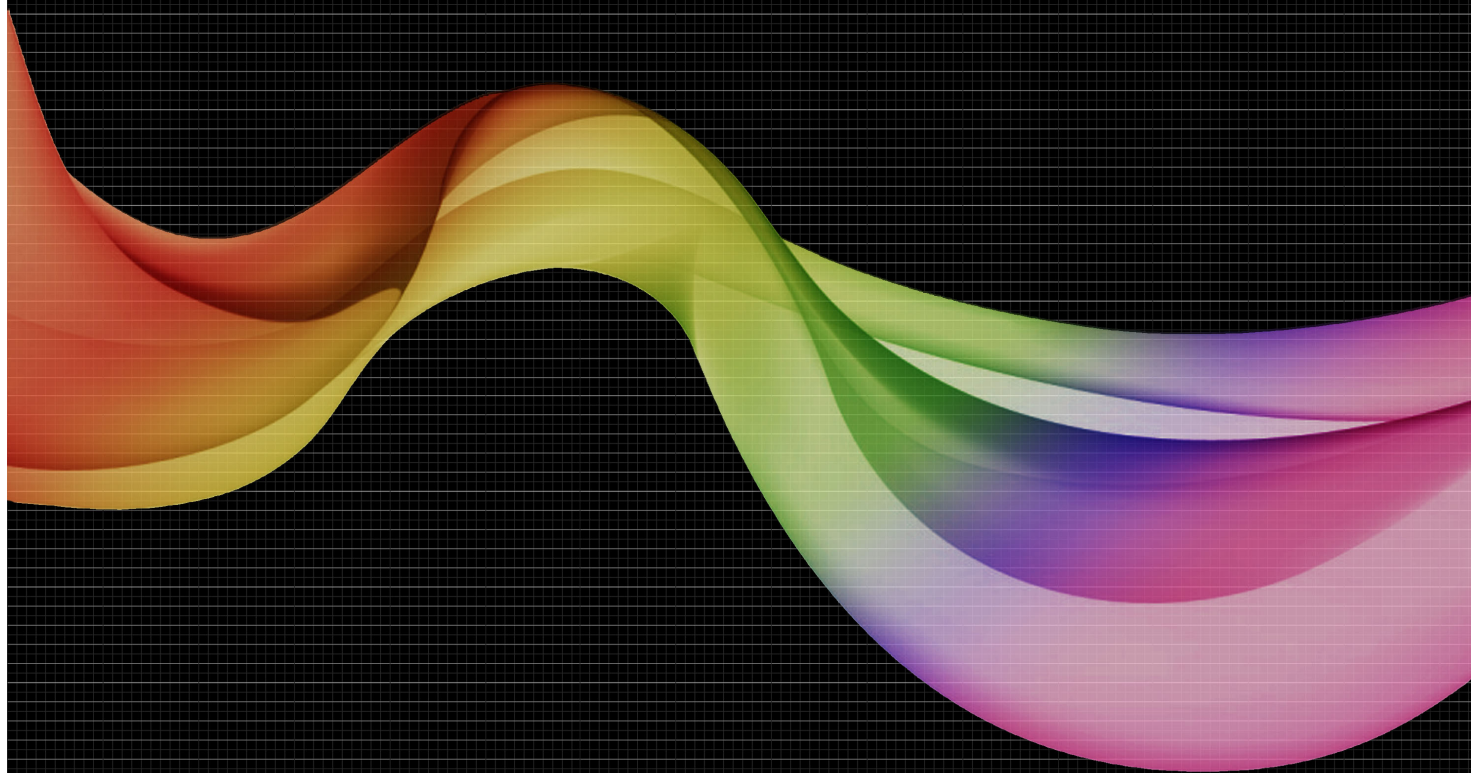


The Philippine Educational Measurement
and Evaluation Association (PEMEA)

ISSN 2094 - 5876



THE EDUCATIONAL MEASUREMENT AND EVALUATION REVIEW

Volume 4 July 2013

The Educational Measurement and Evaluation Review (EMEReview) is the official publication of the Philippine Educational Measurement and Evaluation Association (PEMEA). It is international, refereed, and abstracted/indexed. The EMEReview publishes scholarly reports about contemporary theories and practices 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 and policies on assessment, contemporary approaches in educational and psychological measurement, and other studies with direct implication to assessment in education, social science, and related fields. Emereview is indexed/abstracted in the Asian Education Index, Social Science Research Network, Google Scholar, Open J-Gate, and NewJour.

Copyright © 2013 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 open-access and users may read, download, copy, distribute, print, search, or link to the full texts, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.

The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

The articles in the EMEReview are open access at
<https://sites.google.com/site/emereview/>



Publication Division of PEMEA
Philippine Educational Measurement and Evaluation Association

◆ Editorial Board

Executive Editor: Dr. Carlo Magno, De La Salle University

Associate Editors: Mr. Ryan Cayubit, University of Sto. Tomas
Dr. Marcos Lopez, Centro Escolar University, Malolos Bulacan

Dr. Richard Gonzales, Development Strategists International
Consulting/ University of Santo Tomas, Philippines

◆ Editorial Advisory Board

Dr. Rose Marie Salazar-Clemeña, De La Salle University, Manila

Dr. John Hattie, University of Melbourne, Australia

Dr. Jack Holbrook, University of Tartu, Estonia

Dr. Anders Jonsson, Malmo University, Sweden

Dr. Timothy Teo, The University of Auckland, New Zealand

Prof. Tom Oakland, University of Florida, USA

Dr. Jimmy dela Torre, Rutgers University, USA

Dr. Jose Pedrajita, University of the Philippines, Diliman

Dr. Shu-ren Chang, Department of Testing Services, American Dental Association,
USA

Dr. Karma El Hassan, Office of Institutional Research and Testing, American
University of Beirut, Lebanon

Dr. Milagros Ibe, University of the Philippines, Diliman

Prof. Alexa Abrenica, De La Salle University, Manila

Dr. Marie Ann Vargas, University of Sto. Tomas, Manila

Dr. Leonore Decencente, Center for Educational Measurement, Inc., Makati

◆ Editorial Staff for the Volume

Stephanie Dy

Martin Cyrus Romero

Educational Measurement and Evaluation Review

ISSN 2094-5876 <https://sites.google.com/site/emereview/> ♦ July 2013, Vol. 4

Editorial Note.....	1
Ryan Cayubit	
Towards the Development of the Case Notes Assessment Scale	
<i>Susana A. Estanislao</i>	2
Assessment of Family Functioning among Perpetrators of Physical Abuse	
<i>Aime T. Guarino and Glenda Mae V. Santos</i>	12
Reinforcement Sensitivity as Predictor of Test Anxiety	
<i>Rene M. Nob</i>	26
A Path Analysis of Inspiration, Purpose in Life, Gratitude and Subjective Well-Being Among Filipino Workforce	
<i>Ricardo C. Lo and Gabriel Ngombedua Makundu</i>	37
Parental Involvement in Children's Assessment in Kindergarten	
<i>Joyce Ferro Orillosa and Carlo Magno</i>	47
Development and Standardization of the Anger Expression Type Indicator Test (AETIT)	
<i>Ryan Francis O. Cayubit</i>	66
The use of Western Standardized Psychological Tests in Non-Western Contexts	
<i>Dylan Marshall</i>	78
Determining the Construct Validity of a Critical Thinking Test	
<i>Marcos Y. Lopez</i>	87
Scale on Attitude Toward Mathematics (SATM)	
<i>Ivee Guce and Joy Talens</i>	100



Editor's Note

Ryan Francis O. Cayubit
University of Santo Tomas, Philippines

This fourth volume of the Educational Measurement and Evaluation Review is a testament to the commitment of PEMEA to help in the advancement of research in the Philippines. Just like the previous volumes, we have provided an avenue for the scholarly work of researchers in the Philippines and beyond. It is also in this fourth volume that our partnership with Philippine e-Journals became a reality and through this partnership. We look forward to a wider range of readers for all the published articles and an increase in the number of citations.

The current volume has seven articles that deal with some aspect of measurement and assessment. The works of Estanislao and Cayubit are scale development projects whose outputs would aid in the advancement of the field of testing and measurement in the Philippines. The articles of Guarino, Santos, Orillosa, and Magno deals with the assessment of constructs that are related to families. Nob's article on test anxiety has practical implications to students while that of Lo and Makundu deals with the Filipino work force. This volume ends with a short report on the use of western made tests in non-western culture, a topic on test adaptation that has been subject of many debates throughout the years.



Towards the Development of the Case Notes Assessment Scale

Susana A. Estanislao

De La Salle University - Manila, Philippines

Abstract This study initially developed an instrument that measures the counseling professionals' current case documentation process and outcome with reference to a prescribed way of writing case notes. A 10-step test development method was employed in constructing, analyzing and finalizing the scale using 151 participants. In writing the items, the Subjective, Objective, Assessment and Plan (SOAP) Note Format Standards of Cameron and Turtle-Song (2002) were considered. The instrument was subjected to experts' review and was pilot tested to establish its psychometric properties. Results yielded a Cronbach's Alpha of 0.92, indicating a high level of internal consistency. The factor loadings of 0.40 or higher retained 30 items. The loadings observed ranged from 0.42 to 0.82. The newly derived subscales of the Case Notes Assessment Scale are Record Content (14 items), Brevity of Notes (6 items), and Writing Mechanics (10 items). This newly developed tool serves as guidelines or standards that can be used in evaluating the case documentation process and outcome of counselling professionals.

Keywords: case documentation, assessment, case notes, instrument development

The Guidance and Counseling Act of 2004 (R.A. No. 9258), seeks to professionalize the practice of guidance and counseling in the Philippines. In relation to this, the counseling professionals are expected to provide high quality services to students and other clientele. According to the International Association of Counseling Services, Inc., professional ethical practice forms the cornerstone of the counseling service. In particular, systematic case records, such as intake and assessment information, case notes, termination summary and tests, must be maintained as required by professional standards (Kiracofe et al., 1994).

This calls for accountability from counseling professionals who are expected to accurately document what has transpired during their counseling session or therapeutic hour in the form of case or progress notes (Cameron & Turtle-Song, 2002). Case notes, which are part of the student

records and kept in the Guidance Office, serve as tools in monitoring the clients' progress in counseling and in developing the counselors' case conceptualization and treatment planning skills. These records also serve as reference for other counselors or specialists who may handle the same client in an emergency case through referral procedures (Estanislao, 2010). The need to further improve the quality and continuity of client services in this regard remains to be a pressing concern among counseling professionals. What is and is not important enough to be incorporated in the case notes posed some questions for counselors who desire to be systematic in the delivery of client care.

Prieto and Scheel (2002) presented the STIPS format for organizing case notes that could increase the counselor trainees' case conceptualization skills. It consists of five major sections, such as, documenting clients' current Signs and Symptoms, Topics discussed in Counseling, Counseling Interventions used, Clients' Progress and Counselors' Continuing Plan for Treatment, and Special Issues of Importance regarding clients (e.g., suicidality). Such information can be used to enhance the trainees' ability to acquire relevant facts about the clients, to better understand the clients' presenting problems, to better monitor counseling the processes, and to better evaluate and adjust the treatment interventions.

Following this frame of mind, Travers (2002) emphasized the need of writing high-quality case notes, which can be time-consuming. According to him, these notes serve as a record for other counselors who may meet with the same client being handled in an emergency situation or when the client transacts to another counselor. Thus, case notes should be legible, efficient, and detailed. Other important uses of case notes are for supervision and court purposes. "A state or federal court of law can subpoena client files if legal action is taken against you or your client" (Travers, 2002, p. 36).

In the same light, Cameron and Turtle-Song (2002) reported a number of models that enable counselors to identify, prioritize and track client problems so that they can be attended in a timely and systematic manner. They also provided an on-going assessment of both the client's progress and counselor's treatment interventions. Components of these models include the Data, Assessment and Plan (DAP), Individual Educational Programs (IEP), Functional Outcomes Reporting (FOR), and narrative notes, which are all variations of the original Subjective, Objective, Assessment and Plan (SOAP) note format.

From the foregoing review of literature, it was observed that only a few number of research studies were done in relation to assessing existing case records. There seems to be no standard content or procedures in writing what has transpired during the counseling session. It is for these reasons that this study was conceived. It aimed to initially develop an instrument that will measure the counseling professionals' current case documentation process and outcome with reference to a prescribed way of writing case notes. Findings of this study hope to provide valuable information in mapping out gaps and limitations in the existing case records and practices of professionals given the recommended standards. Validated case note writing standards may also serve as guidelines in the conduct of

the profession and in the delivery of better client care services. This initially developed tool can be used by school counselors, educators, supervisors, and other mental health professionals in evaluating their case documentation in terms of content, brevity, and mechanics of writing notes.

The Problem-Oriented Medical Records Model

This study adapted Cameron and Turtle-Song's (2002) Subjective, Objective, Assessment and Plan (SOAP) Note Format Standards, as described in the Problem-Oriented Medical Records (POMR) Model. This model considers four areas: (a) Clinical Assessment includes intake interview information, tests, reason for help-seeking behaviors, reason for seeking treatment, secondary complaint, presenting concern, personal, family, and social histories; (b) Problem List covers active and inactive resolved list of the client's presenting and underlying problems derived from the clients history, numbered, dated, and entered into the list-attached in clients file, when resolved, dated, and made inactive; (c) Treatment Plan contains statement of possible therapeutic strategies and interventions to be used in dealing with each noted problem, stated as goals and objectives, and written in behavioral terms in order to keep track the client's therapeutic progress, or lack thereof; and (d) Progress Case Notes, using the SOAP Standards, serve to bridge the gap between the onset of counseling services and final session to clearly document and support, through subjective and objective sessions, decisions to modify or to bridge the gap between the onset of counseling services and final session or closure.

Cameron and Turtle-Song (2002) discussed the first component in terms of its contents such as information gathered during the intake interview/s. These generally include the reason the client is seeking treatment, secondary complaints considering the client's personal, family and social histories, psychological test results, if any, diagnosis and recommendations for treatment. They continue that from the Clinical Assessment, the second component emphasized on a Problem Checklist, which includes an index of all the problems, active or inactive, generated from the client's history. The third component of the POMR is a statement of the possible therapeutic strategies and interventions to be used in dealing with each noted problem. Treatment Plans are stated as goals and objectives are written in behavioral terms in order to track the client's therapeutic progress, or lack thereof.

Finally, the fourth component, progress case notes are generally written using the SOAP format and serve to bridge the gap between the onset of counseling services and the final session. In writing case notes, clear and concise document serves as references in order to organize the counseling professional's thinking about the client and to aid in planning of quality client care. An example of case notes based on the objective and subjective data reads: "Death of a loved one or diagnosed with a potentially life-threatening health problem..." By recording this information in the progress notes, it provides justification and documentation for sudden shift in therapeutic decision and is immediately able to address what is now the more pressing issue for the client. The next step, which should be recorded, may state: "Refer client to a domestic violence group for perpetrators". The subjective and objective sections of the SOAP notes would chronicle the

client's history of physical aggression and violent behaviors. Other benefits of case notes include improving the quality of continuity of client services, enhancing communication among mental health professionals, facilitating the counselor in recalling the details of each client's case, and generating an ongoing assessment of both the client's progress and treatment success (Cameron & Turtle-Song, 2002).

The SOAP Note Format

The Subjective, Objective, Assessment and Plan (SOAP) are described with examples and rules in writing case notes. These descriptions are: Subjective (S) - includes what the client tells you and what the significant others tell you about the client. Basically, it is how the client experiences the world. Client's feelings, concerns, plans, goals and thoughts, intensity of problems and impact on relationships, pertinent comments by the family, case managers, behavioral therapist, etc., client's orientation to time, place, and person and client's verbalized changes toward helping are recorded in this section. Objective (O) or Factual -contains what the counselor personally observes and witnessed. These are quantifiable, such as what was seen, counted, smelled, heard or measured outside written materials received. The client's general appearance, affect, behavior, nature of the helping relationship, client's demonstrated strengths and weaknesses, test results and materials from other agencies, etc., are to be noted and attached here. Assessment (A) - summarizes the counselor's clinical thinking, a synthesis and analysis of the subjective and objective portion of the notes. For counselors, inclusion of clinical diagnosis and clinical impressions (if any), are considered in this portion. For care providers, they look into how would you label the client's behavior and the reasons (if any) for this behavior? Plan (P) - describes the parameters of treatment, consists of an action plan and progress format and standards. This section also includes interventions used, treatment progress, and direction. Counselors should include the date of next appointment. In addition, prognosis is determined, whether there can be anticipated gains from the interventions.

The rules in Case Note Writing using the SOAP Note Format consider the following: keep notes brief and concise; maintain quotes to a minimum; use an active voice and in precise and descriptive terms; record notes immediately after each session; start each new entry with date and time of session; write legibly and neatly; use proper spelling, grammar, and punctuation; document all contacts or attempted contacts; use only black inks if notes are hand written and sign-off using signature, plus your title. Moreover, what to avoid in writing case notes are the following: using names of other clients, family members, or other names mentioned by the client, terms like "seems, appears" (uses senses), value-laden language, common labels, opinionated statements, the use of terminology unless trained to do so, erasing, obliterating, using correction fluid or in any way in attempting to obscure mistakes. Finally, blank spaces between entries and squeezing of additional commentary between lines or in margins should not be done.

The abovementioned SOAP descriptions, examples, and rules in case note writing was considered in the construction of the Case Notes Assessment Scale, which went through content and factor analyses.

The Present Study

This study aimed to initially develop an instrument that assesses the current case documentation process and outcome of counseling professionals in accordance to prescribed note format and standards. Specifically, it sought to answer the following questions: (1) What are the factors or subscales that will measure the counseling professionals' current case documentation process and outcome with reference to a prescribed way of writing case notes? (2) Is the Case Notes Assessment Scale (CNAS) a reliable instrument in measuring the counseling professionals' current case documentation process and outcome?; and (3) Is the CNAS a valid instrument?

Method

Participants

One hundred fifty one (151) counseling professionals, who are employed in different schools, colleges, and universities in Metro Manila and provinces across educational levels at the time of this investigation were randomly tapped as participants for this study. Majority of them were female (76.16%) and 14.57% were male. More than half of these participants (52.32%) were married; 31.79% were single; and a handful of them (15.89%) did not indicate their marital status. In terms of age, the oldest was noted at 68 years and the youngest, 20 years, with a group mean age of 37.71. The counseling professionals' profile further revealed that the longest number of counseling years of experience was 28 and the shortest was less than a year, with a mean of 10.53 years. Finally, close to half of the participants (44.37%) have obtained their Bachelor's Degree; 50.34% of them have finished or are still working on their Master's Degree when these data were gathered; and the rest 5.29% obtained or were working on their Doctoral Degree.

Measures

The Case Notes Assessment Scale (CNAS) was administered to the participants of this study in order to establish its psychometric properties. The CNAS contains a 30-item, 5-point Likert type self-report scale following Cameron and Turtle-Song's (2002) Subjective, Objective, Assessment and Plan (SOAP) Note Format Standards. It measures the current case documentation process and outcome of counseling professionals based on the said prescribed standards. The participants were asked to rate their degree of agreement or disagreement to each of the statements using a scale from Strongly Agree (5) to Strongly Disagree (1).

Procedures

This study employed the 10-step procedure in test development, as suggested by Sevilla, Ochave, Punsalan, Regala, and Uriarte (1992). The steps include the search of content domain, item writing and review, development of the pre-try-out, main try-out and final forms, first trial and final test administration and evaluation tests validity and reliability, and plan for the development of norms. During the mid-section stage, copies of

the main try-out forms were fielded to the 151 participants. Some of these counseling professionals were requested to respond online and for those working in nearby universities and colleges from my workplace, they responded through the hand-carried forms.

Data Analysis

Factor analysis was primarily used to validate the Case Notes Assessment Scale (CNAS). It is a statistical procedure of reducing a large number of measures to a fewer number called factors. A factor matrix is derived which is comprised of each of the item correlation coefficients and the groupings of items under individual factors. Thus, the components of a construct are discovered. In this study, exploratory factor analysis was done using several factor solutions. In this analysis, the items were simply entered and the resulting factors were described (Gable, 1986).

A Principal Components Analysis with Varimax Rotation (with Kaiser Normalization) was finally performed using a 3-factor solution. Factor loading of 0.40 or greater was adopted for the screening of the items. The reliability of the CNAS was determined using the Cronbach's Alpha.

Results

Validity

Considering the exploratory factor analysis, factor solutions were done before determining the final selection. The three-factor was chosen because of the highest value obtained in the total variance was accounted for by this solution. An initial principal components analysis with Varimax rotation was performed on the entire sample of the 151 respondents but only 140 records were considered valid. They were included to maximize the possible range of item responses that would affect the inter-correlations between items entering factor analysis. The following factors extracted in the Varimax rotation were given these labels in describing the process and outcome of case documentation, namely, Record Content, Brevity of Notes, and Writing Mechanics.

Factor 1: Record Content. The first factor extracted displayed the most number of items ($n = 14$). All of them were observed to solely load on just this aforementioned factor. A factor loading of 0.40 or higher was adapted as a basis for screening these items. Factor loadings range from 0.43 to 0.78. This dimension describes the written information that should be found in case records indicating how the client experiences the world in terms of orientation to time and place, attitude towards counseling, and what significant others tell about the client. Facts or quantifiable observations in terms of the client's general appearance, affect and behavior, nature of the helping relationship, demonstrated strengths and weaknesses, test results and other materials, etc., are noted and attached here. Finally, a summary of the counselor's clinical analysis, prognosis and interventions are portions of these notes. Sample items include no. 9: "I document what I personally observed and witnessed (ex. nature of our relationship)." and no. 15: "I document action plans including interventions, treatment, progress and direction of counseling".

Factor 2: Brevity of Notes. The second factor extracted in the Varimax rotation consisted of the smallest cluster of six (6) items with loadings ranging from 0.42 to 0.82. Five (83%) of these items solely loaded on this factor. It describes the shortness of case notes indicating only general words and minimum quotes, which are documented in precise and descriptive ways. The rules are to be observed in terms of what should be included and how these notes are written. Sample items include no. 2: "I document only key words or a very brief phrase when using quotations." and no. 16: "My case notes are brief and concise."

Factor 3: Writing Mechanics. The third factor yielded 10 items. This factor describes how case notes are written and presented following the rules, format, and standards prescribed by the POMR - SOAP Model discussed earlier. Likewise, this factor identifies a number of standards or rules that should be followed or avoided in writing case notes. Sample items include no. 24: "My case notes are documented using proper spelling, grammar, and punctuation marks." and no. 30: "My case notes are not squeezed with additional commentary between lines and margins."

Table 1
Accepted Items with Their Factor Loadings

Item Number	Factor 1	Factor 2	Factor 3
3*	0.596		
4*	0.462		
5*	0.558		
7*	0.599		
8*	0.643		
9*	0.636		
10*	0.529		
11*	0.715		
12*	0.783		
13*	0.611		
14*	0.714		
15*	0.652		
19*	0.435		
20*	0.446		
1*		0.422	
2*		0.534	
6*		0.592	
16*		0.740	
17*		0.821	
18		0.506	
22*			0.444
23			0.517
24*			0.728
25*			0.569
26*			0.690
27			0.646
28*			0.519
29*			0.658
30*			0.697
31*			0.441

Note: * items that solely loaded in one factor

Reliability

Measures of internal consistency indicate the extent to which items on a test interrelate and represent similar content. They also provide a check on the content validity of the subscales since items within a subscale should be relatively homogeneous. In determining the reliability coefficient of the Case Notes Assessment Scale based on the 30 items, the Cronbach's Alpha was computed for the total scale. Results yielded a coefficient of 0.92, indicating a very high level of internal consistency.

