graphs in written forms of reports.

Given both the traditional and contemporary approaches to reporting assessment results, it is still highly recommended for reports to be delivered personally. Face to face contact ensures that reports will be looked at and provides the opportunity to address questions that cannot be catered by written documents.

Interpreting Assessment Data

With the assessment data obtained and secured, meanings can now be extracted and applied in the contexts of instruction. Apart from student-driven information, assessment data should also inform the teacher of his or her own teaching. Thus, effective recording can provide a clear insight on the content and

process of the student's learning as well as insight on teacher's pedagogy and effectiveness.

As previously mentioned, there is no single source of information that tells everything about the student, and that the context where learning occurred affects student behaviour in various ways. In analyzing assessment data, therefore, the type of process usually reflects the objectives of the assessment in order to serve as a guide in arriving at explanations and implications. For instance, one strategy in the analytic process is to look for general themes or common issues among students. Perhaps, these commonalities come in the form of concerns, difficulties or small successes in class. The responses, derived from textual or visual data, can also be categorized to a particular teaching task or activity.

It is also efficient to conduct specific analysis of data in response to particular issues at hand. Having a research question in mind can be utilized as a lens to look into the data in a particular focus, where the teacher can be selective and prioritize certain types of responses or information over others. By means of such selectivity, the teacher is able to explore key moments within the lesson in more detail, where students are really engaged and clearly learning effectively.

With regard to this, Fautley and Savage (2008) espouse the use of thick description, a strategy that is used as a guide for data interpretation that reminds the teacher that data is heavily contextualized. The scope of student learning context encompasses within the classroom, the instruction that is being undertaken and the individual lives of the students. Thus, when the teacher focuses interpretation and analysis on individual students, it is helpful to keep in mind that the classroom is a socialized environment and instruction is mediated by classroom environment, pedagogy, other students and previous learning experiences. It is imperative that these variables are considered when analyzing and interpreting assessment data.

Writing Assessment Reports

Apart from the role of context in learning, it is also important to consider the audience or the recipient of the report such as parents or school managers. Individuals who require a write up of assessment data would need different types of information. For instance, a parent may be interested on how the child is faring in school and what other strategies can help the child excel, while a school administration may prefer data that would facilitate curriculum improvement. Clearly, the report audience will have implications to content and eventually, the format of the report.

When it comes to writing assessment reports, however, it is generally acceptable to cover points regarding the consequence of one's teaching and how it can be improved, as well as what students have learnt and the evidence of their learning. For example, a teacher can create simple descriptions of individual student's learning and should be able to justify these by citing accomplished classroom activities. The teacher is expected to illustrate the facts of what a student has achieved in a particular time period with descriptions with their work that is either in progress or finished products.

Comparative and interpretative information should also be a part of the assessment report. Comparisons are important if different audiences are to derive meaning from assessment results, while interpretative information about the implications of the assessment results that will play a role in instruction. Specifically, assessment data derived from interpretative information will aid in the utility of the information in arriving at conclusions that will have an impact in maintaining or improving instructions for learning.

Synthesis

Conclusions derived from assessment data are contextualized in one's way of teaching and in most cases, are tentative and not easily generalizable to other instructor's way of teaching or managing a class. This lack of generalizability should not be mistaken as a weakness of the assessment process, but instead should be viewed as strength. For one thing, it reinforces one's professional role and identity as a teacher in relation to the instruction done and the learning that the students have gained. As an individual in a teaching profession, a key function of a teacher is to obtain, develop and provide a rich account of student learning based on assessment data, where the experience is expressed in a rich manner and not reduced to mere numbers or statistics.

References

- Fautley, M., & Savage, J. (2008). *Assessment for learning and teaching in secondary schools*. Great Britain: Learning Matters.
- Helman, L. A. (2006). How assessments promote learning from diverse students. *English Leadership Quarterly, 28*, 10-15.
- McAfee, O., & Leong, D. J. (2007). *Assessing and guiding young children's development and learning*. Boston, Mass.: Pearson.
- Navarrete, C., & Gutske, C. (1996). *A guide to performance assessment for linguistically diverse students*. Albuquerque, New Mexico.
- Shepard, L. (1980). Reporting the results of statewide assessment. *Studies in Educational Evaluation, 6*, 119-125.