After these initial procedures, the final form entitled Case Notes Assessment Scale (CNAS) was finalized. This 30-item scale assesses the current case documentation process and outcome of counseling professionals in accordance to prescribed note format and standards.

Scoring Procedures

In scoring the CNAS subscales and overall score, ratings for each item per cluster are added: For Factor 1 (Record Content), items included were nos. 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 19, and 20; Factor 2 (Brevity of Notes), items included were nos. 1, 2, 6, 16, 17, and 18; Factor 3 (Writing Mechanics), items included were nos. 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30. In interpreting the raw scores per subscale and overall score, the following arbitrary scaling was computed:

4.500 and above	Strongly Agree (Outstanding)
3.500 - 4.499	Agree (Very Satisfactory)
2.500 - 3.499	Neutral (Satisfactory)
1.500 - 2.499	Disagree (Needs Improvement)
1.499 and below	Strongly Disagree (Poor)

Discussion

It is evident from the results of this study that the Case Notes Assessment Scale (CNAS) is a valid and reliable instrument. This scale was constructed using Cameron and Turtle-Song's (2002) Subjective, Objective, Assessment and Plan (SOAP) Note Format Standards, as described in the Problem-Oriented Medical Records (POMR) Model. The derived factors are Record Content, Brevity of Notes, and Writing Mechanics. Findings suggest that counseling professionals have to consider these three components in assessing their case notes to ensure the appropriate process and outcome of written proceedings of the counseling session. Findings also provide valuable information in mapping out gaps and limitations in the existing case records and practices of professionals given the recommended standards.

In particular, Factor 1, Record Content enumerates all the essential elements covered by the SOAP model in terms of inclusions or substance of case records. The next two Factors 2 and 3, Brevity of Notes and Writing Mechanics, focused on the rules prescribed in writing progress notes - what to do and avoid in documenting what has transpired during the therapeutic hour. This validated the applicability of Cameron and Turtle-Song's (2002) note format standards in the Philippine milieu. Moreover, in terms of the number of items left after the factor analysis procedures, unequal numbers were observed for each factor. Factors 1, 2, and 3 have items that consist of

14, 6, and 10, respectively. However, all of the factors describe the different dimensions or characteristics required from a well documented case notes that are considered precise and brief. This observation also lends support to a study of Hansen (2009), suggesting a simple one-page anecdotal notes or summary of the time spent to a particular student for referral and/or reporting purposes.

The foregoing three factors of the CNAS, which were constructed empirically through factor analysis, maybe used by the counselor educators, supervisors, and counselors as an objective checklist on certain characteristics and standards necessary to help write well-documented case notes. Such information will also be helpful in reviewing the counselor's case records during supervision, in designing continuing education programs, and coming up with a support system for professionals' enhancement in the area of case documentation.

Finally, in the process of establishing the psychometric properties of the CNAS, this study undertook the initial stages of testing its validity (item review and factor analysis) and reliability. Thus, a validation study is hereby recommended to give more detailed properties of the test. Additional test items are also needed for inclusion in the item bank and larger sample size for its norming and interpretation of test scores may be considered. Furthermore, a review of the individual item or item analysis may be done for future research studies. Other helping professional populations in the different settings aside from schools (e.g., community, industrial, and clinical sites) may likewise be tapped as added norm groups since the test items cover case documentation processes and outcomes that are applicable to all those who are engaged in mental health and client care services.

References

- Cameron, S., & Turtle-Song, I. (2002). Learning to write case notes using the SOAP format. *Journal of Counseling & Development*, 80, 286 - 292.
- Estanislao, S. A. (2010). Documenting what has transpired during the therapeutic hour: An exploratory study. *The Guidance Journal*, 38(1), 22-37.
- Gable, R. K (1986). *Instrument development in the affective domain*. Boston: Klumer Academic Publisher.
- Hansen, S. (2009). *Documentation Tips and Guidelines*. School-Counseling-Zone.com. online [Aavailable]
<http://www.school-counseling-one.com/documentation.html>.
- Kiracofe, N. M., Donn, P. A., Grant, C. O., Podolnick, E. E., Bingham, R. P., Bolland, H.R.,..... Yamada, K.T. (1994). Accreditation Standards for University and College Counseling Centers. *Journal of Counseling and Development*, 72, 38 - 43.
- Prieto, L. R., & Scheel, K. R. (2002). Using case documentation to strengthen counselor trainees' case conceptualization skills. *Journal of Counseling and Development*, 80, 11-21.
- Sevilla, C. G., Ochave, J. A., Punzalan, T. G., Regala., B. P., & Uriate, G. G. (1992). *Research Methods* (Revised Edition). Quezon City: Rex Printing Company, Inc.

The Guidance and Counseling Act of 2004, RA 9258 (2004).

Travers, P. (2002). *The Counselor's Helpdesk*. California: Thomson Learning.

Correspondence may be sent to:

susana.estanislao@dlsu.edu.ph or susana.estanislao@yahoo.com



Assessment of Family Functioning among Perpetrators of Physical Abuse

Aime T. Guarino

De La Salle University, Manila, Philippines

Glenda Mae V. Santos

Brokenshire College, Philippines

Abstract The present study assessed family functioning and intergenerational patterns of physical abuse of adult among perpetrators. Family Assessment Measure III and Genogram were used as assessment tools. Fifteen perpetrators of physical abuse participated in the initial assessment and only seven out of the fifteen participants continued with Genogram interview. Respondents were in the early to middle adulthood stage and most of them were females. Majority of the respondents have not reached tertiary education. They belong to the low socio-economic status and have no stable source of income. The assessment conducted was anchored on the Structural Family Theory of the systems perspective. The assessment was made by: (1) filling out of personal data sheet, (2) taking the family assessment tool that measured the perpetrators' strengths and weaknesses in terms of communication, involvement, control, values, and norms, and (3) constructing of three generation genogram focused on physical abuse. Results showed difficulty in communicating effectively as a family and in setting boundaries at home. Moreover, power struggles, inconsistent regard to explicit and implicit rules at home, control issues and discrepant value system of the family were seen. Consistent with these problems in the family system, genogram observed the presence of intergenerational pattern of physical abuse. A report of the result was forwarded for appropriate action and intervention.

Keywords: family dysfunction, physical abuse, intergenerational pattern

In the Philippines, the implementation of The Special Protection of Children Against Child Abuse, Exploitation and Discrimination Act or Republic Act 7610 (Art.1 Sec. 3A) guards children with ages 18 years and below from any form of abuse. The programs of Department of Social

Welfare and Development (DSWD) and other non-government organizations to protect children are commendable. The existence of Bantay Bata 163 in rescuing abused children and attending to their welfare is a significant contribution to the society. However, despite these laws and programs, international and local statistics consistently show that the increase of child abuse cases is alarming. The abused continues to be more abused while the abuser keeps on abusing and the cycle of violence fails to stop (Bancroft & Silverman, 2002).

In an effort to promote mental health that starts from the family, the Counseling and Psychological Wellness Center (CPWC) collaborated with Bantay Bata 163; Davao City in addressing psychological needs among their clientele.

In systems perspective, presenting problems like physical abuse serves as a communication to the family that a dysfunction exist (Young & Long, 2007). This assessment study looked into the family systems structural sources of dysfunction. Genogram is an assessment tool used across family therapy theories (Corey, 2006). The genogram together with paper pencil test on relationship measures or inventories was the first step in the comprehensive intervention program. During the signing of the informed consent, the intake counselor educates the clients of the recommended procedures that would be best for the clients to undergo before therapy. Psychological assessment prior to counseling is encouraged. This assessment is a series of standardized personality measures, which includes mental ability, emotional competence, and relational skills. In addition to testing, the intake counselor invites family members together with the identified client to map a three-generation genogram. Results of the assessment are used to conceptualize the problem and treatment plan. Psychological report was made and feedback is given to the client(s). Counseling/family therapy proceeds after the assessment. Thus the respondents of the present study underwent the same process to come up with an appropriate intervention for the whole family experiencing physical abuse. However, the assessment done for this study was focused on family functioning and intergenerational pattern of physical abuse.

Previous reviews indicate how family violence is assessed using different protocols. Milner and Murphy, (1995) identified assessment approaches to evaluate offenders of violence as discussed below. It includes the use of interviews, observations, general personality measures, and offender specific measures. These evaluations are needed for a variety of purposes, which include screening, report confirmation, treatment planning, treatment evaluation, and recidivism prediction. Interviews attempt to assess offender characteristics using a question and answer format. The success of structured interview depends on a variety of factors such as the competence of the interviewer and client condition. Interview process may have lower reliability and validity than standardized objective tests; however, it allows for idiographic assessment. Observational methods have been used in natural and laboratory settings. The advantage of this approach is the direct observations of the behavior but results may yield mixed levels of sensitivity and specificity ranging from chance to adequate levels. General psychological assessment used abuse-related standardized questionnaires that identify significant areas of dysfunction in the offender

and offenders' family that are in need of remediation. This may be appropriately used for screening, treatment planning, and evaluation. Another evaluation tool used is specialized offender assessment techniques. There are two types of this assessment tool: physiological assessment and self-report questionnaire. To sum up the authors' idea, they gave the responsibility to the test users as to the choice of assessment techniques and further maintained that given a wide variety of measures and applications, it is the responsibility of the test user if the validity and reliability are sufficient for a particular application for a specific population.

In another study, Eckhardt and Utschig (2007) suggested that assessment among perpetrators of violence should begin with assessment of readiness to change before undergoing intervention program for a sustained change. The same authors made use of Prochaska et al.'s (1992) model of the processes experienced by perpetrators as they attempt to change. The path to intentional behavior change consists of five distinct stages of change: (1) pre-contemplation, (2) contemplation, (3) preparation, (4) action, and (5) maintenance. Researches on perpetrators indicated that clients who terminated therapy prematurely are more likely to be in the pre-contemplative stage of change. Movement through stages of change predicted client persistence in treatment and overall therapy outcome (Prochaska et al., 1992 as cited by Eckhardt & Utschig, 2007). Aside from readiness assessment, risk assessment among perpetrators is also a very important assessment considered by helping professionals for safety and readiness purposes (WHPPU, 2007).

On the other hand, family therapists observed that clients have a hard time attending therapy sessions even if the intervention program is summoned by the court or institutions where the clients are affiliated. Uebelacker, Hecht, and Miller (2006) developed a Family Check-up brief intervention to counter barriers of not coming to family therapy. This brief intervention program could be a two to three sessions only. It is interesting to know that almost 50% of the intervention is assessment. Assessment is in itself therapeutic. Battery of tests is conducted to clients but assessment occurs prior to and at the first meeting. Questionnaires are emailed to the clients for them to fill out in private to save time during the first session. Giving feedback of the assessment consumes the first session. Basically, Family Assessment Device (FAD) is administered to clients. It is a 60-item self-report questionnaire with seven subscales, six assess each of six areas of family functioning and the seventh assesses general family functioning. Analyses are often conducted on mean FAD scores. The FAD was used to provide feedback to families and to assess change overtime.

Skinner, Steinhauer, and Barbara's (1995) scales on communication, control, involvement, and values and norms as evaluated by the Family Assessment Measure (FAM III) were used as the areas of concern manifesting structural dysfunction. Communication refers to the achievement of mutual understanding so that the message received by the perpetrator/child is similar with the message intended. Control refers to the process by which the perpetrator influences, manipulates, or exercises power over the child. Involvement refers to both the degree and quality of the perpetrator's

interest in the family. Values and norms refer to the beliefs, ways, and rules explicitly or implicitly followed or manifested by family members.

Coming from the family systems perspective, this study focused on the assessment of family system and intergenerational patterns of physical abuse cases. This study wanted to get the perspective of the perpetrators as baseline data in designing intervention program since there are already available data coming from the victims at Bantay Bata 163 Davao City. Using the family assessment tool and genogram, this study aimed to determine the strengths and weaknesses of the family in terms of communication, involvement, control, values, and norms and to trace intergenerational patterns of physical abuse.

The Present Study

This study was anchored on systems perspective specifically using Structural Family Therapy by Salvador Minuchin. Family therapists explained that Structural family therapy considers three dimensions in conceptualizing family dysfunction (Corey, 2006; Kaslow & Massey, 2004; Young & Long, 2007). Problems could stem from issues in boundaries, power, and alignment. Family is functional or dysfunctional based upon its ability to adapt to various stressors, which relies on the clarity and appropriateness of its subsystem boundaries. The same author described boundaries as characterized along a continuum from enmeshment through semi-diffused permeability to rigidity. In addition, family subsystems are characterized by a hierarchy of power; usually the parental subsystem is on top of the offspring subsystem. In healthy families, parent-children boundaries are both clear and semi-diffused, allowing the parents to interact together with some degree of authority portraying the parenting roles. From the children's side, the parents are sufficiently enmeshed from the children to allow for the degree of autonomous sibling and peer interactions that produce socialization, yet not so rigid or aloof as to ignore childhood needs for support, nurturance, and guidance.

In addition to structural theory, Bowen's theory on multigenerational transmission is incorporated in the assessment using genogram. This assessment tool inquired and validated family structural issues and intergenerational pattern of physical abuse. Multigenerational Family Therapy operates on the premise that a predictable pattern of interpersonal relationships connects the functioning of family members across generation. An individual's problems can be understood by viewing the role of the family as an emotional unit. Emotional problems will be transmitted from one generation to another until resolved emotional attachments are dealt with effectively (Corey, 2006).

This study sought to determine if the structural family functioning issues that includes boundaries, power, and alignment, existed among the families with physical abuse cases as perceived by the perpetrators. The subscales on communication, involvement, control, and values and norms from the Family Assessment Measure III helped in the in-depth exploration of family dysfunction. This assessment can be used by the Counseling and Psychological Wellness Center (CPWC) and Bantay Bata 163 Davao as baseline data for further psychological intervention.

On the other hand, the present study inquired about the existence of physical abuse across generations in the perpetrators' lineage as postulated by the theory on Multigenerational Family Therapy. McGoldrick and Gerson (1985) described genogram as a format for drawing a family tree that records information about family members and their relationships over at least three generations. He further explained that genogram displays family information graphically in a way that provides a quick gestalt of complex family pattern and a rich source of hypotheses about how a clinical problem may be connected to the family context and the evolution of both problem and context over time.

The study aimed to determine the family functioning of the perpetrators of physically abused children in Bantay Bata 163. Specifically, it aimed to answer the following: (1) what is the socio-demographic profile of perpetrators of physical abuse in Bantay Bata 163? (2) What is the level of family functioning among the perpetrators of physical abuse in the following scales: communication; involvement; control, and values and norms? (3) Is intergenerational pattern of physical abuse present among the families of the perpetrators?

Method

Research Design

This study used a descriptive design. It described the socio-demographic profile, the level of family functioning, and the intergenerational patterns of physical abuse among the perpetrators.

This study was conducted in two modes of assessment. The Family Assessment Measure III was administered among perpetrators of abuse in Bantay Bata 163 located in Km.4 McArthur Hi-way, Matina, Davao City. The other mode used in-depth interview using genogram was conducted individually at the Counseling and Psychological Wellness Center, Brokenshire College, Madapo City.

Respondents

The study used purposive sampling due to the dearth of respondents who willingly participated in this study considering the sensitivity of the issue on physical abuse. This study was participated by fifteen (15) perpetrators. They were identified as immediate and extended family members of the victims who were reported to Bantay Bata 163 for committing physical abuse. These fifteen (15) participants took the Family Assessment Measure III. All participants referred by Bantay Bata 163 were present during the test administration. However, only seven (7) participants went on to the interview phase of the assessment.

Instruments

Family Assessment Measure III (FAM III). The FAM III by Skinner et al. (1995) is a standardized instrument used to measure the level of functioning of perpetrators in terms of task accomplishment, role performance, communication, affective expression, involvement, control, and values and norms. It consisted of three (3) components namely: (1) General scale which focused on the family as a system, (2) Dyadic relationship scale which

examined relationship of family members, and (3) Self-rating scale which tapped the individual's perception of his/her functioning in the family. This instrument was validated to people in Canada and United States of America across different ages, genders, and family background specifically to natural and foster families of family members who were suffering from clinical conditions such as depression, alcoholism, anxiety disorders, anorexia, bulimia, manic-depressive, social phobia, schizophrenia, etc. This was also tested to gifted, handicapped, foster children and other participants who came from distressed relationships. The general scale and dyadic relationship scales of FAM III had an internal consistency reliability of .93 and .95 respectively while the self-rating scale had a reliability of .89 (Skinner et al., 1995). Prior to test administration of FAM III, the tool was translated to Cebuano on December 2009 and was back translated to English on January 2010 with the assistance of different experts in Visayan language in order to show consistency in thought and content. The help of three (3) experts in the field of Psychology was sought to validate the Cebuano-translated assessment tool. The tool was pilot-tested to seven perpetrators of physically abused children from Men's Responsibilities in Gender and Development (MR GAD), a non-government organization. The test on reliability analysis provided a Cronbach's alpha of .80, which is considerably high.

Genogram. The genogram is a multigenerational family tree that plots familial relationships and visually records information about social and psychological issues in the family across three (3) generations. This was used in this study to determine the presence of intergenerational abuse among the perpetrators of physically abused children in Bantay Bata 163 (McGoldrick, 1995).

Personal Data Sheet. The personal data sheet is a document that shows the respondents' agreement to participate in this study. It included personal information of the participants like: the participant's name, gender, age, civil status, address, religion, contact number, educational attainment, and occupation. The second part assured confidentiality of information to the respondents.

Procedure

FAM III was administered to the group together with the assistance of other counselors to attend to the participants' testing needs. Genogram was conducted to each respondent at the counseling cubicles of the researchers in CPWC. Fellow help providers in the center helped conduct the in-depth interview. The counselors of the center are competent in using genogram as an assessment tool as it is a standard operating procedure before counseling clients at CPWC.

Data Analysis

Descriptive statistics was used to describe the respondents' socio-demographic profile and FAM III results. Percentage, mean, and standard deviation were used to analyze the data. Constructing of three-generation genogram was used to see the existence of family functioning focusing on physical abuse.

Results

Table 1 presents the participants' age, gender, presence of intergenerational abuse, educational attainment, economic status, religion, and tribe.

Table 1
Profile of the Participants (N=15)

Demographics		f	%
A. Age	Early Adulthood (20-40 years old)	9	60%
	Middle Adulthood (41-60 years old)	6	40%
B. Gender	Female	12	80%
	Male	3	20%
C. Presence of Intergenerational Abuse	Present	10	70%
	Not Present	5	30%
D. Educational Attainment	Elementary		40%
	Undergraduate	6	
	Elementary		27%
	Graduate	4	
	High School		20%
	Undergraduate	3	
E. Economic Status	College	2	13%
	Undergraduate		
F. Religion	Low*	15	100%
G. Tribe	Roman Catholic	14	93%
	Iglesia ni Cristo	1	7%
G. Tribe	Cebuano/Bisaya	10	70%
	Ilonggo	2	10%
	Mandaya	3	20%

Note: *Based on Department of Labor and Employment Wage Order No. RB XI-14

Table 1 shows that participants who belong to the middle adulthood stage are relatively higher in proportion compared to those in the early adulthood. The oldest participant is 62 years old while the youngest is 28 years old. Eighty percent are female and only twenty percent are male participants in the FAM III assessment. Seventy percent said they experienced to be physically abused as children. Most of them have not reached college level. They belonged to low socio-economic status and majority are Cebuano and Roman Catholic.

Table 2 presents the respondents' family functioning in terms of communication, involvement, control, values and norms. The four aspects

of family functioning are shown in each of the three scales namely general, dyadic, and self-rating scales.

Table 2

Family functioning of the perpetrators of physically abused children in Bantay Bata 163 (N=15)

		Communication	Involvement	Control	Values and Norms
General Scale	Mean	57	57	61	59
	SD	6.79	10.22	11.93	7.17
	Description	Average	Average	High Average	Average
Dyadic Scale	Mean	59	61	62	62
	SD	9.25	8.84	6.52	16.70
	Description	Average	High Average	High Average	High Average
Self-Rating Scale	Mean	59	62	57	54
	SD	9.41	8.69	8.41	5.37
	Description	Average	High Average	Average	Average

Table 2 under the general scale row shows average scores in communication ($M = 57$, $SD = 6.79$), involvement ($M = 57$, $SD = 10.22$), and values and norms ($M = 59$, $SD = 7.17$) and a high score in control ($M = 61$, $SD = 11.93$). The respondents looked at their overall family system as unhealthy. Scores 57 to 61 indicate problem areas in the family relationship. The higher the score one gets the more problems the family has. High scores mean weakness in the family interaction.

Secondly, when it comes to the respondents' relationship to a family member (refer to dyadic scale), they obtained high scores in involvement ($M = 61$, $SD = 8.84$) control ($M = 62$, $SD = 6.52$) and values and norms ($M = 62$, $SD = 16.70$). Results show that the perpetrators view the relationship with one another as dysfunctional.

Lastly, in the self-rating scale, respondents scored high in involvement ($M = 62$, $SD = 8.69$), which means weaknesses in how each member of the family engages with each other. High score in involvement validated by the consistent average to high scores in all other scales demonstrate disengagement.

Constructing a genogram aimed to see patterns of physical abuse of the respondents' respective families within the three generations. Intergenerational pattern of physical abuse existed within the families whose present problem involves physical abuse of adults. Results show presence of intergenerational physical abuse. Majority of the perpetrator respondents experienced being physically abused as children. In turn they also physically abuse their children/grandchildren. At least three among the respondents shared that they are also being abused by their present spouses / partners. The abuse comes in different forms or a combination of all forms: Physical, verbal, and emotional abuse. A sample written report of

genogram result is shown in Appendix A. These Seven (7) respondents willingly submitted themselves for this assessment.

Discussion

Perpetrators of physical abuse are in the early to middle adulthood stage and mostly are females. Majority of the respondents have not reached tertiary education. They belong to the low socio-economic status and have no stable source of income. They are mostly Roman Catholics and majority is Cebuano. Data on educational attainment show that all the respondents of the study failed to finish the tertiary level of education. Respondents point out poverty as the reason for not being able to finish school and lack of motivation to pursue their education because their primary concern is to put food on the table in order to satisfy their physical needs more than pursuing their education. This confirms the study of Gil as cited in Corby (2000) that most of the reported parents of physically abused children belong to low social classes and depend on the support of institutions. All of the respondents come from low economic status. Classification was based on the Department of Labor and Employment specifically Wage Order No. RB XI-14, which indicated the ranges of the minimum wage in Davao City between 219 to 250 pesos depending on the sector. Six of the respondents have no stable source of income while one falls below the minimum wage range. This implies that all respondents lack the capacity to live comfortable lives. Three of them are jobless while the rest are scavenger, street sweeper, cosmetic and coffee dealer. In fact, there were, initially, 15 respondents; however, only seven respondents continued with the in-depth interview using genogram in a scheduled face-to-face individual session. Burdened with household tasks, lack of person to entrust their children, insufficient husband- or partner-support to attend the sessions were the reasons stated for not continuing with the assessment. The researchers and Bantay Bata 163 exhausted different measures to communicate with the respondents; however, they did not show up.

Respondents view their family system as generally unhealthy (dysfunctional) as revealed by the results of the general scale of FAM III. The four areas measured by FAM III exhibited weaknesses in communication, involvement, control, and values and norms. In terms of communication, respondents express a fluctuating pattern. Communication pattern at home is inconsistent which creates confusion in the family. This supports Murphy and Eckhardt (2005), which state that perpetrators, who usually belong in dysfunctional families, have difficulties on how to express feelings and reactions appropriately in times of conflict. Using the examples of the interview with the clients in this study, one of the clients with four children was left with the sole responsibility of rearing their children. She and her first husband had no communication but they stayed in the same house. Involvement is another area of concern. There is an absence of involvement among members in the way respondents view their family relationship. They seem to be distant to each other. Alexandrov, Cowan, and Cowan (2005) explained that when physically abused children begin to make sense of the abuse, they may begin to develop a notion that people are not safe, predictable or trustworthy, thereby, making attachment a dangerous risk

that they are no longer willing to take. In Structural theory, it is described as having disengaged interaction or disengagement (Corey, 2006; Young & Long, 2007). Disengagement is the opposite extreme of enmeshment in the continuum of the family's task to set boundaries in the system. Disengagement is very evident in one of the respondents in this study. She was separated with her husband due to extra-marital affair between her husband and her sister. They separated with her husband without resolving the conflict. The client, her husband, and her sister have not been talking to each other up to the time of the interview. They just seem not to know each other. In the same way, Bancroft and Silverman (2002) stated that perpetrators could cause divisions among family members. As per report of Bantay Bata 163, most of their physical abuse cases are by a family member. Some family members decided to be an ally of the perpetrator while the other family members side with the victim. This is evidence of what Bancroft and Silverman (2002) referred to as "great splits and alignments among family members." Control scale scores indicate a problematic area. Control issues are characterized by overt and covert power struggles when it comes to parenting and decision making (Skinner et al., 1995). The same authors described this family to exhibit rigidity and lacking of flexibility or spontaneity or extreme chaos. This supports what Tower (2006) mentioned that abusers have not learned to make decisions and are afraid to feel out of control. Using the experiences of the respondents, most of them exhibit a chaotic relationship. For example, one respondent narrated that she has many children from different partners. She is presently cohabiting with her fourth live-in partner for a year with whom she had a child. Almost all of her children except the youngest were under the custody of Bantay Bata. She always falls out of control in dealing with any of the family members at home. Similar to her growing up years and experiences in relationships, she was also physically abusive to her children. She slaps, pinches, and hits her children with a stick when she becomes irritated.

Values and norms scores indicate that the value system is unclear at home. Explicit and implicit rules that should and should not be followed by family members can be inconsistent. Feinauer (2006) mentioned that rules provide stability, manage involvement of family members and guides communication; however, these cannot be seen from the respondents' families. Skinner, et al. (1995) described high average scores of values and norms as dissonant family value system resulting in confusion and tension. Family's values may also be in conflict with those of the culture as a whole. From the interviews of the present study, one respondent grew up without her mother. Her mother left their family for another man. Hence, it was her eldest sister who took the role of being a mother to her and her siblings. She is the fourth child in the brood of nine. Her father used to be a photographer and was usually out of the house. It was their eldest sister who managed the household. Her father was physically beating her and eldest sister through slaps, pinches, and hitting of guava stick that resulted in bruises and hematoma. As a child, she admitted that she was naughty and was not interested in going to school.

Relationship between and among family members indicates presence of ineffective communication, distant relationship, demanding or rigid interaction, and unclear value system.

The respondents keep distance from any member of the family. Respondents also do not communicate their thoughts and feelings to the family but instead demonstrate their disappointments through physical punishments. A respondent in this study shared that her children grew up seeing her as the disciplinarian in the family. She used to hit her children with stick and pinched them in their waists to express her disagreements to children's misbehaviors. She also became physically abusive by using a stick to hit her granddaughter. She was also verbally abusive to her children and granddaughter. They have difficulty in communicating effectively with each other. They seem to be distant to each other. Adjusting to the changing life demands and cultural values and norms is hard and challenging and thus uses physical abuse to cope with the transition. Consistently, relationships with family members cannot achieve total harmony. Respondents of the study who are all parents fail to establish a predictable control style to their children. Bancroft and Silverman (2002) state that part of the issue on control is the family's capacity to maintain ongoing functions and adapt to change. However, the confusion on what parenting style to use to their children signifies failure to perform existing roles and adapting to change. In addition, perpetrators can develop a belief system that makes aggression an acceptable behavior as mentioned by Bancroft and Silverman (2002). Hence, the values and norms passed on to their children can be the negative views towards them as parents and on relationships. This causes perplexity that allows children to question certain beliefs that can further result in arguments. This is also similar to their attachments with their family members. At some point, they can be attached with each other; however, there are also instances that they can be aloof to other family members (Skinner et al., 1995; Young & Long, 2007).

Perpetrators exhibit disengaged relationship with other family members. Results show that the respondents keep distance from any member of the family. Respondents also do not communicate their thoughts and feelings to the concerned member of the family but instead demonstrate their disappointments through physical threats and punishments. A respondent in this study shared that her children grew up seeing her as the disciplinarian in the family. She used to hit her children with stick and pinched them in their waists to express her disagreements to the kids' misbehaviors. She also became physically abusive to her children and grandchildren.

Intergenerational pattern of physical abuse existed within the families whose present problem involves physical abuse of adults. This means that the pattern of physical abuse in the participants' families has been passed on from generation to another. Using a genogram, it has been known that their father, mother, grandmother, or elder siblings have physically abused these participants as a child. They experienced being pinched, hit by objects such as a stick or belt that caused bruises and wounds. They grew up getting used to being beaten and believed that those were the correct ways to instill discipline. This supports the studies of Buerkel-Rothfuss as cited in Galvin, Bylund and Bronnmel (2008), Steel and

Pollock as cited in Corby (2000), Tower (2002) that parents who abuse their children have once experienced abuse themselves. These studies consistently state that early childhood exposure to aggression is influential in committing physical abuse in the later years of life. On the other hand, 43% of the participants have not experienced physical abuse from generations before them. However, their parents were reprimanding them when they committed mistakes. Despite the absence of intergenerational abuse, the genograms revealed that the respondents experienced being abused by their spouses or live-in partner. All of them have experienced raising their children alone at a certain time in their lives. Research shows that individuals who experienced physical abuse in childhood were more likely to report physical illness, depression, anger, and anxiety than non-abused individuals and those who have been victims of childhood physical abuse are impeded in their ability to develop social and emotional skills. Larsen, Sandberg, Harper, and Bean (2011), subscribed to empirical findings that childhood physical abuse has been found to have negative effects on adult mental and psychological functioning. Genogram results further showed presence of ineffective patterns of communication, rigid parental control, and distant and conflicting relationships between and among members of the family. These genogram findings validated the results of FAM III.

References

- Alexandrov, E.O., Cowan, P. A., & Cowan, C. P. (2005). Couple attachment and the quality of marital relationships: Method and concept in the validation of the new couple attachment interview and coding system. *Attachment and Human Development*, 7(2), 123-152.
- Bancroft, L., & Silverman J.G. (2002). *The batterer as parent*. California: Sage Publications, Inc. Thousand Oaks.
- Corby, B. (2000). *Child abuse towards a knowledge base*. Buckingham: Open University Press.
- Corey, G. (2006). *Theory and practice of counseling and psychotherapy*. Australia: Brooks/Cole.
- Feinauer, I. D. (2006). The relationship of implicit family process rules to adolescent psychological symptoms (Doctoral dissertation). Retrieved from <http://contentdm.lib.byu.edu/ETD/image/etd1328.pdf>
- Eckhardt, C. I., & Utschig, A.C. (2007). Assessing readiness to change among perpetrators of intimate partner violence: analysis of two self-report measures. *Journal of Family Violence*, 22, 319-330.
- Galvin, K. M., Bylund, C. L., & Bronnmeel, B. J. (2008). *Family communication cohesion and change*, (7th ed.). USA: Pearson Education, Inc.
- Kaslow, F., & Massey, M. (2004). *Comprehensive handbook of psychotherapy*. New York: John Wiley & Sons.
- Larsen, C. D., Sandberg, J. G., Harper, J. M., & Bean, R. (2011). The effects of childhood abuse on relationship quality: Gender differences and clinical implications. *Family Relations*, 4, 435-444.
- McGoldrick, M., & Gerson, R. (1985). *Genogram*. New York: Penguin Books.

- Milner, J. S., & Murphy, W.D. (1995). Assessment of child physical and sexual abuse offenders. *Family Relations*, 44, 478-488.
- Murphy, C. M., & Eckhardt, C. I. (2005). *Treating the abusive partner*. New York: The Guilford Press.
- Skinner, H. A., Steinhauer, P. D., & Barbara J. S. (1995). *Family assessment measure version III*. New York, USA: Multi-Health Systems, Inc.
- Tower, C. C. (2002). *Understanding child abuse and neglect*. USA: Pearson Education Co.
- Tower, C. C. (2006). *A clergy guide to child abuse and neglect*. Cleveland, Ohio: Pilgrim Press.
- WHPPU (2007). *Guidelines for responding family and domestic violence*. National Library of Cataloguing in Publication.
- Uebelacker, L. A., Hecht, J., & Miller, I. W. (2006). The family check-up: a pilot study of a brief intervention to improve family functioning in adults. *Family process*, 45, 223-236.
- Young, M., & Long, L. (2007). *Counseling and Therapy for Couples*. USA: Brooks/Cole Publishing.

Correspondence may be sent to:
aime.guarino@dlsu.edu.ph

Appendix A

(Sample Genogram Report)

Client X:

Family functioning

Client X has ten children with her husband. None of them finished elementary education. Their children had no interest in school and would prefer to work in the streets like looking for parking spaces or taxi cabs in exchange for pennies. Similar to their children, Client X's husband failed to finish elementary education while Client X finished elementary. Her husband used to be a construction worker and was often not home. Client X was the one who raised the children and managed the household. When her husband was dismissed from work, Client X had to work as street sweeper to provide for their physical needs. Client X admitted that she does not want to have many children; however, her husband would physically hurt her when she does not give in to his sexual urges and the use of contraceptive is not allowed.

Relationship pattern

Members of the family do not openly expressed what they think and how they feel. This lack of communication is also seen in the way the couple interacts with each other. But, once problem arises, verbal, emotional, and physical abuse comes into the fore. Client X expressed that she loves her family but she finds it difficult to control her emotions when angry. Client X views her relationship with her family as distant and conflicting.

Physical abuse

Intergenerational abuse is present in her family. Client X and her husband grew up having disciplinarian parents who also believed in using physical force to correct misbehaviors. Hence, they both use physical force as a way to discipline their children. Their granddaughter, who is reported to have been physically abused by Client X, was taken by Social Workers and was under the temporary custody of Bantay Bata 163 Davao at the time of interview. The child will be returned as soon as safety is ensured at home. Client X welcomes the idea of regular home visitations by Bantay Bata 163 representative to ensure the safety of the child. In addition, to how physical abuse was used by Client X to communicate disappointments at home, she is also physically abused by her husband. Her husband hit her using his fist, slapping her, and cursing her especially when he is drunk.

Findings:

The following themes were found: (1) Extended family system with many children, (2) Less value on education, (3) Presence of ineffective communication, (4) Rigid parental control, (5) Existence of distant and conflicting relationship, and (6) Manifestation of intergenerational pattern of physical abuse.



Reinforcement Sensitivity as Predictor of Test Anxiety

Rene M. Nob

St. Paul University - Manila, Philippines

Abstract In an attempt to advance the revised Reinforcement Sensitivity Theory (RST), this study investigates the implications of RST to test anxiety. 217 college-level students participated in the study. The Behavior Inhibition System (BIS) and the Fight, Flight, Freeze System (FFFS) of the RST correlate with test anxiety, but not the Behavior Activation System (BAS). Using multiple regression, findings reveal that FFFS, specifically freeze, best predicts test anxiety, worry, and emotionality. Moreover, fight also predicts worry, while flight predicts emotionality. These results generally support the current conceptualization of the revised RST.

Keywords: Reinforcement Sensitivity Theory, test anxiety

The idea that one's personality is a reflection of biological processes is nothing new. Early theories, such as Murray's Personology, have long argued that personality is rooted in the physiological activities of the brain (Schultz and Schultz, 2005). Intuitively, it does seem logical to assume the validity of this position, given that the brain is generally understood to be responsible for most activities of a human person, including the manifestation of personality.

Among the theories that explain personality in the light of biological-based systems, Gray's Reinforcement Sensitivity Theory (RST) remains to be one of the most influential (Smillie, 2008a). Created during the 70s, RST has undergone a considerable revision in the past decade (Pickering & Corr, 2008; Smillie, Pickering & Jackson, 2006). In its current conceptualization, it proposes that there are three brain-based systems concerned with motivation, emotion, and learning that are believed to mediate individuals' responses to stimuli that are rewarding (i.e., incentive), punishing (i.e., threat or non-reward), or conflicting (i.e., stimuli that signify both incentive and threat) (Smillie, 2008a). Respectively, the systems are termed as the following: behavioral activation system (BAS), fight-flight-freeze system (FFFS), and behavioral inhibition system (BIS) are neurological-based, (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006) but may also be viewed behaviorally.

According to Pickering and Corr (2008), Smillie (2008a), and Smillie et al. (2006), behaviorally, activities of the BAS is observed in an individual's approach response to appetitive or rewarding stimuli, both conditioned and unconditioned. The FFFS mediates defensive avoidance responses to conditioned and unconditioned threats or punishing stimuli. Furthermore, the specific response (i.e., fight, flight, or freeze) depends on certain ecological characteristics of the threat. If the threat is proximate and escape is not an option, 'fight' is the likely response. If the threat is distant, and there is a possibility for escape, the likely response is 'flight'; but if escape is not possible, 'freezing' is more likely. It is important to note that the three behavior manifestations under FFFS are different from one another, but they are functionally woven into a single system (Eilam, as cited in Perkins, Kemp, & Corr, 2007). Meaning, all three behaviors serve the general purpose of defensive avoidance. If the stimuli present conflicting valences (i.e., a goal that presents both possible incentive, and threat or non-reward), the BIS mediates a defensive approach response. According to Smillie et al. (2006), BIS can be seen as sensitivity to situations characterized by uncertainty, social comparison, and goal conflict.

It is generally thought that individual differences in the sensitivity of these systems are reflected in individual personalities (Smillie, 2008a). Theoretically, some traits are assumed to be the manifestations of the systems, such that, extraversion is thought to manifest the reactivity of BAS (i.e., reactivity to rewards); fearfulness for FFFS (i.e., reactivity to punishment), and anxiety in the case of BIS (i.e., reactivity to conflicting stimuli) (Smillie, 2008a). However, more recent investigations suggests that there is a need for a new conceptualization of BIS of the revised RST since anxiety seems to correlate more with FFFS, than it does with BIS (Jackson, 2009). There has been much difficulty in separating the constructs of fear and anxiety, since both are negative emotions aroused when confronted with perceived threats (Smillie et al., 2006). A review by Perkins et al. (2007) of by various studies has found consistent relationship between anxiety and fear. While there is no finality as regards this issue, it is best to understand BIS and FFFS in terms of their behavioral manifestations and functional value, rather than assuming them to represent known trait labels.

Given the abovementioned theoretical descriptions, we may anticipate that an individual with high BAS sensitivity can be readily attracted to perceived rewards, and has the tendency to be impulsive, given certain appropriate situations. Individuals with high FFFS sensitivity readily perceive, and are easily incapacitated by punishments and threats. They tend to be avoidant of discomfort, and evasive of perceived dangers, whenever escape is possible, and will react with defensive aggression when cornered or desperate. As for individuals with high BIS sensitivity, they are likely to be tentative in engaging what might be an incentive-gaining behavior, due to perceived threats. They are inhibited by risk assessment and are careful in their actions (Mitchelle et al., 2007).

One critical area in RST is measurement. Being based on neurological systems, direct measures of the reactivity of the systems are possibly better captured using neurological or biological paradigms (Smillie, 2008a). While neurological methods seem promising, they are still in its infancy, and are confronted with their own challenges (Smillie, 2008a). Most psychologists,

on the other hand, who are trained as social scientists, rely on the development of purpose-built scales to assess these constructs (Jackson, 2009). Corr (cited in Smillie, 2008b) argues that questionnaires that assess reinforcement sensitivity should be theoretically faithful. As with current convention, the use of self-reports is still an invaluable tool in measuring surface behaviors produced by biological systems. Perkins et al. (2007) suggests that the use of questionnaires remain suitable and informative at these early stages of reinforcement sensitivity theorizing. One of the more recent purpose-built psychometric measurements constructed in the light of the most recent revision of the RST is Jackson's (2009) *Jackson-5 scale of Revised Reinforcement Sensitivity Theory*.

The personality construct is ubiquitous, and is thus relevant to a wide array of behaviors in a variety of contexts. The appreciation of personality has always been in behavioral terms, in such a way that if someone is identified to be strong in a particular trait, another may expect specific behavioral manifestations of that trait. The RST is no exemption. For example, Van Der Linden, Beckers, and Taris (2007) found out that work stress is partly determined by one's sensitivity to punishment (i.e., FFFS).

In order to advance revised RST, it is necessary to examine its usefulness and implications in other areas. To date, the revised RST has not been well advanced in terms of its impact to real-life contexts. One area that is not well explored is the implication of RST on behaviors in the context of education. There are limited studies that directly implicate RST to student performance, behaviors, motivation, or emotions. Those that are available were done in the context of the original RST, and not the revised theory. For example, in the study of Heimpel, Elliot, and Wood (2006), BIS positively correlated with avoidance achievement goals, while BAS negatively correlated to achievement goals.

Being a personality theory concerned with motivation, emotion, and learning (Smillie, 2008a), it would be interesting to know the extent to which reinforcement sensitivity determines education-related emotions. That is, can individual differences in responding to potential rewards, threats, and conflicts predict patterns in context-specific student emotions?

Among the different emotion variables educational psychologists are concerned about, one that has been extensively examined is test anxiety. In simple terms, test anxiety is defined as a kind of situation-specific trait anxiety that is experienced by students when they are subjected to evaluation, such as an examination, or test (Putwain, Woods, & Symes, 2010).

From a practical standpoint, interest in studying test anxiety has been focused on its implications to academic achievement of students, given that achievement is largely measured through their performance in tests. Several studies have identified small to moderate negative relationships between test anxiety and academic performance (Wong, 2008). This is important because it suggests that evaluation in school, through the use of tests, may not be reflective of the actual skills and knowledge of students. Moreover, since current convention in most educational systems is the reliance on evaluative tests, such may result to increase in anxiety, particularly test anxiety, among students (Casbarro; Milloy, Winans, Jehlen, Loschert, & O'Neil as cited in Sena, Lowe, & Lee,

2007). Methia (cited in Sena et al., 2007) reports that 33% of children and adolescents experience test anxiety.

Most researchers recognize the multidimensionality of test anxiety. Wong (2008) cites that different models of test anxiety have suggested that the construct is composed of cognitive, emotional/physiological, and behavioral components. From a psychometric perspective, popular measure of test anxiety, such as the Test Anxiety Inventory, focuses on two dimensions: worry and emotionality (Spielberger, 1980). Identification of these two components is based on the 1967 conceptualization of Liebert and Morris, where worry is referred to as a cognitive component about the consequences of failure in the test, while emotionality pertains to a host of physiological responses in wake of assessment and evaluation situations (Spielberger, 1980).

Current Study

The current study investigates the relationship between students' individual differences in their sensitivities to reinforcement and their test anxiety. In the context of schools, there are rewards and threats a student may perceive. Among these perceived rewards and threats are academic evaluations in the form of tests. In one hand, a test may be perceived as partly rewarding, since it is instrumental in obtaining good grades which is an outcome that can validate one's self worth. However, the nature of tests presupposes that the aforementioned rewarding scenario is inevitably tied with a threat - the prospect of the difficulties in reviewing, the possibility that the student may fail the test, or perform less than what he expects of himself. The typical response to the perception of this threatening aspect of tests is test anxiety. As cited by Putwain et al. (2010), Meijer and Zeidner emphasized fear of failure as an important characteristic of the construct, while Spielberger (1980) highlights negative social evaluation and threat to position. Moreover, arriving at the reward is only after the hurdles of reviewing for and being subjected to evaluation are dealt with.

Not everyone who takes tests experience test anxiety. While test anxiety may be prompted by evaluative situations, individuals vary in their responses. Some are more prone to test anxiety others are not. It is critical to identify those who are more likely to be test anxious as it was estimated by Schwazer (cited in Putwain et al., 2010) that two thirds of students who have low test anxiety would outperform an average high-test anxious student. Furthermore, it is of interest if RST, as a theory of personality, can predict test anxiety variations.

One possible predictor of anxious response to tests is perhaps an individual's dispositional sensitivity to threats, or threat-reward conflicts, as explained by RST. As mentioned earlier, tests may be perceived as rewarding and threatening at the same time. If that is the case, we may expect that sensitivities of neurological systems to such may be correlated to test anxiety. For one, the BIS, or the RST system that monitors conflict between incentives and threats are expected to correlate with test anxiety, since tests can be perceived as having both. Therefore:

(H1) BIS will correlate with test anxiety, and its subscales.

However, if one thinks of which among the reward or threat component is more salient in tests, as perceived by students, it is intuitive that they may lean towards the latter. After all, we rarely, if at all, hear students become excited over the prospect of taking a test. This leads to:

(H2) Test anxiety, and its subscales will have a stronger correlation with FFFS, than BIS.

(H3) BAS will not correlate with test anxiety and its subscales.

Since FFFS can be manifested in three ways (i.e., fight, flight, and freeze), we also take into account the ecological nature of tests in terms of the proximal-distal continuum, as well as the possibility of escape. It is difficult to define whether tests are proximal or distal, particularly since items in the measure of test anxiety used in this research draws from several temporal perspectives: before, during, and after the test. For certain, the items do not strictly refer to during-tests scenarios only. Similarly, the prospect of future tests is prevalent in school situations. Hence, tests may be construed more often than not, as distal threats. At the same time, tests are not necessarily escapable. There are typically no alternatives to it, except if the student chooses to simply ignore the test, or to walk out in the middle of it, which is rather unlikely. With these conditions, that is, tests are often regarded as distal threats, and that there is virtually no option to escape, it is predicted that:

(H4) Freeze will correlate more to test anxiety and its subscales compared to fight or flight.

Method

Research Design

This study is a cross-sectional, predictive study. Johnson (2001) describes this design as a study that attempts to determine capacity of a set of variables to predict other set of variables, using data from a single point in time. This method is deemed appropriate since the researcher intends to determine how sensitivity to reinforcement predicts students' level of test anxiety.

Participants

217 college level students, predominantly female (89.5%) adolescents ($M=18.5$, $SD=1.51$), from a university in Manila, participated in study.

Instruments

Informed consent was sought from the participants and two self-report questionnaires were administered.

Jackson-5 Scales of Revised Reinforcement Sensitivity Theory. A purpose-built measurement for reinforcement sensitivity composed of 5 scales (i.e., BAS, BIS, Fight, Flight, and Freeze). There is also a total score for FFFS, which is the sum of the Fight, Flight, and Freeze scales. Each scale is comprised of six items and are rated and scored using a five point scale: 1 = completely disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = completely agree. All scales have reasonable internal consistency with Cronbach's Alpha of 0.70 and higher (Jackson, 2009).

Test Anxiety Inventory (TAI). This instrument intends to measure individual differences in a situation-specific trait, test anxiety. There is a total score for Anxiety, but there are subscales for two components, Worry (TAI/W) and Emotionality (TAI/E). The instrument has 20 items that are rated using a four-point scale: 1=almost never; 2=sometimes; 3=often; 4=almost always. Item number 1 is reverse-scored. It is reported that the TAI has an Alpha of 0.92 or higher, while TAI/W and TAI/E subscales have median alphas of 0.88 and 0.90, respectively (Spielberger, 1980).

Data Analysis

The retrieved data were analyzed using SPSS. Pearson r correlation and multiple regression analysis were used to meet the objectives of the study.

Results

Descriptive and Correlation

Table 1 shows the zero order correlation matrix and descriptive statistics of the different variables. Results generally support the hypotheses stated above. As predicted in H3, BAS did not correlate with any of the text anxiety scales. On the other hand, BIS had low correlation with total test anxiety, as well as worry and emotionality, as predicted in H1 ($r = .227, p < .01$; $r = .199, p < .01$; $r = .206, p < .01$, respectively). Furthermore, as predicted in H2, FFFS had moderate correlation with total test anxiety, worry, and emotionality ($r = .457, p < .01$; $r = .407, p < .01$; $r = .412, p < .01$, respectively). As for the primary scales of FFFS (fight, flight, and freeze), it is freeze that consistently had the highest correlation coefficients, having moderate relationship with total test anxiety, worry, and emotionality ($r = .452, p < .01$; $r = .410, p < .01$; $r = .401, p < .01$, respectively), as predicted in H4. Other primary scales of FFFS also had low to moderate correlation with test anxiety and its subscales, with the exemption of fight and emotionality.

Table 1
Descriptive Statistics and Correlation

	M (SD)	Correlations (r)								
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 BAS	25.11 (2.83)	-								
2 BIS	23.32 (3.09)	.26**	-							
3 FFFS	58.16 (8.09)	.08	.31**	-						
4 FIG	18.93 (4.08)	.17*	.29**	.63**	-					
5 FLI	19.48 (3.91)	.01	.17	.76**	.12	-				
6 FRE	19.75 (3.54)	-.13	.25**	.74**	.12	.50**	-			
7 TA	44.66 (9.66)	.06	.23**	.46**	.20**	.33**	.45**	-		
8 WOR	18.02 (4.27)	.06	.20**	.41**	.22**	.24**	.41**	.91**	-	
9 EMO	17.73 (4.40)	.03	.21**	.41**	.12	.36**	.41**	.92**	.69**	-

Note: N=217; BAS = behavioural activation system; BIS = behavioural inhibition system; FFFS = fight-flight-freeze system; FIG = fight; FLI = flight; FRE = freeze; TA = Test Anxiety; WOR = Worry; EMO = Emotionality

Regression Analysis

Taking from the results of the correlation, subsequent simultaneous regression analyses were made among RST variables that showed relationship with test anxiety and its subscales. Initial regression analyses (not in the table) with test anxiety total, worry, and emotionality were made by entering BIS and FFFS (composite) as a block ($R^2 = .217$, $F [2,214] = 29.580$, $p < .001$; $R^2 = .171$, $F [2,214] = 22.148$, $p < .001$; $R^2 = .177$, $F [2,214] = 29.950$, $p < .001$, respectively). All three models were significant. Consistently, FFFS, rather than BIS, predicted test anxiety total ($B = .428$, $t [214] = 6.718$, $p < .01$), worry ($B = .382$, $t [214] = 5.837$, $p < .01$), and emotionality ($B = .385$, $t [214] = 5.901$, $p < .01$).

Tables 3, 4, and 5, provide more detailed models, wherein the correlated primary scales of FFFS are entered, instead of the FFFS composite.

Model 1: Test Anxiety and its Predictors

The first model tested for the collective and unique contributions of BIS, fight, flight, and freeze in predicting test anxiety as a whole, as presented in Table 3. It indicates that the model is significant ($R^2 = .245$, $F [4,212] = 17.224$, $p < .01$), and approximately explained 24.5% of the variance in text anxiety. However, if accounts of the individual variables are considered, freeze is the lone predictor of test anxiety ($B = .351$, $t [212] = 4.970$, $p < .01$).

Table 3
Summary of Regression Analysis for Model 1

	Unstandardized B	Std. Error	Beta
BIS	0.278	0.200	0.089
FIGHT	0.270	0.148	0.114
FLIGHT	0.327	0.171	0.132
FREEZE	0.958	0.193	0.351**

Model 2: Worry and its Predictors

Table 4 shows the multiple regression results for worry, wherein the predictors BIS, fight, flight, and freeze were entered as a block into the model. Analysis indicates that the model is significant ($R^2 = .202$, $F [4,212] = 13.440$, $p < .01$), explaining 20.2% of the variance of worry. It is interesting, however, that a slightly different pattern emerged as significant predictors of worry. The model suggests that worry is predicted by both fight ($B = .158$, $t [212] = 2.450$, $p < .05$) and freeze ($B = .359$, $t [212] = 4.939$, $p < .01$).

Model 3: Emotionality and its Predictors

As for the last model, which analyzes the prediction of emotionality, Table 5 shows that only BIS, flight, and freeze were entered since BAS, and fight did not correlate with emotionality.

Table 4
Summary of Regression Analysis for Model 2

	Unstandardized B	Std. Error	Beta
BIS	0.080	0.091	0.058
FIGHT	0.165	0.067	0.158*
FLIGHT	0.041	0.077	0.038
FREEZE	0.432	0.088	0.359**

Table 5 shows that, again, the model is significant ($R^2 = .208$, $F [3,213] = 18.619$, $p < .01$). The multiple regression suggests that the model accounts for 20.8% of the variance in emotionality. Among the three predictors in the block, it is flight ($\beta = .216$, $t [213] = 3.072$, $p < .01$) and freeze ($\beta = .265$, $t [213] = 3.667$, $p < .01$) that yield significant results.

Table 5
Summary of Regression Analysis for Model 3

	Unstandardized B	Std. Error	Beta
BIS	0.162	0.090	0.114
FLIGHT	0.243	0.079	0.216**
FREEZE	0.329	0.090	0.265**

Discussion

The aim of this study is to advance the revised RST by understanding its implications on test anxiety experienced by college students. Results support the status of BAS as a system that facilitates responses to rewards and incentives (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006) as evidenced by its lack of relationship with test anxiety. While tests may be instrumental in eventually obtaining rewards such as good grades and esteem, there is nothing inherently rewarding about being evaluated. On the contrary, society's emphasis on performing well in a test places students under stress and pressure in meeting these expectations (Casbarro; Milloy et al., as cited in Sena et al., 2007). Test anxiety is experienced in the context of threats that accompany evaluative situations; hence, reactivity to rewards, as indicated by BAS, is by no means associated with the experience of test anxiety.

Although there may be perceived threats in the case of evaluations or tests, these academic exercises are unavoidable. Meaning, the education system entails that it is necessary for students to go through such evaluations regularly. Here, the student is, in a way, obliged to approach the situation despite the perceived threats. Hence, it is not surprising the BIS somehow had a low correlation with test anxiety and its subscales. The BIS is identified as a conflict detection system, and facilitates the attempt to find solutions (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006). In this case, the conflict may be between the goal of taking a test,

as required by the educational system, and the pressure doing well enough so as not to be perceived as incompetent, or the sheer difficulty in preparing for and being subject to evaluation. Moreover, the BIS scale is said to measure sensitivity to uncertainty and social comparison (Jackson, 2009); that which is likewise highlighted as, in part, a determinant of test anxiety (Liebert & Morris, as cited by Spielberger, 1980). According to Mitchell, et al, (2007), those who are sensitive BIS are wary due to risk assessment.

Among the three RST systems, it is FFFS that had the most significant relationship with test anxiety and its subscales. When BIS and FFFS were entered into the regression model, FFFS also emerged as the significant predictor of total test anxiety. The FFFS is a system that facilitates responses to threats or punishing stimuli (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006). As mentioned earlier, threats in evaluation seem to overshadow the instrumentality of test in obtaining rewards and that these threats lead to test anxiety (Spielberger, 1980). It is because of this, that FFFS, indicative of sensitivity to threats, became a significant predictor of test anxiety. Findings, such as that of Van Der Linden et al. (2007) have linked sensitivity to punishment to other negative emotions, such as work stress. Similarly, Heimpel et al. (2006), found that sensitivity to punishment positively correlated with the adoption of maladaptive motivations, such as avoidance achievement goals.

Looking more closely, it is freeze, one of the primary scales of FFFS, that mostly accounts for the prediction of test anxiety and its components (i.e., worry and emotionality). Freezing, as a defensive behavior, happens in the context of threats that are perceived to be distal or non-desperate, yet, inescapable (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006). Since the taking tests does not necessarily lead to desperation and at the same time is often oriented in the future, while being inescapable for the most part, it is logical that anxiety responses to such is well accounted for by freezing. In the context of tests, freezing is likely to be observed as confusion in terms of which to prioritize in reviewing, procrastination, or the classic experience of having mental block as the test draws near, or even during the test itself.

Together with freezing, fight is also a significant predictor of worry, while flight is a predictor of emotionality, albeit to a lesser degree. As a component of test anxiety, worry pertains to cognitions about the untoward consequences of tests. On the other hand, emotionality refers to physiological responses, such as nervousness, that are reported when confronted with evaluation (Liebert and Morris, as cited by Spielberger, 1980). As stated above, it is not surprising that freezing is a typical response that predicts test anxiety and its subscales because of the ecological nature of tests (i.e., distal/non-desperate and mostly inescapable). But how is it possible that fight and flight responses are also related to test anxiety components, worry and emotionality, respectively? How does fighting or fleeing behaviors manifest in the context of tests?

The study shows that freeze and fight are significant predictors of worry. One commonality between freeze and fight responses is that they are done so when the threat is inescapable (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006). When a threat is inescapable, as it is in tests,

it is expected that a student would cognitively wallow about possible consequences, or perhaps even think about possible maneuvers to deal with the situation. Hence, worrying may serve as an attempt to deal with the situation and may be observed more among students who demonstrate freezing (described above) and fighting responses when faced with threats. Fighting or warding off threats that come along with evaluations may be in the form of actively confronting the situation; especially, among those who may perceive tests more desperately. Seriously preparing for exams, cramming a few minutes before the test, or employing focused determination during the test in order to succeed, are likely manifestations of this response.

On the other hand, freeze and flight responses are typical for threats that are distal or non-desperate (Pickering & Corr, 2008; Smillie, 2008a; Smillie, et al., 2006). The model indicates that freeze and fight responses are predictive of emotionality, or a set of physiological responses in the context of tests. If that is so, emotionality may be an experience that is more salient in situations outside the actual test, or most likely, during the anticipation of the test, or test results. Those who are likely to be nervous are those who display freezing (described above) and fleeing behaviors. Flight responses may be in the form of active attempts to ignore the prospect of taking tests, or ultimately, those who deliberately do not show up for the testing schedule.

Conclusion

As a whole, the RST maps out well with the construct of test anxiety and its subscales. While BIS and FFFS are related to test anxiety, it is FFFS, particularly freeze, which accounts much of the prediction of test anxiety. On the other hand, BAS is not related to test anxiety. These findings are in support of the revised RST. Test anxiety is, in a way, related to RST's conflict detection and resolution system (BIS) since tests are needed to be engaged with, despite possible threats that accompany it. The need to clarify the conventional thinking that BIS manifests as anxiety, and that FFFS manifests as fear, needs to be further clarified by future researchers, as findings reveal that test anxiety is more closely related to FFFS. Threats in the context of tests, that lead to test anxiety, are well accounted for by the significant contributions of the RST's threat response system (FFFS) into the models. Moreover, the finding provides evidence for the variations among fight, flight, and freeze responses, despite having similarity in general function (Eilam, as cited by Perkins et al., 2007), as indicated by the different models for total test anxiety, worry, and emotionality.

References

- Heimpel, S. A., Elliot, J. A., & Wood, J. V. (2006). Basic personality dispositions, self-esteem, and personal goals: An approach-avoidance analysis. *Journal of Personality*, 74(5), 1294-1320.
- Jackson, C. J. (2009). Jackson-5 scales of revised reinforcement sensitivity theory (RST) and their application of dysfunctional real world outcomes. *Journal of Research in Personality*, 43, 556-569.

- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30(2), 3-13.
- Mitchell, J. T., Kimbrel, N. A., Hundt, N. E., Cobb, A. R., Nelson-Gray, R. O., & Lootens, C. M. (2007). An analysis of Reinforcement Sensitivity Theory and the Five-Factor Model. *European Journal of Personality*, 21, 869-887.
- Perkins, A. M., Kemp, S. E., & Corr, P. J. (2007). Fear and anxiety as separable emotions: An investigation of the revised Reinforcement Sensitivity Theory of Personality. *Emotion*, 7(2), 252-261.
- Pickering A., & Corr P. (2008). J.A. Gray's reinforcement sensitivity theory (RST) of personality. In G. Boyle, G. Mathews, and D. Saklofske (Eds.) *The SAGE Handbook of Personality: Theory and Assessment Personality Measurement and Testing* (Volume 2, p. 239-255). London, New Delhi and Thousand Oaks: SAGE.
- Putwain, D. W., Woods, K. A., & Symes, W. (2010). Personal and situational predictors of test anxiety of students in post-compulsory education. *British Journal of Educational Psychology*, 80, 137-160.
- Schultz D. P., & Schultz S. E. (2005). *Theories of Personality* (8th ed.). USA: Wadsworth.
- Sena, J. D. W., Lowe, P. A., & Lee S. W. (2007). Significant predictors of test anxiety among students with and without learning disabilities. *Journal of Learning Disabilities*, 40(4), 360-376.
- Smillie, L. D. (2008a). What is reinforcement sensitivity? Neuroscience paradigms for approach-avoidance process theories of personality. *European Journal of Personality*, 22, 359-384.
- Smillie, L. D. (2008b). The conceptualization, measurement and scope of reinforcement sensitivity in the context of a neuroscience of personality. *European Journal of Personality*, 22, 411-425.
- Smillie, L. D., Pickering, A. D., & Jackson C. J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review*, 10(4), 320-335.
- Speilberger C. D. (1980). *Test Attitude Inventory, Preliminary Professional Manual*. USA : Mind Garder, Inc.
- van der Linden, D., Beckers, D. G. J., & Taris, T. W. (2007). Reinforcement sensitivity theory at work: Punishment sensitivity as a dispositional source of job-related stress. *European Journal of Personality*, 21, 889-909.
- Wong, S. S. (2008). The relations of cognitive triad, dysfunctional attitudes, automatic thoughts, and irrational beliefs with test anxiety. *Current Psychology*, 27, 177-191.

Correspondence may be sent to:
 renenob@yahoo.com



A Path Analysis of Inspiration, Purpose in Life, Gratitude and Subjective Well-Being among Filipino Workforce

Ricardo C. Lo

JM Elizalde Holdings Corporation, Philippines

Gabriel Ngombedua Makundu

Congregation of the Immaculate Heart of Mary, Taiwan

Abstract The relationship among inspiration and subjective well-being as mediated by purpose in life and gratitude has been postulated (Thrash & Elliot, 2010). In the present study, a sample of 254 Filipino employees was used to verify whether some types of inspiration influence purpose in life and gratitude. Satisfaction With Life Scale (SWLS), Inspiration Scale (IS), Purpose in Life Scale (PLS), and the Gratitude Questionnaire with Six Items (GQSI) were used to respectively assess subjective well-being, inspiration, purpose in life, and gratitude. The results of path analysis showed that purpose in life mediated the relationship under investigation but contrary to Thrash and Elliot's (2010) theory, gratitude did not.

Keywords: inspiration, subjective well-being, purpose in life, gratitude

The concept of subjective well-being has gained attention where different factors have been related to it. Most of these studies showed that subjective well-being is influenced by internal and external factors of the individual, with the preponderance of external factors (Lynbomirsky, Sheldon, & Schakade, 2005). Given this assumption, Thrash and Elliot (2010) investigated inspiration's influence on well-being. Inspiration was composed of three core characteristics: (1) *transcendence* that refers to a concrete and vivid awareness of better possibilities, the gained awareness resembling an illumination or insight in the vision metaphors; (2) *evocation* that refers to the fact that inspiration is experienced as evoked in the sense that the inspired person ascribes responsibility to something beyond the self; and, (3) *approach motivation* which translates the fact that one is compelled to actualize one's new idea or vision. The three core characteristics reflect the combination of two component processes of inspiration: (1) being inspired by; and, (2) being inspired to. The former involves the

appreciation of, and accommodation to the perceived intrinsic value of the evocative object, and the latter involves the motivation to transmit or actualize the qualities exemplified in the evocative object.

Thrash and Elliot (2010) were not the only ones who examined the relationship between inspiration and well-being. Other studies (Haidt & Keltner, 2004; Hymer, 1990; Kris, 1952; Maslow, 1968; Peterson & Seligman, 2004; Poloma & Pendleton, 1991) also pointed to this relationship. However, it was Thrash and Elliot's (2003) construct validation that operationalized both inspiration and well-being. From their construct validation, they theorized that inspiration influences subjective well-being through the mediation of purpose in life and gratitude. In other words, inspiration enhances purpose in life and gratitude, which, in turn promotes well-being. The concepts of purpose and gratitude have also been the subject of many researches.

Some researchers argue that purpose derives from connecting to something that transcends the self (e. g., Buber, 1996; Frankl, 1992; Nozick, 1989; Seligman, 2002). Others (Emmons, 1999; Klinger, 1977; Scheier et al., 2006) support the idea that purpose derives from pursuit of valued goals. McKnight and Kashdan (2009) argued that purpose offers direction in life and living according to one's purpose and offers a self-sustaining source of meaning through goal pursuit and goal attainment. This is corroborated by Burrow and Hill's (2011) study with adolescents and emerging adults. These authors found that greater purpose commitment mediates the association between identity commitment and well-being, supporting the assumption that affirming who one is may include establishing a particular direction in life, which may result in more selective and positive experiences. Another study by Ishida and Okada (2011) with university students comparing purpose in life to approval motivation indicated that purpose in life has a greater influence on emotional response and the autonomic nervous system response during psychological stress. The authors postulated that both purpose in life and approval motivation grow through experiences that include exposure to supportive and empathic understanding from parents and teachers during various developmental stages. Based on these studies, it appears that inspiration includes life with sense of purpose because it affords opportunities for transcendence of limits and for striving toward valued goals. This is further supported by Diener, Sapyta, and Suh (1998), attainment of one's life purpose influences a person's subjective well-being.

As for gratitude, it has been a subject of investigations by many researchers. McCullough, Kilpatrick, Emmons, and Larson (2001) defined it as a positive affective response to receiving a gift. They concluded their study with the observation that gratitude as a typical response to perceived benevolence from other moral agents fosters prosocial behavior among beneficiaries and benefactors alike and might have important interfaces with people's well-being and spirituality. In the same line, Emmons and Crumpler (2000) presented gratitude as an emotional state and an attitude toward life that is a source of human strength in enhancing one's personal and relational well-being. Their study mentioned an empirical experiment by Emmons (1999) that supported the supposition that gratitude is a source of human strength since participants in the gratitude condition reported more progress toward their goals than those in the non-gratitude condition.

A corroboration of this supposition came from an empirical study by Emmons and McCullough (2003) suggesting that a conscious focus on blessings may have emotional and interpersonal benefits. This study revealed that gratitude affects subjective life appraisals and appears to create increases in positive affect as well as reductions in negative affect. Solomon (1983) saw in gratitude an estimate of gain coupled with the judgment that someone else is responsible for that gain. Polak and McCullough (2006) argued that gratitude results from two cognitions: that one has obtained a positive outcome and that an external agent is responsible for it.

Since inspiration involves the gains associated with transcendence and approach motivation, as well as the perception that something beyond the self is causally responsible, inspiration leads to feelings of gratitude toward its source.

The studies about purpose and gratitude lend a support to Thrash and Elliot (2010) theory that inspiration influences well-being through the mediation of purpose in life and gratitude. Besides establishing the dual mediation in their model, these two authors also raised the question of whether some types of inspiration might produce more purpose in life and gratitude than others.

There seems to be a dearth of studies that investigated this “contextualized” aspect of the inspiration - subjective well-being relationship as mediated by purpose and gratitude. Thus in the present study, the authors contribute to the literature on the area by investigating whether or not the mediation of purpose in life and gratitude in the relation between inspiration and subjective well-being will remain significant regardless of the type of inspiration. This study conceptualized the model by Thrash and Elliot (2010) by considering the effects of inspiration provided by the superiors to their subordinates’ (employees) expressed sense of purpose and feeling of gratitude. In other words, the study intends to determine whether employees would report a significant sense of purpose in life and gratitude, which affect their subjective well-being as a function of being inspired by their immediate superiors in the company. Specifically, the researchers hypothesized that when employees are inspired to work by their immediate superiors in the company, they gain purpose in life and experience a sense of gratitude, which in turn will heighten their subjective well-being. Support for the hypothesis allows organizational analysis and interventions may be introduced in companies to ensure a happier and more effective workforce through the inspirations provided by their leaders.

Method

Research Design

The study used a cross-sectional explanatory design. Following Johnson (2001), the cross-sectional explanatory design is used when the primary objective of the research is to develop or test a theory about a phenomenon and to explain how the phenomenon operates by identifying the causal factors that produce change in it; and when the data collection is done at a single point in time. The present study was intended to test Thrash and Elliot’s (2010) theory about the mediating effect of purpose in life and gratitude in the relation inspiration - subjective well-being. The

theory was tested using a sample of Filipino workforce and the data were collected during a single period of time.

Respondents

The target population of the study was the Filipino workforce. To obtain the sample, the researchers made use of non-probability sampling method, specifically, the convenience sampling. The researchers resorted to this method due to time constraint. The researchers conducted the survey in the Company where one of them is presently employed.

The participants were from a retail company in Manila and Boracay from different levels of the organization (rank and file to managers). Of the 254 employees who participated in the survey, 126 were males, and 128 were females. There were 195 employees in the staff level while 59 were in the supervisory/managerial level. The average age for males and females were 27.2 and 29.7, respectively, with an overall mean age of 28.5.

Using Daniel Soper Statistics Calculator (a-priori sample size calculator for Structural Equation Models), the sample size of 254 yielded a statistical power of 99%.

Instruments

Each participant was asked to complete the Satisfaction with Life Scale, the Purpose in Life Questionnaire, Inspiration Scale, and the Gratitude Questionnaire-Six Items.

Subjective well-being. Subjective well-being was measured using the Satisfaction with Life Scale (SWLS). The SWLS consists of five items that are rated on a 1-7 scale (1- Strongly disagree; 2- Disagree; 3- Slightly disagree; 4- Neither agree nor disagree; 5- Slightly Agree; 6- Agree; 7 - Strongly agree). Because it asks sweeping questions, such as “The conditions of my life are excellent” and, “I am satisfied with my life”, the SWLS shows good stability over time (Biswar-Diener, 2010). Initial and subsequent studies have examined the internal consistency of the SWLS, and alpha-coefficients have repeatedly exceeded .80.

Inspiration. Inspiration was assessed using an adapted version of the Inspiration Scale (Thrash & Elliot, 2003), contextualized in a corporate setting. The inspiration scale consists of four items that are rated on a 1-5 scale (1 no; 2 yes, slightly; 3 yes, somewhat; 4 yes, very; and 5 yes, extremely). The inspiration items were as follows: “Have you felt inspired? (and if so, how inspired did you feel?)”; “Have you encountered or experienced anything that inspired you? (and if so, how inspiring was it?)”; “Have you been inspired to do something? (and if so, how inspired [to do something] were you?)”; and “Would you say that you experienced inspiration? (and if so, to what extent?).” To contextualize the instrument in a corporate setting, the phrase “by your immediate superior” was added at the end of questions 1, 2, and 4, and the phrase “in your work” was inserted in question 3.

Purpose in Life. Purpose was assessed using three items from the Purpose in Life (PIL) scale (Crumbaugh & Maholick, 1964). McGregor and Little’s (1998) factor analysis of the PIL items yielded distinct Meaning and Happiness factors. The items selected for this study were those that loaded most strongly on McGregor and Little’s meaning factor. The PIL items have a

bipolar response format and were rated on a 5-point scale (for example, in my life: 1 -“I have no goals or aims at all”, and 5- “I have very clear goals and aims”). Responses were summed to form a purpose index.

Gratitude. Gratitude was measured using the items grateful and thankful, which were drawn from an established measure (Emmons & McCullough, 2003). The Gratitude Questionnaire-Six Items were rated on a 1-7 scale. (1- Strongly disagree; 2- Disagree; 3- Slightly disagree; 4- Neither agree nor disagree; 5- Slightly agree; 6- Agree; 7 - Strongly agree). It includes questions such as: “I have so much in life to be thankful for” and, “I am grateful to a wide variety of people”. Cronbach’s alpha for the six-item totals have ranged from .76 to .84.

Data Analysis

To test the factorial validity of the four-factor model and to find out if the data will fit the model, a confirmatory factor analysis (CFA) was done. The CFA has to be conducted since the Inspiration Questionnaire was contextualized in a work setting.

In ascertaining whether purpose in life and gratitude mediate the effect of inspiration on subjective well-being, a path analysis was deemed appropriate. Inspiration is the independent variable, purpose in life and gratitude are the mediators, and subjective well-being is the dependent variable.

Results

Descriptive Statistics

Table 1

Descriptive Statistics and Cronbach’s Alpha for the Scores on Satisfaction with Life, Inspiration, Purpose in Life and Gratitude

	M	SD	Cronbach’s Alpha
Satisfaction in life	4.73	1.09	.80
Inspiration	3.32	.78	.77
Purpose in life	3.93	.68	.74
Gratitude	4.99	.85	.69

Table 1 shows the descriptive statistics for the different variables in the study. Gratitude and Satisfaction in life were both rated on a 1-7 scale while Inspiration and Purpose in life were both rated on a 1-5 scale. The results show that Gratitude has a higher mean than Satisfaction in life, and Purpose in life has a higher mean than Inspiration.

The Cronbach’s alpha of four instruments that were used in the study is also presented in table 1. Except for the Gratitude instrument, the obtained Cronbach’s alpha for the instruments used in the study were higher than .70 which indicates that the items for each instrument are within the acceptable range of consistency and that the items are correlated with each other. This signifies the acceptability and the reliability of the instrument.

Correlation

Table 2 shows the correlation matrix of all the variables in the study. The data in the correlation matrix confirm the results of past studies about the relationships of most of the variables. There were significant relationships between variables being studied, except for the relationship between inspiration and gratitude.

Table 2

Intercorrelations of the Scores on Satisfaction with Life, Inspiration, Purpose in Life and Gratitude

	(1)	(2)	(3)	(4)
1 Satisfaction in Life	-			
2 Inspiration	.14*	-		
3 Purpose in Life	.38**	.31**	-	
4 Gratitude	.24**	.06	.50**	-

Note: * $p < .05$; ** $p < .001$

Mediation Analysis

The main goal of the study was to test the hypothesis that when individuals are inspired to work by the Company and their immediate superior, the employees feel grateful and gains purpose in life, which in turn will heighten their subjective well-being. This effect was tested by doing a path analysis using subjective well-being as the dependent variable, inspiration as the independent variable, and purpose in life and gratitude as the mediators as shown in the path diagram below.

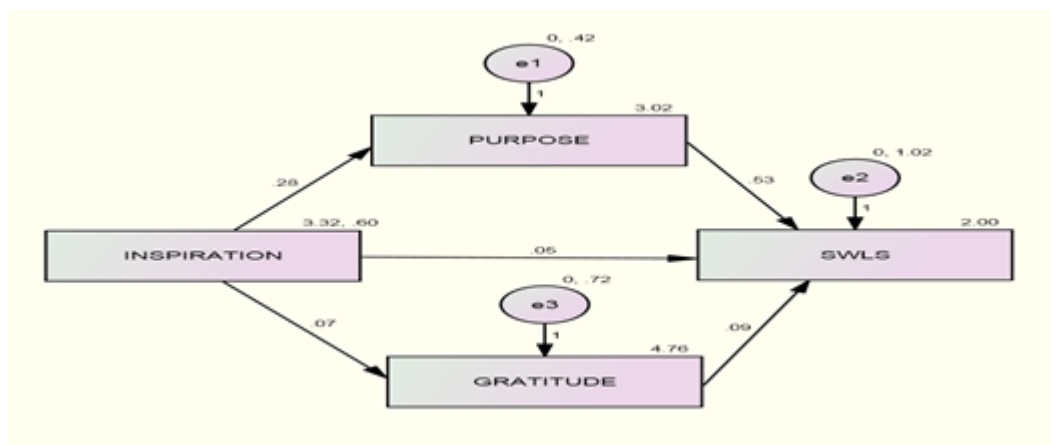


Figure 1. Mediation Path Diagram.

In the case of purpose in life as a mediator of inspiration and subjective well-being, the result of the path analysis shows that inspiration significantly predicts purpose in life (path coefficient = .28, $p < .001$) and purpose in life significantly predicts subjective well-being (path coefficient = .53, $p < .001$). This is consistent with the theory of Thrash and Elliot (2010) that inspiration promotes well-being because it enhances purpose in life, which in turn promotes well-being, and shows that purpose in life is a significant mediator of inspiration and subjective well-being. Furthermore,

since inspiration does not significantly predict subjective well-being (path coefficient = .05, $p > .05$), then purpose in life fully mediates the effect of inspiration on subjective well-being.

In the case of gratitude as a mediator of inspiration and subjective well-being, the result of the path analysis shows that inspiration does not significantly predict gratitude (path coefficient = .07, $p > .05$) and gratitude also does not significantly predict subjective well-being (path coefficient = .10, $p > .05$). This does not support the theory of Thrash and Elliot (2010) that inspiration promotes well-being because it enhances gratitude, which in turn promotes well-being. This shows that gratitude is not a significant mediator of inspiration and subjective well-being.

Although the results of the path analysis showed that purpose in life fully mediated the effect of inspiration on subjective well-being of the Filipino workforce, gratitude did not mediate the relationship between inspiration and subjective well-being.

Discussion

The results only confirmed one part of the researchers' hypothesis. That is, the effect of inspiration in the subjective well-being of the Filipino workforce, as significantly mediated by purpose in life. When inspired to work by the Company and their immediate superior, the employees gain purpose in life, which in turn heightens their subjective well-being. Thus, the

attainment of one's life purpose influences the subjective well-being. This is in agreement with Diener et al. (1998) who argued that purpose is a path to subjective well-being. Baumeister's (1991) study suggesting that purpose is a source of happiness and meaning also concurs with this finding. The finding is also in alignment with other studies showing that purpose is related to (Keyes, Shmotkin, & Ryff, 2002; Reker, Peacock, & Wong, 1987; Ryff, 1989; Scheier et al., 2006; Zika & Chamberlain, 1987, 1992) or predicts improvements in (Smith & Zautra, 2004) positive affect, negative affect, and other aspects of well-being.

The above results have implications in the world of work as having a workforce that is inspired and happy often translate into productivity; thus, companies should work towards having a more inspired employees by constantly looking into their job satisfaction and employee welfare. Factors like the work environment, corporate values, culture, stress management, conflict resolution, team building, and the like could be focused on via the different employee interventions. In addition, companies should attract, hire, train and or develop good leaders that will inspire the general workforce through leading by example and good work.

The researchers' other hypothesis is that purpose in life would function as a partial rather than a full mediator because it primarily concerns the transcendent and motivational aspects of inspiration, was disproved. Purpose in life fully mediated the effect of inspiration on subjective well-being among the Filipino workforce. Inspiration did not have a direct effect on the subjective- well-being of the Filipino workforce. According to Thrash and Elliot (2010), the fact that inspiration is ascribed to someone or something beyond the self suggests that gratitude functions as a

second partial mediator. The results of the present study, however, failed to prove that assumption. Gratitude did not mediate the effect of inspiration on the subjective well-being of the Filipino workforce. In other words, even when inspired to work by the Company and their immediate superior, Filipino employees seem not be grateful for it. More so, even if they feel grateful, it will not heighten their subjective well-being.

One possible reason why inspiration did not directly influence well-being was the contextualization of the instrument. An individual is “inspired by” and “inspired to” by several aspects of his life, including his family, and inspiration provided by his work, specifically his immediate superior, may comprise a “negligible” portion of inspiration that it does not influence the subjective well-being of the Filipino employees. The collectivist nature of Filipinos, especially within their group, may provide an inspiration that will influence their subjective well-being. Similarly, future studies may use the concept of family context as one source of inspiration.

References

- Baumeister, R. F. (1991). *Meanings of life*. New York, NY: Guilford Press.
- Biswar-Diener, R. (2010). *Practicing positive psychology coaching*. New Jersey: John Wiley and Sons Inc.
- Buber, M. (1996). *I and thou*. New York, NY: Simon & Schuster.
- Burrow, A. L., & Hill, P. L. (2011). Purpose as a form of identity capital for positive youth adjustment. *Developmental Psychology*, 47(4), 1196-1206.
- Crumbaugh, J. C., & Maholick, L. T. (1964). An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *Journal of Clinical Psychology*, 20, 200-207.
- Diener, E., Sapyta, J., & Suh, E. (1998). Subjective well-being is essential to well-being. *Psychological Inquiry*, 9, 33-37.
- Emmons, R. A. (1999). *The psychology of ultimate concerns: Motivation and spirituality in personality*. New York, NY: Guilford Press.
- Emmons, R. A., & Crumpler, C. A. (2000). Gratitude as human strength: Appraising the evidence. *Journal of Social and Clinical Psychology*, 19, 56-69.
- Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology*, 84(2), 377-389.
- Frankl, V. E. (1992). *Man's search for meaning* (4th ed.). Boston, MA: Beacon Press. (Original work published 1956).
- Haidt, J., & Keltner, D. (2004). *Appreciation of beauty and excellence: Character strengths and virtues*. Washington, DC: American Psychological Association.
- Hymer, S. (1990). On inspiration. *Psychotherapy Patient*, 6, 17-38.
- Ishida, R., & Okada, M. (2011). Factors influencing the development of “purpose in life” and its relationship to coping with mental stress. *Psychology*, 2(1), 29-34.
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30(2), 3-13.

- Keyes, C. L. M., Shmotkin, D., & Ryff, C. D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology*, 82, 1007-1022.
- Klinger, E. (1977). *Meaning and void: Inner experience and the incentives in people's lives*. Minneapolis: University of Minnesota Press.
- Kris, E. (1952). *Psychoanalytic explorations in art*. New York, NY: Schocken Books.
- Lynbomirsky, S., Sheldon, K. M., & Schakade, D. (2005). Pursuing happiness: the architecture of sustainable change. *Review of General Psychology*, 9, 111-131.
- Maslow, A. H. (1968). *Toward a psychology of being* (2nd ed.). New York, NY: Van Nostrand Reinhold.
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127, 249-266.
- McGregor, I., & Little, B. R. (1998). Personal projects, happiness, and meaning: On doing well and being yourself. *Journal of Personality and Social Psychology*, 74, 494-512.
- Mcknight, P. E., & Kashdan, T. B. (2009). Purpose in life as a system that creates and sustains health and well-being: An integrative, testable theory. *Review of General Psychology*, 13(3), 242-251.
- Nozick, R. (1989). *The examined life: Philosophical meditations*. New York, NY: Simon & Schuster.
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Washington, DC: American Psychological Association.
- Polak, E. L., & McCullough, M. E. (2006). Is gratitude an alternative to materialism? *Journal of Happiness Studies*, 7, 343-360.
- Poloma, M. M., & Pendleton, B. F. (1991). *Exploring neglected dimensions of religion in quality of life research*. Lampeter, Wales: Edwin Mellen Press.
- Reker, G. T., Peacock, E. J., & Wong, P. T. P. (1987). Meaning and purpose in life and well-being: a life-span perspective. *Journal of Gerontology*, 42, 44-49.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57, 1069-1081.
- Scheier, M. F., Wrosch, C., Baum, A., Cohen, S., Martire, L. M., Matthews, K. A., Schulz, R., & Zdaniuk, B. (2006). The life engagement test: assessing purpose in life. *Journal of Behavioral Medicine*, 29, 291-298.
- Seligman, M. E. P. (2002). *Authentic happiness*. New York, NY: Free Press.
- Smith, B. W., & Zautra, A. J. (2004). The role of purpose in life in recovery from knee surgery. *International Journal of Behavioral Medicine*, 11, 197-202.
- Solomon, R. C. (1983). *The passions*. Notre Dame, IN: University of Notre Dame Press.
- Thrash, T. M., & Elliot, A. J. (2003). Inspiration as a psychological construct. *Journal of Personality and Social Psychology*, 84, 871-889.

- Thrash, T. M., & Elliot, A. J. (2010). Inspiration and the promotion of well-being: Tests of causality and mediation. *Journal of Personality and Social Psychology*, 98(3), 488-506.
- Zika, S., & Chamberlain, K. (1987). Relation of hassles and personality to subjective well-being. *Journal of Personality and Social Psychology*, 53, 155-162.
- Zika, S., & Chamberlain, K. (1992). On the relation between meaning in life and psychological well-being. *British Journal of Psychology*, 83, 133-145.

Correspondence may be sent to:
ricky@budgetmart.com.ph or ngombedua@yahoo.com



Parental Involvement in Children's Assessment in Kindergarten

Joyce Ferro Orillosa

De La Salle University, Manila, Philippines

Carlo Magno

De La Salle University, Manila, Philippines

Abstract The present study described the various practices of parental involvement in the assessment of their children in kindergarten. More specifically, the present study looked into the role of parents in three stages of assessing their child: pre-assessment (before instruction), assessment during instruction, and post assessment in their children's schooling. There were 20 parents interviewed and their specific practices in getting involved in the child's tests, grading, and other forms of marking, and assessment were asked. It was found in the study that before assessment, the parents organizes review materials for the child by purchasing references, prepares reviews, and further inquiry with teachers. During assessment, the parents balance between being a mother and delivering the academic practices (oral questioning and revisiting previous assessment results). After the assessment, the parents praise and reward the child.

Keywords: assessment, parental involvement, kindergarten assessment, early grade assessment

Parents play an important role on their child's academic achievement. Once the child is placed in schools, the parents' role is deepened due to the various modes of assessments used by schools. Given this situation, parents must provide opportunities at home where children can further hone their skills to increase academic performance. This is especially true among parents who are full time in taking care of their own child. They are much concerned about the kind of assessment that their child will engage. The seriousness of schools in the conduct of assessment places the parents in a situation to mimic the kind of assessment done by schools. These practices need to be documented in order to situate parental involvement factors within a more specific context in education such as assessment.

The available literature on the role of parents in the assessment of their child can be explained in two areas. One is the rich reviews on parental involvement that specified ways that parents interact with schools and with their children to promote academic success (Hill & Taylor, 2004). Another is the concept of the Early Grade Learning Assessment (EGLA) that specifies appropriate levels and kinds of assessment for the early grade including kindergarten (Early Learning Standards Task Force, 2005).

Theories available on parental involvement explain that involvement goes on two levels: School-based involvement (Conners & Epstein, 1995; Epstein & Sanders, 2002) and home-based involvement (Seginer, 2006). Examples of school-based involvement are involvement in the parent-teacher conferences, volunteering in school activities, and joining school governance. On the other hand, home-based activities include assisting the child in school activities, facilitating reviews for the child, and assisting other parents. In a meta-analysis conducted by Hill and Tyson (2009), they reported that parental involvement showed to have a strong positive association with students' achievement across studies.

The concept of parental involvement was distinguished by Magno (2007) with parenting styles. Different factors of parenting characteristics were clustered using multidimensional scaling which distinguished parental involvement and parenting styles. The factors of parental involvement include communication, support, bonding, interaction, giving material rewards, monitoring, and affection. On the other hand, the factors that congested for parenting styles include authoritarian, psychological control, behavior control, autonomy, authoritative, and discipline. Given the set of factors under parental involvement, the variable is further described as specific socializing activities in which parents engage such as guiding the child in homework. Likewise parenting styles are further described as attitude towards the child that is manifested through communication and creation of an emotional climate.

The meta-analytic study by Hill and Tyson (2009) showed that parental involvement impacts the child's performance in the academic setting. In the same way, the clusters of parental involvement in the study by Magno (2007) provided the specific characteristics that parents show when getting involved in their child.

Typically, teachers in schools are the primary players seen as the key agents involved in the learning of children. A significant amount of research may suffice that there are actually concrete good results that when parents get involved, it also benefits the children's learning (Brinke, 2002). Parents' role in making appropriate and high-quality environment are important to be noted in order to assess learning in its totality.

Parents' involvement in children's early academic development is necessary to investigate in order to establish the influence of family in a child's academic life. Parents' involvement is malleable variable that can be changed. This malleability has been its distinct quality, which makes the parents' engagement more reasonable to be interrogated (Arnold, Zeljo, Doctoroff, & Ortiz, 2008).

In particular, the involvement of parents in evaluating child's academic performance has been studied by several scholars and found that it is actually associated to some positive outcomes. Most of these findings

are correlational, but there are also some causal relationships (see Anderson & Minke, 2007). Moreover, for preschoolers, increased parent school involvement and more active types of parent involvement were both associated with more positive development in all domains and greater mastery of early basic school skills in all subject areas (Marcon, 1999). In addition, parents' involvement contributes to the social and cognitive development in the process of evaluating learning outcomes. There is a superfluity of evidences that might prove parents' significance in making them involved in the assessment of education of children, and there must be strong positive outcomes that may be associated to such theoretical linkage, but the interrogation of such research field does not end here.

Many scholars somehow posit a different interpretation to parent's involvement, if not, little benefits are seen and sometimes they are even immeasurable. Fan (2001) makes a compelling case in which to present a critical way of understanding parent's involvement and bringing out the issue of definition and dimension of such theoretical construct. Brinke (2002) also asserts that there is somehow a lack of confidence to parents' capacity, in particular, their conception of high-quality environment for young children. There seems to be a question also of professional judgment and the validity of information that comes from them in assessing their child's behavior. Arnold et al. (2008), further argue that, there may be a wide belief on the importance of parents' involvement to early childhood learning, but there are quite few studies that deal, empirically, with evaluation of parents' involvement particularly relating on the phase of a child's emerging academic performance like pre-school and the like. Furthermore, Anderson and Minke (2007) claim that there is a variation among each family's level of resources when it comes to the compliance to teachers' expectations. It is necessary to take into consideration that the capacity of each family depending on their socio-economic status still is a factor in the degree of inducing formalized mechanisms of engagement to their child's learning development.

There might be issues surrounding the vitality of parent's involvement in early childhood learning, but it is important to consider that its positive effects are not to be easily dismissed. This significant linkage, indeed, is already in itself worthy of further scientific investigation because of its potential contribution to the literature and to the practice of early childhood development.

Parents become involved with their child's schooling in a variety of ways in the academic area. These manifestations of involvement involve the parents guiding the child in doing their homework, giving examples of exercises, and reviewing the child for a test (Tam & Chan, 2009). There is a pattern that the involvement is focused on assessing the child. It is noteworthy to look at this specific area in education where parents get involved.

The assessment of the child's academic performance in the early grades requires a holistic perspective. This reasoning necessitates further investigation of plethora of variables that may affect child's development.

Kindergarten Assessment

Pre-school programs are potential arenas that situate the assessment of early childhood learning. Kindergarten assessment, in particular, is important to look into because of some distinct manifestation of parents' engagement as to how they contribute to the process of the child's learning.

Kindergarten assessment processes are somewhat complex because it requires making priorities and decisions based on building knowledge, best practices, and designing a process that is feasible based on the availability of resources. Moreover, it is sophisticated because there are many interrelated decisions to be made about its purpose, focus, methods, and implementation. The Washington State Department even distinguished kindergarten assessment as an organized way to learn what children know and are able to do, including their disposition toward learning, when they enter kindergarten and possibly at other points in time (SRI International, 2008).

In assessing the competencies of children, it calls for a collection of information from multiple areas of development. Many research studies suggest that in the process of assessing what a child knows and can do when they are about to enter kindergarten, it is important to look at the multi-faceted and multiple ranges of skills or areas of learning. In other words, gathering information from multiple sources such as teachers, parents, and trained assessment specialists is recommended to understand better the young children's skills and competencies (Snow, 2011). Included in this source of assessment are the parents who are important participants in producing a valid and useful assessment. The parents act as informants about their child's skills and abilities and as recipients of assessment information as well.

Some forms of kindergarten assessment include measures of language and literacy, social, emotional, and general knowledge. In general, some kindergarten assessment are inclined and interested to provide information across multiple facets of learning (Fan, 2001).

The purpose of assessment in early childhood is to support learning. Assessment sets a definition about what a child can do and what he/she is ready to learn next (Shepard, Kagan, & Wurtz, 1998). It also becomes a guide and a turning point for teachers to improve their ineffective ways of teaching or improving the curriculum and classroom management to fully maximize a child's learning development. It is important to identify the child's special needs for many reasons such as developmental assessment, knowing if the child has special needs and if they are physically or mentally impaired.

Principles of Assessing Children

The National Association for the Education of Young Children (NAEYC) with the Division for Early Childhood (DEC), the Head Start Program Performance Standards, American Educational Research Association (AERA), and the National Association for School Psychologists have come up with six principles of assessing children in the early grade. These principles serve as a guide for teachers and other school personnel that directly conduct assessment for children. The principles include the purpose for conducting

assessment, the appropriate method, context, process, standards-based, and parents as partners.

Purpose. The purpose of assessing early childhood should “identify both the well-developed and the least well-developed skills of children so that beneficial learning experiences and teaching can be individually planned and carried out” (NAEYC, p. 7). The purpose for this level should be diagnostic in nature and the assessment results are used for further planning. The diagnosis should identify the interests, difficulties and strengths, changes overtime, and appropriate instruction.

Method. The assessment method appropriate for children is authentic and done through naturalistic observation. Assessment is closely integrated with instruction where the teacher observes children during instruction and class activities. The teacher uses portfolios, observations, anecdotal notes, and checklists. Wiggins (1998) described that tasks are authentic if (a) it replicates the ways in which the child’s knowledge and abilities are tested in real world situations, (b) the child uses knowledge and skills to solve unstructured problems, and solution involves more than following a set of rules, (c) the child has to carry out exploration and work within the discipline of the subject area, rather than restating what is already known, (d) the contexts involve specific situations that have particular constraints, purposes, and audiences, (e) the child is required to integrate all knowledge and skills needed, and (f) learning should be focused through cycles of performance-feedback-revision-performance.

Context. Assessment is conducted while instruction or class activities are conducted. Assessing takes place in the naturalistic setting while the child is learning (like when engaging in class activities such as singing, dancing, playing, writing, storytelling, and others). The naturally occurring behavior typically describes the child’s strengths and weaknesses.

Process. The process of assessment is conducted regularly over time to determine the progress and growth of the child. The process of assessment is individualized rather than comparative. Each performance of the child is a reference point to his/her progress overtime.

Standards-based. Assessment is closely integrated in the educational program of schools for preschool or K to 2. The assessment provides feedback whether the curriculum is appropriate or not. It also informs teachers about developing their instruction. The assessment and the curriculum should be aligned with some national or international standards.

Parents as partners. One of the unique principles of assessment in the early grades is the close involvement of parents in the activity. The assessment of the child does not only occur in school but also must be continuous at home. The parents also provide feedback for the child’s improvement. The teacher and the parents work together in order to help facilitate the child’s skills and improvement.

Assessment Domains

The child is assessed on three areas: Curriculum-based, early learning skills, and diagnosing disabilities.

Curriculum-based assessment measures the capacity of the child within the school program. The school program specifies some required competencies and the child is diagnosed on how much further instruction is

needed to develop the targets for the child. This typically includes assessing reading, writing, and mathematics skills.

The assessment of early learning skills includes an in-depth analysis of determining what the child can do at a specific level. By identifying the early learning skills of the child, programs are created to suit their learning needs.

The diagnosis of disabilities concerns in identifying some delayed skills of the child. The child is referred to specialists if difficulties are recurring.

Individual Approach in Assessment

Individual approach in assessment is important in assessing children and young people most especially in determining their special educational needs and/or disabilities. Each children's needs in the early grades are varied that needs to be addressed in the individual level. Individual pupil assessment is composed of using pupil records, classroom observations, interviews and quantifiable baseline data, and a collection of materials. An in-depth report is made with recommendations based on pupil assessment for schools, parents and other professionals. Personal education plans, individual education plans, behavior support plans, pastoral support plans, and risk assessments are made to suit individual pupil needs. A recommendation is given on how to implement the plans created.

Assessment for early childhood is done like the usual process. First, a teacher or whoever responsible for the assessment observes and takes evidences about the child. Second, they analyze and evaluate the evidences to check if there are some issues needed to be improved. Third, they have a plan for action with regard to the child's needs. The action plan should be congruent with what they found out during the second step. Last, they implement the action plan and check on the child's progress then go back to step 1.

Subject Area Assessment

Mathematics and Reading are subjects that are taught to students in order to develop certain skills and competencies of an individual. Mathematics is offered to the students who are at young age in order to speed up the cognitive development of the individual. According to Clements and Sarama (2008) that schools use a curriculum in order to effectively mold students to become good in mathematics. An example would be the building block curriculum where the implementation and facilitation uses materials and technological tools in order to teach the student mathematical concepts (Clements & Sarama, 2008). Students who were tested to be under the building blocks math curriculum were found to show positive effects in their performance in mathematics. Tests that are rooted under this building block curriculum are used in order to assess the mathematical abilities of the students. In a study by Xue and Meisells (2004), it was said that standardized tests are used to assess the performance of early childhood in the reading subject. The subject in math and reading are assessed based on their respective curriculum assessment tools.

The literacy skills of children can be measured using the (ELSA) Early Literacy Skills Assessment. ELSA assesses the comprehension, phonological awareness, and alphabetic principle of children. In order to execute this assessment, the teacher has to read a story to his/her students then he/she will have to stop in order to ask questions to them. The evaluation process would entail the students to undergo the same procedure again through the process of teacher reading the story again to his/her student after one year (Purpura, Hume, Sims, & Lonigan, 2011).

The Present Study

The present study described the various practices of parents in the assessment of their children in kindergarten. More specifically, the study looked into the role of parents in the pre, during, and post assessment practices in their children's schooling.

The study used qualitative approach where 20 sets of parents were interviewed. They have children in kindergarten level ages 5.5 to 6.5 who are in K to 12 schools. These parents that were interviewed did not employ any tutors for their children's academic needs. Both or one of the parents took the sole responsibility of teaching their children in all academic subjects.

Method

Participants

The study involved 20 parents who voluntarily participated in the study. The parents had children who are in the preschool level and at least one of the parents is full time (house wife or house husband) in assisting the child in school. All the 20 parents had either a junior or a senior kindergarten child who is currently studying in a private school in Metropolitan Manila.

Instruments

An interview guide was constructed to determine the general practices of parents in getting involved in their child's assessment. The parents were asked about the forms of assessment conducted in school, how they guide the child when studying specifically before exams, during periodic exams or major assessments, and after major assessment are conducted when the results are obtained.

Procedure

The parents in three private schools were recruited by the referral of school personnel. The school personnel referred the full time parents who regularly accompany their preschool child in school. These parents stay near the vicinity of the school while the child is attending his/her class. Both parents (mother and father) were requested for an interview. Once the parent agreed, an appointment was scheduled for the best time to conduct the interview.

The interview took about one hour for each parent couple. The interview session was recorded. The parents were informed about the

purpose of the study. At the end of the interview the parents were thanked for giving their time in participating in the interview.

Data Analysis

The responses from each of the parents were transcribed. Extracting the necessary information for each question asked among the parents further reduced the transcribed data. The specific practices that were taken from the transcriptions were coded. The codes were classified into meaningful categories that best explain assessment practices of parents. The categories were used to integrate the codes derived and they typically explain the assessment practices. The codes were shown to experts in parenting studies and qualitative analysis to make sure that they were mutually exclusive. They also arrived at an agreement on the placement of the codes on the categories derived.

Results

A total of 414 codes were identified, these codes were classified into meaningful categories to come up with the parental practices that reflects assessment of the kindergarten child's learning.

The responses were organized into the presentation of the various assessments used in the kindergarten level, the grading system, and role of parents before, during, and after assessment.

Forms of Assessment Used in Kindergarten

Formative Assessment. The first type used in the kindergarten level is informal assessment that includes the homework, exercises, worksheets, using the workbooks, pen and paper quizzes, and board works. Most of the homework were done on textbooks and worksheets prepared by the teacher that allowed children to master a particular skill. Others come in the form of a project that needs to be completed at home such as arts and craft. These forms of assessment are used to check up whether the child can perform the task while teaching is going on. In this level, it can be observed that parents can recognize how formative assessment is conducted in the classroom setting. The exercises and homework given to the child reflect formative assessment as part of the instruction. It serves to inform the teacher by adjusting teaching and learning while they are happening.

I can see at times that my child writes in the pad paper in doing exercises and gets a star when the work is well done. At certain times the teacher allows my child to redo the work to give chance for my child to perform better. I asked once that these exercises are not graded because they are meant to check if the child has acquired the skill.

Performance-based Assessment. The second type of assessment includes the performance tests like counting and reading, performance tests like reciting, writing and psychomotor activities that includes writing, singing and physical exercises. These assessments are mostly conducted one on one with children that focus on a single skill like reading or counting. This form of assessment may or may not be announced at times. This

category of assessment that the parents describe is performance-based. Parents describe this type of assessment as practical and test students' specific competencies that are not through the usual paper and pencil tests.

Some assessment that the teachers give are unannounced. This assessment involves recitations where my child is asked to read aloud words in a flash card or on the board. There are also times that the teacher lets them do exercises on their writing book. During the PE time, they are tested on their physical fitness like jumping and running. At times, they are asked to count from 1 to 10 individually.

Traditional Assessment. The last form of assessment is the paper and pen tests, which was the most common form of assessment. The paper and pen tests were given periodically and all of these tests were announced and scheduled in advance. Examples include quizzes, long tests, periodical, and quarterly tests. These tests are described by the parents as major types that the school puts greater weight. The pointers for review are provided and the teachers announce it two weeks in advance.

The serious tests are taken each month and before the quarter ends. These are the quarterly test and the monthly test for each month. We generally review the child on this because the teacher sends pointers for review and it is 40% of their grade.

Awareness of the School's Grading System

The grading system in schools would vary. Some schools would use numerical grades, others give letter grades, and others would only give a narrative feedback about the child.

Numerical Grade. For the assessments that provide an equivalent numerical value, the highest numerical grade a child can get is 100% and the passing grade is 75%. The number grade is generally the kind of grade understood by parents. They are more pressured for the child to get grades with 90 onwards because the numbers reflect the amount of achievement obtained by the child.

We help our child all the time especially on assignments and projects that are taken home. During quarterly tests, we really schedule the time of our child for review because this takes a large percent of the grades. We are concerned and do not want our child to get a line-of -seven grade.

Letter Grade. There are schools whose assessment used letters. Some would use "E" for excellent as the highest grade and the passing grade is "MS" for moderately satisfactory. Some schools would give A+ as the highest grade and C- as the passing. The parents would generally inquire about the status of the child to better understand the letter grades.

My child is new in the school and a letter grade is given. We are not accustomed so we ask the teachers what it means and the class standing of our child. We got used to the numerical grade in the previous school.

Anecdotal Records. Other schools would not give numerical nor

letter grade. They provide a narrative description of the child's performance. The feedback is supported with specific facts and anecdotal records. The assessment is not limited to the cognitive skills but there is equal reference to the social and emotional development of the child as well. This type of assessment is appreciated by the parents because they can see that the teacher prepared for it. The teachers support their narratives with the portfolios of the child to view their progress and the parents get an idea how the child is improving.

We listen to the teacher's narrative assessment. This is discussed with us and the teacher shows our child's portfolio. We can see in the portfolio how our child performs. What we appreciate is that the teacher includes feedback about how our child deals with classmates in the class. We get an idea if my child is shy or active in class. We appreciate this because the teacher spent a lot of time for the feedback.

Before Assessment

The parents before assessment practice the following: Self-reviewing, purchasing identical copies of books used in school, preparing home based academic activities, organizing academic requirements/school sheets, communicating with the teachers, and engaging in non-academic activities. These practices were done in preparation for their child's assessments.

One of the common practices is self-reviewing where the parents themselves study topics in advance to teach their children. This is practised so that the parents can pace the review with the child in an identical manner the way the teacher teaches. This is ensured to pace the level of learning for the child and to make the review suitable for the child. The parents are concerned about giving the same kind of review in school so that it will help the child understand the lesson taken up with to ensure mastery.

I myself review the lesson so that I will know what to teach my child. This helps me get an idea what is being taught in the classroom and I could follow-up the learning of my child.

Reviewing helps me know what kind of test I can prepare for my child.

Another common response on parental practice before assessment is the purchasing of identical copies of books and learning resources used in the school. This practice ensures that the kind of guidance that parents provide is at par with the standards of the lesson in school. Having the same kind of reference materials used in school allows the parents to provide a wider array of exercises that are not limited on the book that the child has.

I ask the teacher the other references that they use in school so that I can buy them. This helps me provide other exercises for my child to work on.

I buy the same kind of books that the teacher uses so that I will never ran out of sample work for my child to answer at home.

Majority of the parents also reports preparing home-based academic activities. Examples would be drills, exercises, and the kind of items in the book and previous tests. This activity allows the child to further master the skills taught in the classroom. Having the child answer a series of exercises practices the child to become more prepared for the tests. The constant exercise helps the child to be familiar with the kind of test items that will come out in the actual test. The parents use the pointers for review provided by the teacher to come up with their sample items.

We prepare a set of items similar with the test to review our child. We use the pointers to review as our guide.

The items we make are similar to the test given by the teacher so that the child will be familiar with the test.

In order to prepare for the assessment, the parents engage in organizing the academic requirements/school sheets. This includes the parents collecting all the previous assessment such as quarterly tests, long tests, quizzes, and worksheets. These are compiled so that the parents are guided to the test types provided and the kind of skills included in the test.

We collect the test folders and put in an envelope all quizzes and worksheets. This gives us an idea on what will come out in the test.

When it is time for us to make the reviewer for the child, we bring out the precious tests so that it can guide us. We also make our own outline of the lesson to guide the child.

The parents mentioned that communication with the teachers is an important part of preparation. In this practice, the parents have an opportunity to clarify some information with the teacher. The clarification involves asking the current class standing of the child, some behavior concerns, and the child's general learning capacity. Some schools would require an appointment and some would not. For other parents, this opportunity comes during the card distribution and the feedback is used to make some improvements for the following quarter of the school year.

We are lucky at times that we have some moment of conversation with the teacher. We ask about our child's current standing in school. We also ask if the teacher is having difficulty in handling our child.

Meeting with the teacher needs to be done through an appointment in the school of my child. In the meeting we discuss some of my concerns like performance in projects and tests and I ask for suggestions how to help further my child.

The parents also engage in non-academic activities in order to know about the school. By knowing more about the school, they get to know what learning facilities the school has and the resources used in the school. The parents indicated that learning about what the school has gives them idea what the child is learning and how to help the child further.

It is not only through the teacher that we get to know about what is happening in the school. When we join parent-teacher conferences, meetings, and volunteer activities, we see what the school can provide to our child.

By joining school activities we know what is happening in the school. Then we know more the schedule of our child in school and adjust our own schedule for it.

During Assessment

The parents during assessment practised multiple strategies, which were categorized into two: the academic orientation and motherly orientation. The academic orientation includes analyzing reviewers from school, inspecting previous notes from the teachers, creating new reviewers, revisiting previous tests/quizzes, and oral questioning. On the other hand, the motherly orientation includes reinforcing an early bed time, preparing good breakfast preferred by their child, allowing play time in between review sessions including the use of the computer, making promises for good performance, and preparing a conducive learning environment at home.

The academic orientations are those practices that are consistent with what the teacher and the school do in assessing the child. These are similar to the academic activities given by the school and the teacher.

The common academic orientation is analyzing reviewers from school. The reviewers are analyzed by looking at the formative assessment done in school. These are the quizzes about the topics, exercises from the books, and worksheets. The parents are informed what kind of items to create for their own review with the child. The parents in analyzing review materials are guided on what kind of items and type of test to make for their own review.

I look at the review materials that my child worked with in school. I use this to make my own review.

I make sure that the review materials I make are similar to what the teacher gave.

Part of the academic orientation is inspecting previous notes taken up in class. The inspection involves knowing what topics will be covered in the exams. The parents browse the books and the notebooks to know the lessons that will be covered in the tests. Sometimes, the sequence is not directly followed in the book and the parents will look at the pattern of the pages answered so that they will know what to follow up in the child.

To know the topics taken up, I browse the notebook and books. I will know what topics are covered for the exams. I mark the ones taken up in the previous test so that I am properly guided.

The notes and books tell me what topics will be taken up in the test.

The parents as part of the academic orientation create new reviewers for the child. Majority of the parents make their own exercises and test during the week of the test dates. They realize that when they practice their child, the child will come prepared for the real test in school. In this

aspect, the parents use their previous resources such as the notes, books and reviewers to write their own mock test for the child.

I make my own reviewer for my child. This is an effective way to prepare for the quarterly and long tests.

We make items similar to what the test will be. We are guided by the pointers to review and the kind of items in the quizzes.

The parents revisit previous tests/quizzes in order to review the child. The parents use the old test and quizzes and let the child answer them again. The old tests and quizzes review the child of the lessons that will be covered in the future test. This is usually done during the week the major tests are given. This technique ensures that the child is able to recall the topics taken up and that will be covered for the quarterly test.

We use the previous tests and quizzes to review our child. We rewrite the old tests and let the child answer them again.

The old tests are helpful. It guides us on what will be taken up for the quarterly test.

The parents during the test week engage the child in oral questioning. The oral questions are asked to review the child often. The questions are asked once in a while to check if the child can still recall the important things to be remembered. The oral questioning is helpful in different settings. Even if the child and the parents are outside of the house or doing other things, the oral questions allow to review the child.

We ask our child oral questions about what we have reviewed. This can be done repetitively once in a while to make sure the child does not forget what we have reviewed.

Even when we are in bed already or eating we ask our child questions about the lessons to review.

The next set of practices refers to motherly orientation. This refers to the practices that involve activities showing how the parents regularly socialize with the child. They differ with the academic orientation because these sets of practices with the child can only be done at home with the parents.

One of the common responses is reinforcing an early bedtime. This is done in order to make sure that the child has enough sleep and be able to wake up early in the morning. This strategy also ensures that the child will not be sleepy while answering the test in school.

We make sure that our child goes to bed early. This makes sure that he will not be sleepy in school.

The parents also prepare a good breakfast preferred by their child. This helps the child to be more motivated to take the test and perform well.

The child sees that the parents are exerting effort and appreciates the child. In return, the child needs to perform well for the parents.

I prepare a good breakfast that my child likes during the test week. My child shows their appreciation by performing well in the test.

I cook the favorite food of my child for breakfast so that they will be more energetic in taking the test.

The parents allow playtime in between review sessions. This includes the use of the computer. The child generally has limited attention span and the break gives time for the child to rest and have fun. The time of the child is scheduled in a way that there is time for play and time for studying.

I give break time for my child when studying. They will not enjoy studying too much if there is no fun.

I notice that my child becomes uneasy when reviewing for a long time so I give time for breaks.

The parents prepare a conducive learning environment at home for studying. This involves setting an area in the house for the child to study, provision of tables and chairs to study, maintaining a quiet place, and having school supplies in the house. Having a study place in the house allows the child to take time to study and not forget about school.

I make sure that my child has a place to study in the house. We set aside a place where they go and study.

My child has a corner where there is table and chair and this is the place where my child studies. During test weeks, we maintain silence in the house for the review.

The parents make promises to the child for good performance. This practice is done to reward the child and that the child will work hard in order to get the reward in the end. This practice is consistent with extrinsic motivation to learn where rewards are given to attain learning goals.

We promise to our child new toys when they get high grades in test or during the card distribution.

We often get our child something new when they good grades in school.

After Assessment

The parental involvement does not end after the major tests. There are still assessment practices that occur after. These practices after assessment involve asking for feedback about the performance in the test taken, praising the child, and giving the material rewards promised to the child. These after-assessment practices focus on parents praising and providing the material promises without the results of the assessments.

Upon the release of positive assessment results, parents still provide more material benefits.

The common response after assessment is the parents asking for feedback about the performance of the child in the test taken. The parents generally ask their child about the details of the test. They inquire about the level of difficulty, the type of test, and if the child did well in the test. These responses indicate that parents are conscious about test difficulty, types of test, and the confidence level of child in the test. Reporting the details of the test to the parents would require a great deal of ability for the child to make judgment about the tests difficulty and their performance. Children who makes accurate judgment of their performance tends to perform well on the test.

I ask my child how if he was able to answer correctly the test items and if he was able to answer most of the items or not.

I ask if the topics we reviewed came out in the test. This ensured the confidence of my child in my expertise as a parent.

Majority of the parents praised their child after taking the major assessment such as the quarterly test or any major performance-based activity. The parents made sure to capitalize on the strength and positive aspect of the child's performance. They made sure not to criticize the child so that the child is further motivated in future assessments. Performance-based assessment impacts parents' positive emotions. After the parents have seen their child in performances, the parents take pride in the talent shown by the child.

I make sure that my child gets to be praised after the exams. When I see the quarterly test, even if the scores are not very high my child deserves praise for the effort.

When I see my child joins contests such as poems, singing, and dancing, I feel proud that my child is not shy to showcase her talent. I praise my child for this.

The parents' after seeing the results of the child's performance, they give material rewards that they have promised. Initially, the parents promise these rewards to extrinsically motivate the child. After the performance, the child is given material rewards for the work done. The material rewards are given to the child to show appreciation of the hard work shown in the entire assessment cycle.

I give the materials that I have promised my child. This reward is not for the results of the test but for the effort shown by the child in studying.

I saw that my child becomes happy when they are bought new things after the test. When this happens, they look forward for the next test week and study again.

Discussion

A parent's involvement in their kindergarten child's assessment is important, considering that a parent has the ability to accommodate the cognitive, socio-emotional, and physical needs of their child during the early years. The parent's role in their child's kindergarten assessment is viewed as crucial in the development of academic skills. Providing these various needs, the academic achievement of a child increases and at the same time the relationship between parents and the child is nurtured. Moreover, when parents' capitalize on the moral development of their child in an attempt to increase academic performance, parents view this attempt as motivational in nature thus strengthening the motivation of their child towards learning. The findings in the study highlights that parental involvement in assessment is not only limited to a set of practices that parents engage with their preschool. The involvement in assessment extends to the consciousness of the parents in knowing the different forms and functions of assessment, the grading system, and practices before, during, and after assessment. This result contradicts the findings of Brinke (2002) that parents lack consciousness on the rigors of assessment in schools.

The results of the study highlight that parents when deeply involved in the child's assessment becomes conscious of the forms and functions of assessment. Generally, the ability to use different forms and functions of assessment is part of teacher competencies. However, in the present study, it was found that when parents are engaged in the child's assessment they also acquire this competency. This competency is not formally learned but through their experience in teaching their own child. The parents' acquiring the assessment skills leads them to acquire assessment literacies. Based on the responses, the parents possess three important assessment literacies that include (1) how to make use of a variety of assessment methods to gather evidence of the child's learning, (2) how to provide appropriate feedback to the child, and (3) engineering an effective assessment that boosts their child's motivation to learn. Possessing a certain degree of assessment literacy reflects that the parents that are highly involved with their child have a sophisticated notion on assessment and how it is conducted. This competency allows them to become effective in teaching and guiding their own child.

The results also highlight the function of the grades to determine how much effort is needed in enforcing the study skills of the child. This result is made known in the responses on their awareness of the grading system in the school. The parents work on the needed effort to help their child improve their grades and performance in school. For the parents, the grades are not simply marks given to the child. They inform parents on how much effort they need to help their child to perform better. This practice is supported in the principles of assessment 'for' learning. In this principle, assessment is used to help the child learn better. The parents show this by the specific practices mentioned in the second cluster: giving constructive feedback, inquiry from the teachers, and asking for recommendations from the teachers.

It was also found that though parental involvement, the parents are able to distinguish the different stages of assessment: before, during, and

after instruction. This sophisticated awareness in the assessment stages show the parents in their progression towards expertise in educating their child. Having distinguished manifestations for the stages of assessment allows the parents to pace themselves in the child's learning process. The child as an expert learner uses different learning strategies when reviewing for tests, taking the test, while assessment is going on and after assessment. The parents are able to adjust themselves also with the pace that the child goes through. The parents act as a scaffold in the child's preparation for assessment. As a scaffold, support is given in order for the child to demonstrate the acquired skills.

The parents assume two roles during the assessment stage: as a mother (motherly orientation) and as a teacher (academic orientation). The findings show that the parents are able to balance two orientations. They see to it that their academic guidance for the child is balanced with how they show their affection for the child. They balance the cognitive engagement of the child with the affect by becoming parents. This balance is uniquely done by the parents because they are able to blend two aspects that the child needs, which is the need for cognition and the need for affection.

The parents realize the value of extrinsic motivation for the child to perform. Extrinsic motivation is evidenced by giving material rewards for the child as the parents have promised. This practice is acceptable for parents because part of their role is to provide for the material needs of the child as well. It should also be emphasized that the material rewards are given not for the test outcome but rather for the effort that the child exerted in studying. The value of the material reward is taken by the child as something to look forward to in the next assessment cycle. In this way, the child enjoys engaging in the assessment process. The kind of willingness of the child to go through the extrinsic rewards manifests the inherent power of this type of motivation. The child is externally propelled into studying in order to obtain a high grade because of the reward that follows.

The findings further support the principles of early grade assessment where parents are partners in the child's assessment. This principle is further expanded in the present study where parents do not only provide support, but rather, they take a more active role in facilitating the assessment in schools. In this regard, parents must be given the opportunity to appreciate assessment as a tool for understanding their child's development. The parents have the potential to acquire assessment competencies and literacies. They are also adaptive in pacing themselves in the different stages of their child's assessment. They have acquired two areas of expertise in assessment, which made them not only as parents who provide emotional support, but in addressing the academic needs of their child as well.

References

- Anderson, K., & Minke, K. (2007). Parent involvement in education toward an understanding of parent's decision making *Journal of Educational Research*, 1(5), 311-323.

- Arnold, D., Zeljo, A., Doctoroff, G., & Ortiz, C. (2008). Parent involvement in preschool: predictors and the relation of involvement to preliteracy development. *School Psychology Review*, 37(1), 74-90
- Brinke, M. (2002). Involving parents in early childhood assessment: perspectives from an early intervention instructor. *Early Childhood Education Journal*, 29(4), 251-257.
- Clements, D. H., & Sarama, J. (2008). Focal points-Pre-K to Kindergarten. *Teaching Children Mathematics*, 14, 361-365.
- Conners, L. J., & Epstein, J. L. (1995). Parents and school partnerships. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. 4, pp. 437-458). Hillsdale, NJ: Erlbaum.
- Epstein, J. L., & Sanders, M. G. (2002). Family, school, and community partnerships. In M. H. Bornstein (Ed.), *Handbook of parenting: Practical issues in parenting* (pp. 507-437). Mahwah, NJ: Erlbaum.
- Early Learning Standards Task Force and Kindergarten Assessment Work Group (2005). *Early childhood assessment for children from birth to age 8*. Harrisburg, PA: Pennsylvania's Departments of Education and Public Welfare.
- Fan, X. (2001). Parental involvement and students' academic achievement: A growth modelling analysis. *The Journal of Experimental Education*, 70(1), 27-61.
- Hill, N. E., & Taylor, L. C. (2004). Parental school involvement and children's academic achievement: Pragmatics and issues. *Current Directions in Psychological Science*, 13, 161-164.
- Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology*, 45(3), 740-763.
- Magno, C. (2007). Exploratory and confirmatory factor analysis of parental closeness and multidimensional scaling with other parenting models. *The Guidance Journal*, 36(1), 63-87.
- Marcon, R. (1999). Positive relationships between parent school involvement and public school inner-city preschoolers' development and academic performance. *School Psychology Review*, 28(3), 395-412
- National Association for the Education of Young Children (NAEYC) & the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2003, November). *Position Statement: Early childhood curriculum, assessment, and program evaluation, building an effective, accountable system in programs for children birth through age 8*. Washington DC: NAEYC.
- Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, 110(4), 647-658.
- Seginer, R. (2006). Parents' educational involvement: A developmental ecological perspective. *Parenting: Science and Practice*, 6, 1-48.
- Shepard, L., Kagan, S. L., & Wurtz, E. O. (1998). *Principles and recommendations for early childhood assessments: Goal 1 early childhood assessments resource group*. Washington, DC: National Education Goals Panel.

- Snow, K. (2011). *Developing kindergarten readiness and other large-scale assessment systems*. NAEYC Center for Applied Research. Washington, D.C.
- SRI International (2008). *Kindergarten assessment process planning report*. Washington State Department of Early Learning.
- Tam, C. V., & Chan, M. R. (2009). Parental involvement in primary children's homework in Hong Kong. *The School Community Journal*, 19(2), 81-100.
- Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.
- Xue, Y., & Meisells, S. (2004). Early literacy instruction and learning in kindergarten: Evidence from the early childhood longitudinal study: Kindergarten Class of 1998-1999. *American Educational Research Journal*, 41(1), 191-229.

Correspondence may be sent to:
crlmgn@yahoo.com



Development and Standardization of the Anger Expression Type Indicator Test (AETIT)

Ryan Francis O. Cayubit

University of Santo Tomas, Philippines

Abstract As a scale development project, the present study intends to locally develop a standardized measure that focuses on the different modes of anger-expression. A total of 5,369 university students participated in this Four-stage project wherein data gathered were subjected to content validation, item analysis, reliability testing, exploratory factor analysis and confirmatory factor analysis. Based on the results, the EFA yielded three factors namely: anger manifest, anger suppress and anger manage. Based on the reliability testing and confirmatory factor analysis, the Anger Expression Type Indicator Test is a valid and reliable scale

Keywords: anger, anger-expression, anger manifest, anger suppress, anger manage

Emotion is said to be a universal part of man's life. Man is often viewed as bounded by emotions because his emotional reactions to events and things usually determine many of his actions and behaviors. Thus the study of human behavior may involve emotions. Emotion can also be described as something that is pleasant or unpleasant and the most common that we are familiar with would include joy, love, fear, jealousy, sadness and anger. Emotions are also considered as vast and complex and more often than not, some type of emotions are less popular than others. One such emotion is anger. One would rarely hear another say that "you have to learn how to be angry" or "it is okay to be angry" as often as he would say, "you have to learn to love" or "it is okay to be sad." Perhaps this is due to the fact that anger is one of the most misunderstood emotions and most often viewed as something negative (Villar, 1998). One is taught that it is all right to express anxiety, depression, moods or a whole lot of other emotions but not to express anger, at least in its original state.

Anger is a natural and normal emotion. Everyone experiences something that is not to his or her liking and the disappointment can be strong enough to create angry feelings. It also refers to an emotional state that varies in intensity from mild irritation to intense fury and rage (Spielberger, 1999) and may be accompanied by physiological and biological changes that is often caused by both external and internal events.

The instinctive and natural way to express anger is to respond aggressively. But experts believe that one can't just keep on physically lashing out at every person or object that irritates or annoys him as laws, social norms, and common sense place limits on how far one's anger could be expressed. This could probably have been the springboard of people who specializes in anger to seek for ways in which individuals can deal with anger in a constructive and appropriate manner. According to Spielberger (1999), anger-expression is conceptualized as having the following types: expressing, suppressing, and controlling (calming down). Expressing has been coined as one of the healthiest ways of dealing with anger because people are encouraged to express feelings of anger in an assertive manner rather than aggressive. Another type of anger-expression is suppression. This is done by keeping in the feelings of anger or ignoring or denying it. But doing such is not healthy as the anger could be turned and directed towards the self. Therefore, Spielberger (1999) recommends that it have to be converted and redirected. This is done by holding of one's anger, thinking about it and then focusing on something positive. The aim would be to convert it into a more constructive behavior. Last, is control a type of anger-expression wherein one could try to calm down or control feelings of anger both outward and inward by relaxing and allowing the feeling to subside.

Nature of Anger and its Expression

According to Spielberger (1999), anger as a psychobiological emotional state is generally accompanied by muscular tension and arousal of neuroendocrine and the autonomic nervous system. He theorized that in order to understand the phenomenon of anger, one could focus on one of its components: the state, trait, expression, and control of it. He referred to state anger as the psychobiological emotional state or condition marked by subjective feelings that may vary in intensity from mild irritation or annoyance or intense fury and rage. It basically describes the anger of an individual in a particular situation. Trait anger is defined in terms of individual differences in the disposition to perceive a wide range of situations as annoying or frustrating and by their tendency to respond to such situations with elevations. He also conceptualized that anger-expression and control could still be broken down to different components. The first component was labeled as anger expression-out, which involves the expression of anger toward other persons or objects within the immediate environment. The next component is anger expression-in, wherein anger is directed inward or towards the person experiencing the feeling. Another component is the anger control-out; this would be based on the control of angry feelings by preventing the expression of anger toward other persons or objects in the environment. The last component is anger control-in, which is related to the control of suppressed angry feelings by either calming down or cooling off when angered.

His work on anger-expression resembles that of Madow (1985) and McKellar (1985), cited by Burney and Kromrey (2001), according to them, anger can be expressed through the following: (a) modified expression on anger, these indicate anger that is fairly directed towards another person but in modified form; (b) indirect expression on anger, which, described anger as more hidden. It attempts to disguise the anger from the individual

and the perpetrator; and (c) variation on depression, these are expressions characterized by feelings of sadness and the blues. Likewise, it can be expressed through: overt, non-overt, and delayed. Overt anger would refer to an immediate and impulsive reaction to a stimulus event, as demonstrated by verbal and or physical aggression and is also one of the most common forms of expressed anger. Non-overt anger was defined as a passive expression of internalized emotions. While delayed anger was referred to as a planned aggression of an individual. He further stated that individuals engaging in delayed anger are often meticulous in planning an attack for revenge on the individuals or circumstances provoking the anger event. He also proposed that these three forms of anger may reflect behavioral responses that vary in frequency, intensity and rate of response.

Aside from looking at anger as a negative emotion, it can be best understood via the examination of its three main domains: cognitive, physiological and behavioral. Cognitive dispositions of anger would include knowledge structures, such as expectations and beliefs, and interpretation processes, which are organized in set patterns or schemas. Anger schemas would refer to cognitive structures that pertain to environment-behavior relationships entailing rules that govern responses to threatening situations. Physiological dispositions of anger would include high hormone level and low stimulus thresholds for the activation of arousal. Anger is marked by physiological activation in the cardiovascular, endocrine, and limbic systems, and by tension in the skeletal muscular. Behavior dispositions include variously learned repertoires of anger-expressive behavior, including aggression but also avoidance behavior (Burney & Kromrey, 2001; Deffenbacher, 1999; Robins & Novaco, 1999).

The social aspects of anger were also highlighted by Deffenbacher (1999). Research shows that anger can be viewed as something arising from the interactions among one or more eliciting events, an individual's pre-anger state, and the person's appraisal of the eliciting events and his coping resources. Anger has also been designed preserve personal worth, essential needs, and basic convictions. It operates on the premise that when an individual perceives rejection or invalidation, he or she feels that his or her dignity has been demeaned and therefore his or her personal worth is not recognized. They further stated that anger is kindled when one's essential needs, such as desire for appreciation of accomplishments, expression of interests, validation of feelings, are not responded to. Furthermore, they hypothesized that the absence of empathy or sentiment, or sensitivity to a person's most fundamental or basic convictions triggers anger.

Assessment of Anger

In 2001, Burney and Kromrey developed and validated an instrument designed to measure two distinctive types of anger: instrumental and reactive. The instrument is known as the Adolescent Anger Rating Scale. It was developed primarily, aside from measuring anger, to assist researchers and practitioners in identifying specific types of anger in adolescents. The scale is composed of forty-one items that measures Instrumental Anger, Reactive Anger, and Anger Control whose psychometric properties were established via internal consistency, item level statistics, test-retest reliability, content validity, criterion-related validity, and construct

validity. To serve as its framework, the researchers used two basic theories to explain the cause and construct of anger, which is cognitive and behavioral. The cognitive theory focuses on how an anger experience is appraised, what expectations are inherent in that appraisal, and the perceived attributions associated with that appraisal. By contrast, the behavioral theory suggests that when anger is experienced some external or internal behavioral responses may occur. The instrument makes use of T-score values in its interpretation making use of the following descriptive interpretation: very high level of anger, moderately high level of anger, average level of anger, moderately low level of anger, and low level of anger.

In 1999, Spielberger developed the State-Trait Anger Expression Inventory 2; this is the revised version of the original that was initially developed in 1988. The new version of the test was based on his extensive research that lasted for the past 10 years. The new version provides concise measures of the experience, expression, and control of anger. It was developed for two primary reasons: to assess components of anger for detailed evaluations of normal and abnormal personality, and to provide a means of measuring the contributions of various components of anger to the development of medical conditions, particularly hypertension, coronary heart disease, and cancer. Its development is rooted on the concept that anger has two major components: the state and the trait. He described in detail that state anger is a psychobiological state or condition marked by subjective feelings that vary in intensity from mild irritation or annoyance to intense fury and rage. Concerning the trait component, according to Spielberger (1999), it is the individual differences in one's dispositions to perceive a wide range of situations as annoying or frustrating and by the tendency to respond to such situations with elevations of state anger. The instrument is psychometric in nature. It consists of six scales with five subscales, and a total anger-expression index, which provides an overall measure of the expression and control of anger. The instrument is suitable for use for both adolescents and adults wherein an individual is to rate him or herself based on a four point scale that assesses either the intensity of their angry feelings at a particular time or how frequently anger is experienced, expressed, suppressed, or controlled. It is composed of 57 items that makes use of T-scores and/or percentile ranks and is interpreted as either high or low.

In 1998, Smith, Furlong, Bates, and Laughlin embarked on an extensive research that resulted in the development of the Multidimensional School Anger Inventory. It is a scale designed to measure affective, cognitive, and behavioral components of anger among the youth. The subscales include: Anger Experience, Cynical Attitudes and Destructive Expression. It is for the use of students attending middle and high school. The MSAI is largely patterned after two previously developed anger instrument, the School Anger Inventory (Smith, Adelman, Nelson, & Taylor, 1988) and the Children's Inventory of Anger (Nelson, Hart, & Finch, 1993).

Sharkin and Gelso (1991) also created the Anger Discomfort Scale, a 15 item self-report inventory designed to assess anger discomfort. The authors of the test conceptualized anger discomfort as an internal experience based on one's perception and attributions.

Siegel (1986) developed the Multidimensional Anger Inventory for the purpose of measuring the duration, frequency and magnitude of anger. It has ten factors, namely: frequency, duration, magnitude, anger-in, anger-out, guilt, brood, anger-discuss, hostile look and range of anger-eliciting situations. The scale was originally designed for adults but can be used with students in grades seven and up.

Another scale is the Novaco Anger Inventory. The inventory was originally developed in 1975 with a short form released in 1993. Both forms describe situations that are related to anger arousal. The long form is composed of 90 items while the short form has 25 items. The psychometric properties of both scales were established via convergent validity, test-retest reliability, factor analysis, and inter-item correlation, Cronbach's coefficient alpha and the split-half reliability.

Anger and Health

According to Spielberger (1999), there are research findings that indicate that anger and hostility contribute to the pathogenesis of many common medical disorders, including hypertension, coronary heart disease and cancer. In 2001, McDermott, Ramsay and Bray found out that expressed anger is a correlate of coronary artery disease after examining 153 patients. Similar results were obtained by Martin et al., (1999) and Siegman, Anderson, Herbst, Boyle and Wilkinson (1992).

Davidson, MacGregor, Stuhr, Dixon, and McLean (2000) found out that in 1,862 participants, constructive anger behavior-verbal is a significant predictor of lower resting blood pressure when controlling for the effects of standard hypertension factors. Anger was also seen as a correlate of high blood pressure and stress.

Anger and the Filipino Nature

Filipinos view and handle anger in a very unique way. More often than not, Filipinos are confused whether it is a good or bad thing. Religious and adult teachings and admonitions have led them to think that anger is not only bad but also sinful. Anger is also misunderstood in the area of its expression. Filipinos often do not know whether it is better to put anger under restraint or express it. According to Villar (1998) Filipinos are hesitant in expressing angry feelings that is based on maintaining and preserving smooth interpersonal relationships, which remains high in the value system of the Filipinos. She further stated that the common theme among Filipinos in relation to anger-expression is avoidance of feelings or controlling it and that they could be characterized as stuffers or passive aggressors in terms of anger-expression. This tendency could be attributed to the image consciousness spawned by their parents or the admonitions they got as children. Another source of non-assertion or difficulty in directly expressing anger is the feeling of fear or offending the other person.

This tendency is often exuberated in their behavior wherein they are often concern with what people will say or think about him or her when he or she is angry. She also said that aside from being stuffers and passive aggressors, Filipinos as a people often employs the technique of dropping anger and setting limits on things or situations that is viewed as something less personal. This could be the reason why "political sins" or "sins of

society” grip many Filipinos only when they are directly and personally affected by them or contaminated by a person who is highly influential to them. Likewise, not expressing ones anger is often viewed as a threat to ones reputation. Though there are those who choose to express what they really feel, majority of Filipinos remain avoidant in terms of expressing anger.

Daly (<http://www.newfilipina.com>, n.d.) appears to share the views of Villar (1998). According to her, it seems that majority of Filipinos are not comfortable with getting mad. Most Filipinos are especially uneasy with the act of confronting others with ones anger even if it is only about a small issue. This could be because the average Filipino family does not really teach their siblings that it was okay to even speak your mind to other people. Most Filipinos avoid, at all costs, arguments and confrontations even when they are in the right. Angry behaviors are shunned, looked down on and discouraged it in ones children. Most Filipinos have been taught that it is more polite to hold your tongue and contain your anger than to make a commotion.

The Present Study

Despite the existence of many instruments that measure anger and its mode of expression, majority of them was developed making use of Western samples thereby resulting to a question as to whether they would be appropriate for use for Non-western samples particularly Filipinos, who are known to be one of the happiest race on earth and whose notion and concept of anger and its expression appears to be different from that of others and because of this, I believe that developing a local scale that would measure anger-expression is imperative. I intend to use the concepts of emic and etic of cultural anthropology as a theoretical basis for creating the local scale because according to Pike (1954), behaviors and constructs that is inherent in man and is universally or widely accepted are known as etic while those that are culture-specific are known as emic and I believe anger-expression is an emic as it is largely influenced by culture.

Method

Participants

During the four-stage research project, a total of 5,369 university students participated in the study.

Table 1

Breakdown of Participants across the Stages of the Project

Stages of the Project	Total Number of Participants
Stage I	500
Stage II	1,278
Stage III	2,290
Stage IV	1,301

Procedure

Content Domains and Item Writing (Stage I)

The item writing for the Anger Expression Type Indicator Test (AETIT) involved two phases. For the first phase, I created three general questions, based on the theory of Spielberger (1999), which focused on how individuals would express anger. The three general questions made up the Opinion Questionnaire that was administered to the 500 participants. From their responses, I was able to create 138 items that made up the initial version of the AETIT. The 138 items focused on what an individual would do if he or she wants to express anger, hide it or control it and was subjected to content validation by five experts.

Pretesting or Initial Trial Run (Stage II)

The pretesting was done using different colleges and universities within the National Capital Region of the Philippines. The initial version of the AETIT was administered to 1,500 participants. However, not all data was used, as 222 of which was considered invalid. All data that was gathered were subjected to data analysis to establish the initial psychometric properties of the new scale. This resulted to a reduction of test items from 138 to 94.

Second Trial Run (Stage III)

The second version of the AETIT was again administered to 2,290 university students. The purpose of the second trial run was to further refine the new scale and to determine its initial reliability and validity via the examination of its factor structure. A total of 2,500 questionnaires were administered but only 2,290 were considered valid. Data gathered was subjected to factor analysis for data exploration.

Final Trial Run (Stage IV)

The items in the third version of the AETIT were reduced from 94 to 69. This was administered to 1,500 university students but only 1,301 were considered valid. The data gathered were further subjected to data analysis to examine its factor structure.

Measures

The Opinion Questionnaire. This is a researcher made instrument that contains three general questions designed to illicit responses related to how an individual would express, hide and control angry feelings. This was used in Stage I of the project.

Anger-Expression Type Indicator Test. This scale has three versions and was used from Stage II to IV of the research project. The first version of the scale has 132 items while the second version contained 94 items. The third version of the AETIT has 69 items. The purpose of the scale was to assess how Filipinos would express their angry feelings.

Data Analysis

Item Analysis. Item Discrimination (U-L Index Method) was computed to examine the ability of the new scale to discriminate between the high scorers and the low scorers. This was used during the second stage of the project.

Reliability. Item consistency was established via Cronbach's Alpha for all the three versions of the AETIT. The Inter-rater agreement or consistency was also computed based on the results of the expert validation.

Validity. Exploratory Factor Analysis was used in the Stage III of the project and Confirmatory Factor Analysis was used in Stage IV. A three-factor measurement model (based on the results of EFA) was constructed for the CFA with anger manifest, anger suppress and anger manage as latent variables. Items under each latent variable were parceled during data analysis.

Results

Reliability

Inter-rater agreement or concordance for the results of the expert content validation was computed in Stage I of the project, $r = .93$ and indication of high agreement among the five experts concerning the content domains of AETIT.

Cronbach's Alpha was used to determine AETIT's internal soundness or consistency. Alpha coefficient for Stage II is .91 indicating high reliability. For Stage III, Cronbach's Alpha was computed twice, one for the entire scale and the other for the different factors as revealed by the EFA. Just like in Stage II of the project, AETIT generated high alpha values an indication that it is a reliable instrument.

Table 2

Cronbach's Alpha obtained in Stage III

Scale	Cronbach's Alpha
Overall	.89
Anger Manifest	.93
Anger Suppress	.90
Anger Manage	.83

Index of Item Discrimination

The U-L Index Method was used to determine the ability of the test items in the initial version to discriminate between those who make use of a particular mode of anger-expression from those who do not. Data analysis shows that a total of 94 items (68%) out of 138 were judged as good items and was retained and made part of the second version of AETIT while 44 or 32% were discarded. In summary, majority of the items that made up the second version are considered good (64 items). See Table 3.

Table 3
Index of Item Discrimination

Item Number	Index of Item Discrimination	Description	Decision
30 Items: 49, 52, 58, 82, 85, 88, 91, 100, 115, 133, 136, 32, 44, 68, 77, 80, 83, 89, 92, 95, 101, 119, 131, 137, 48, 69, 84, 102, 108, 123 64 Items: 10, 13, 19, 22, 31, 34, 37, 40, 46, 55, 61, 64, 67, 70, 76, 79, 97, 103, 106, 109, 118, 5, 11, 17, 20, 23, 26, 29, 35, 38, 41, 47, 50, 53, 59, 65, 71, 86, 98, 104, 107, 110, 116, 122, 125, 128, 134, 6, 12, 39, 60, 66, 72, 75, 78, 93, 96, 99, 105, 111, 114, 126, 135, 138 30 Items: 1, 4, 7, 25, 28, 73, 94, 130, 2, 8, 14, 62, 74, 113, 9, 15, 18, 21, 27, 33, 36, 45, 51, 54, 57, 87, 90, 120, 129, 132 14 Items: 16, 43, 112, 121, 124, 127, 56, 3, 24, 30, 42, 63, 81, 117,	.40 and up	Very Good	Retain
	.30 to .39	Good	Retain
	.20 to .29	Marginal	Discard
	.19 and below	Poor	Discard

Exploratory Factor Analysis

Based on the EFA with Varimax Rotation, three factors were extracted and were labelled as anger manifest, anger suppress and anger manage respectively.

Factor 1: Anger Manifest: Composed of 36 items whose factor loading ranged from .30 to .72. This factor describes the behavior of an individual when he expresses angry feelings through verbally or physically aggressive actions.

Factor 2: Anger Suppress: Made up of 15 items with factor loadings from .31 to .71. This factor describes the behavior of an individual as he attempts to hide or suppress his angry feelings.

Factor 3: Anger Manage: Consist of 18 items with factor loadings from .31 to .60. This factor describes the behavior of an individual when he

attempts to manage feelings of anger by controlling its outward expression by calming down, cooling off, or relaxation.

Confirmatory Factor Analysis

A three-factor model of anger expression was tested via CFA. The latent variables were anger manifest, anger suppress and anger manage respectively with items that has been parcelled as manifest variables. 12 parcels were created for anger manifest, five for anger suppress and six for anger manage.

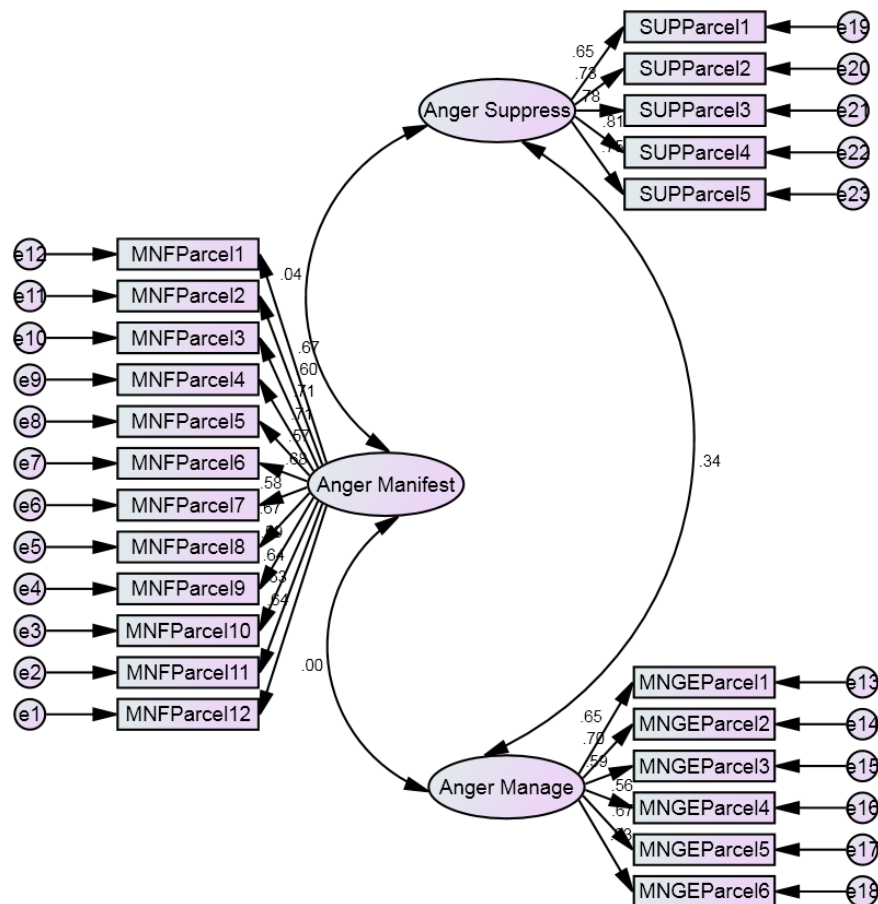


Figure 1. 3-Factor Model of Anger Expression

The following are the different fit indices obtained via CFA RMSEA = .054, RMR = .028, χ^2 (1,299) = 1095.33, NFI = .903, CFI = .922, GFI = .929, and AGFI = .913. The above results appear to support the 3-factor model. RMSEA and RMR values indicate a close and good fit. While the other indices did not meet the required cut-off of .95 and above but may still be judged as a fair fit considering its distance from .95 (Hu & Bentler, 1998). The χ^2

value however indicates a poor fit because of its significance but this could be attributed to larger number of respondents.

Discussion

This is a scale development project that spanned through four stages. The output is the Anger Expression Type Indicator Test (AETIT). A scale designed to assess and identify the modes of expression of anger among Filipinos. It has three factors namely: anger manifest, anger suppress and anger manage.

Results of the data analysis show that AETIT is both a reliable and valid instrument. Its reliability indicates that the scale can produce consistent results over time, which is particularly an important characteristic of a good psychological test. In terms of application, psychologists would now have an alternative assessment measure to look into the modes of anger expression especially since it has been established that anger is a direct correlate and predictor of heart disease, stress and hypertension. The scale can also be used as a preventive tool in schools and organizations where those who are prone to outwardly manifest their anger can be given the appropriate intervention.

Construct wise, AETIT appears to conform to the established theories of expressing anger despite its emic nature as evident by the three factors that was extracted. This shows that Filipinos also choose to outwardly express or suppress or control their anger similar to their western counterparts. The emic nature of the construct could be on the actual or specific ways of outwardly expressing, suppressing or controlling anger. However, the correlation of the three factors appears to be a concern as it ranged from negligible to low nonetheless when the overall fit of the model was examined, it can be concluded that construct validity was achieved as evident by the adequate to good fit of the various fit indices.

References

- Burney, D., & Kromrey, J. (2001). Initial development and score validation of the Adolescent Anger Rating Scale. *Journal of Educational and Psychological Measurement*, 61(3), 446-460.
- Daly, P. (n.d.). The anger ability and Filipino nature. Online [Available] <http://www.newfilipina.com/members/halo/malakas/articles/99.09/BP.M.M.Anger1.html>
- Davidson, K., MacGregor, M. Wm., Stuhr, J., Dixon, K., & MacLean, D. (2000). Constructive anger verbal behavior predicts blood pressure in a population-based sample. *Journal of Health Psychology*, 19(1), 55-64.
- Deffenbacher, J. (1999). Cognitive-behavioral conceptualization and treatment of anger. *Journal of Clinical Psychology: Psychotherapy in Practice*, 55(3), 295-309.
- Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.

- Martin, R. et al. (1999). Style of anger expression: Relation to expressivity, personality and health. *Personality and Social Psychology Bulletin*, 25, 1196-1207.
- McDermott, M.R., Ramsay, J.M., & Bray, C. (2001). Components of the anger-hostility complex as risk factors for coronary artery disease severity: A multi-measure study. *Journal of Health Psychology*, 6(3), 309-319.
- Nelson, W.M., Hart, K.J., & Finch, A.J. (1993). Anger in children: A cognitive behavioral view of the assessment therapy connection. *Journal of Rational-Emotive & Cognitive Behavior Therapy*, 11, 135-150.
- Novaco, R.W. (1975). *Anger control: The development and evaluation of an experimental treatment*. Lexington, MA: DC Health
- Pike, K. L. (1954). *Language in relation to a unified theory of the structure of human behavior* (2nd. ed.). The Hague: Mouton.
- Robins, S., & Novaco, R. (1999). Systems conceptualization and treatment of anger. *Journal of Clinical Psychology*, 55(3), 325-337.
- Sharkin, B.S. & Gelso, C.J. (1991). The Anger Discomfort Scale: Beginning reliability and validity data. *Measurement and Evaluation in Counseling and Development*, 24(2), 61-68.
- Siegel, J.M. (1986). The Multidimensional Anger Inventory. *Journal of Personality and Social Psychology*, 51(1), 191-200.
- Siegmán, A.W., Anderson, R., Herbst, J., Boyle, S., & Wilkinson, J. (1992). Dimensions of anger-hostility and cardiovascular reactivity in provoked and angered men. *Journal of Behavioral Medicine*, 15, 257-272.
- Smith, D.C., Adelman, H.S., Nelson, P., & Taylor, L. (1988). Anger, perceived control and social behavior among students with learning problems. *Journal of Child Psychology and Psychiatry*, 29, 517-522.
- Smith, D.C., Furlong, M.E., Bates, M., & Laughlin, J. (1998). Development of the Multidimensional School Anger Inventory. *Psychology in Schools*, 35, 1-15.
- Spielberger, C. (1999). *State-trait anger expression inventory-2: Professional manual*. Florida: Psychological Assessment Resources
- Villar, I. (1998). *Self-empowerment through anger and burnout management*. Manila: St. Scholastica's College

Correspondence may be sent to:

rocayubit@mn1.ust.edu.ph or ryanfranciscayubit@gmail.com



The use of Western Standardized Psychological Tests in Non-Western Contexts

Dylan Marshall

International School Manila, Philippines

Abstract The aim of this literature review is to examine the issues related to the use of standardized psychological tests developed in the West within non-Western contexts. The central thesis is that it is vital to be aware of the cultural differences and how these may affect performance on standardized psychological tests. Psychologists must be aware of this when they are assessing and evaluating students in schools in non-Western cultures. The review first presents a theoretical framework for understanding cultural differences in psychology. It then reviews researches that highlight the methodological and ethical controversies related to the use of standardized tests within non-Western contexts. The review then examines the steps that can be put in place to ensure equivalence when using standardized tests outside of the context that they were developed. This paper concludes by arguing that culturally relevant tests should be developed in a local context.

Keywords: culture, cultural bias, standardized testing, non-western contexts

Psychology has become a globalized discipline but it is important to understand that modern psychology has its roots in Europe and North America, and the majority of research still takes place on these two regions. For example, Smith and Bond (1998) carried out a review of psychology textbooks and found that only 10 per cent of the world's population was represented in samples used by psychological researchers. There is therefore still a Western dominance in psychological research and although this is gradually diminishing it is still a pertinent issue today. The problem arises when psychologists develop theories and models, which they claim to be 'universal' to all human behavior when they are only really applicable to the cultural context in which they were developed.

The cultural context of theories is highly relevant to the development and use of standardized psychological tests in schools. It is of utmost importance for school psychologists to be aware of this because when standardized psychological tests developed in Western countries are used

for evaluation purposes in non-Western countries, unfair and incorrect decisions could be made. In this review standardized tests are defined as tests that was designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent and applied in a predetermined, standard manner, likewise, it also involves making comparison of the test taker to a set of norms from a particular population (Popham, 1999).

Theoretical framework for the relation between culture and standardized testing

Culture is a very difficult concept to define. Matsumoto (2003), for example, mentions a book from 1998 that analyzed as many as 128 different definitions of culture. One of the most overarching definitions of culture was proposed by Herskovits (1948) who saw culture as the ‘human-made part of the environment’. Culture is a concept that is used to describe many different things, from food and eating habits, clothing and communication patterns to rituals and religion. Triandis (1990) claims that culture has an objective (physical materials) and subjective (cultural norms and beliefs) elements. According to Shaules (2007) the actual culture cannot be seen but it is only its manifestations that can be witnessed. In his view, there is ‘deep culture’ which is related to beliefs, attitudes, and values that underpin cultural manifestations. In addition, Hofstede (2001) describes culture as ‘mental software’, which are cultural schemas that have been internalized and has an influence on thinking, emotions and behavior. According to Hofstede (2001), members of a sociocultural group share this mental software and it is learned through daily interactions and feedback from other members of the group.

Given the definitions of culture, it is reasonable to come to a conclusion that culture affects the way we think and behave to a large extent. However, this factor has often been ignored by psychologists when they carry out their research, and continues to be ignored by psychologists when they administer standardized tests developed in the Western countries to participants from non-Western countries. Considering the issues related to this, test users should consider the conflict between cultural ‘absolutism’ and ‘relativism’. Cultural absolutism is associated with the ‘mainstream’ scientific psychology that has been conducted in North American and European universities during the twentieth century as such, it is no surprise that the major names of standardized tests such as Raven (1936); Otis and Lennon (1957), Flanagan (1960) and Cattell (1973) were all from this tradition while relativism is the approach which has been more central to the field of anthropology. This tradition tends to give more weight to social and cultural factors than to biological ones a direct reverse of the absolutists. Relativists have little interest in studying intergroup similarities while absolutists believe that there is a similarity between groups and in their research they look for ‘the psychic unity of mankind’ (Segall, Dasen, Berry, & Poortinga, 1999). When it comes to psychological testing relativists argue that it is impossible to have context -free categories of measurement. They avoid comparisons whereas absolutists attempt to use context-free measurements and standardized psychological instruments.

Since most of the psychologists who developed the standardized psychological tests were absolutists in their approach they have taken an 'etic' approach to understanding human behavior as opposed to an 'emic' one. Taking his cue from cultural anthropology, Pike (1954) made an important distinction between emic and etic. Etic are universal behaviors and constructs which all human societies have, such as marriage, concepts of intelligence, time, and the education of children. On the other hand, emic are culture-specific behaviors and constructs. This distinction is of importance when considering the development and use of standardized psychological tests. The Western psychologists who developed these tests thought they were measuring etic that are universal characteristics that are common to all human beings. However, when such tests were carried out in non-Western cultures, it became clear that different people approached problems in different ways, suggesting that they had developed a measure of their own culturally specific emic. The problem arises when these standardized measures developed in the West are directly applied to non-Western cultures. This known as 'imposed etic'.

Brislin (1993) clearly illustrates the problems that can arise from such an approach. For the concept of intelligence, he suggests that an etic definition of intelligence as the capacity to excel in 'solving problems, the exact form of which hasn't been done before' is a definition likely to be generally acceptable to different cultures. However, intelligence also has emic aspects, the emic of 'mental quickness' (for example measured by timed IQ tests) has been found not to be valid in all cultures. For example, for the Baganda people of Uganda, intelligence is associated with slow, careful and deliberate thought (Wober, 1974).

An example of 'imposed etic' in psychological research comes from Glick (1975) who wanted to measure intelligence in the Kpelle tribe from rural Liberia, West Africa. Glick asked participants to sort twenty objects into groups. They did this using 'functional' grouping (knife with orange, potato with hoe). The experimenter was expecting normal 'category' sorting (knife, hoe, potato, orange) and when challenged the participants said 'that was how the wise man would do it'. When asked how a fool would do it they sorted the objects into the categories the experimenter was originally looking for. These examples show how the use of a standardized test of intelligence such as the Raven's (1936) Progressive Matrices would not be a relevant or fair measure for the Baganda or Kpelle people, as the test is based on a Western emic concept of intelligence, which is radically different to the Baganda, or Kpelle's concept of intelligence.

It is also important to consider cultural differences when testing for personality. The most widely accepted model of personality is the Five Factor Model (FFM) developed by Costa and McCrae (1992) which consists of five broad domains of personality which are *openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*, assessed using the Revised NEO Personality Inventory (NEO-PI-R). McCrae and Terracciano (2005) reviewed research carried out in fifty cultures on six different continents and concluded that in most cultures around the world personality traits can be summarized by the FFM. This suggests that FFM is a universal etic construct, present in all cultures. McCrae (2004) states that "trait structure, age and gender differences, and cross-observer agreement are all universal"

(p.3) and Digman and Inouye (1986) argued that the FFM was applicable non-Western cultures such as Japan, China and the Philippines.

However, Guthrie and Bennett (1971) carried out research with Filipino participants and found profound differences in personality structure when comparing Filipino participants with American norms. The Filipino sample had similar conception of *extraversion* to the American sample, but some traits factored down differently to the American sample. For example, the factor of *emotional stability* was split into two factors in the Filipino sample, which were *worry/anxiety* and *somatic symptoms*. Furthermore, McCrae and Terracciano (2005) found that Filipinos, score low on *neuroticism* when compared to other cultures.

There is a debate over whether or not the FFM accurately captures the construct of personality in all cultures. Some alternative indigenous measures of personality have been developed in non-Western countries. For example the Chinese Personality Assessment Inventory (CAPI) includes a trait known as *interpersonal relatedness*, which is not included in the FFM (Funder, 2010). Katigbak, Church, Guanzon-Lapeña, Carlota, & del Pilar, (2002) carried out a study of personality in the Philippines using indigenous Filipino personality scales as well as the NEO-PI-R (Costa & McCrae, 1992). They found that there was an overlap between the Filipino scales and the FFM, but they also found indigenous factors such as *pagkamadalal* (social curiosity) and *pagkamapagsapalaran* (risk-taking), which do not feature in the FFM.

Even though there is a large amount of research, which supports the FFM's cross-cultural validity, research also highlights the existence of other personality traits apart from those included in the FFM. More research is needed in this field to further our understanding of personality across different cultures. Psychologists should be aware that the intelligence and personality tests they use might not be able to fully capture personality in non-Western contexts.

Issues related to the use of psychological tests in non-Western contexts

It is probably no surprise that intelligence, memory and critical thinking are influenced by the cultural context in which people live. Humans faced different challenges around the world in order to survive and this was particularly true before the modern era. Industrialization and the development of modern technology resulted to an increase in the need for people with specialized education and this continues up to the present time thus resulting to the adjustment of most of the education system in world in order to meet this need. Students now have to learn, understand and memorize, whether it be in mathematics, languages, or social studies. Students need to learn how to organize, assess and memorize information, and how to retrieve the appropriate knowledge from their memory whenever they need it.

Nevertheless, it is important for psychologists to recognize that this type of education is a component of one's culture and that culture varies in different countries and regions. Standardized psychological tests developed by Western psychologists are done so within a specific cultural context and to automatically apply them directly to non-Western samples can be misleading and problematic. When western researchers have used

standardized tests of cognitive abilities to participants from non-Western countries, almost all the time, the said participants performed poorly in many of the tests (Cole & Scribner, 1974). This clearly raises ethical and social implications since it can lead to people being judged as 'less able' or of 'low intelligence', not because they really are but because of the nature of the standardized test which they were asked to take. The 'local' validity of standardized tests, taking into consideration language and culture specific behavior, beliefs and attitudes, needs to be established first before psychologists use these tests in non-Western cultures.

The effect of culture on the assessment of cognitive abilities is clearly demonstrated by Cole's and Scribner's (1974) study, which investigated memory strategies in different cultures. The researchers compared the recall of a series of words for elementary school children from the USA and the Kpelle people of rural Liberia in West Africa. They understood that they could not use exactly the same list of words with the Liberian and American participants and in order to come up with a 'culture fair' test they observed the everyday cognitive activities of the Kpelle tribe in order to develop a memory tests that consisted of culturally relevant tasks. They made sure that the words used in their memory tests were familiar to the participants and collaborated with local college-educated people who took the role experimenters in the study. In spite of these precautions, their findings showed profound cultural differences in the way the Kpelle people remembered information and solved problems compared to the American participants.

In their first study the researchers asked Kpelle children from different age groups to recall as many items as possible from four categories: Utensils, Clothes, Tools and Vegetables. When researchers analyzed the data, they found that illiterate (e.g. non-schooled) children did not use strategies such as 'chunking'. Chunking is the grouping bits of information into larger units, to help them remember. The Kpelle did apply the mnemonic strategy of mental rehearsal, because the position of a word in the word list did not have an effect on the rate of recall. In their second study, the researchers changed the recall task so that the objects were now presented in a meaningful way in the form of a narrative story. They found that the illiterate (non-schooled) children recalled the objects easily and actually chunked them according to the roles they played in the story. This is of great significance to psychologists, as it highlights the importance of recognizing how culture influences performance on standardized tests of cognitive abilities. The example from Liberia suggests that it is vital for tests to be contextualized and made understandable to those who are taking the tests.

Further cross-cultural research comes from Kagan, Klein, Finley, Rogoff and Nolan (1978) who found that the memory abilities of Mayan children from Guatemala were significantly lower than their American counterparts based on standardized tests. However, Rogoff and Waddell (1982) decided to re-examine these findings in their study of cognitive abilities of Mayan children from Guatemala and American children from Salt Lake City. They aimed to show how placing a test of cognitive ability into the appropriate context could enable better performance. In their study each child watched as a local experimenter select 20 miniature objects from

a set of 80 objects and placed them in the 3D model diorama. The objects that were put in the diorama included animals, cars, people and furniture. The 20 objects were then returned to the group of 60 others on the table. The children were then asked to reconstruct the full scene they had been shown. They found that under these conditions, the memory performance of the Mayan children was slightly superior to their American counterparts. They concluded that children of any culture are skilled at remembering if information is presented in a meaningful context.

Critical thinking is another construct where cultural differences are present. Lun, Fischer and Ward (2010) analyzed differences in critical thinking between Asian and European students in New Zealand. They found that the European students performed better than the Asian students on the Halpern Critical Thinking Assessment and the Watson Glazer Critical Thinking Appraisal. The researchers concluded that these differences are due to Asian students relying more on *dialectical thinking* than their Western counterparts, due to differences in the way they process information. According to Nisbett, Peng, Choi, and Norenzayan (2001) Asians tend to 'perceive more changes, are more tolerant to contradictions and see things as more interrelated than westerners', they also concluded that westerners are more likely to use formal logical rules in reasoning and Asians use more experience based intuitive reasoning (Lun et al., 2010, p.605).

The above studies suggest that memory, intelligence and critical thinking have both etic and emic aspects. It is therefore vital for psychologists to understand this when administering standardized tests. Past researches demonstrates that people solve problems, learn and remember things in ways that are relevant to their own everyday lives, which do not always fit in with the way the concept is articulated in Western psychological tests.

Overcoming issues related to the use of psychological tests in non-Western contexts

When it comes to the use of standardized tests outside the context in which the test was developed, the fundamental issue is *equivalence*. If psychologists are to adapt or use standardized tests in different cultures fairly and accurately they need to ensure that translation equivalence, conceptual equivalence and metrical equivalence have been fulfilled (Brislin, 1993). It is also important to develop local norms for these standardized tests and to promote the development of locally made tests that are relevant to the test takers cultural background.

The first step in translation equivalence is to ensure that the concepts being measured in the items of the test are easily expressed in the language of the non-Western culture. If the test items do not translate well it suggests that the items are not relevant to that particular cultural context. A common method to ensure translation equivalence is *back-translation*. For this process to be carried out the original test items are translated from English into the target language by a bilingual person and a second bilingual person then translates it from the target language back into English. If the two English versions are equivalent it is assumed that the target language version is adequate.

Conceptual equivalence begins with the assumption that there will be different aspects of a concept, which serve the same purpose in different cultures. In order to ensure conceptual equivalence the etic and emic aspects of the constructs needs to be identified. For example, for the etic of 'intelligence' the emic might be 'solving problems the exact form of which have not been seen before' while the Western emic might include 'mental quickness' in the USA and Western Europe while it is 'slow careful and deliberate thought' for the Baganda people of Uganda (Wober, 1974). The different emic here are conceptually equivalent because they all form part of the definition of intelligence used in the two cultures. When it comes to standardized testing, in order to ensure conceptual equivalence both the emic and etic aspects of the construct need to be assessed for the cultural context for which the standardized test was created and also in the target culture in which the test will be used.

Metric equivalence focuses on the analysis of the same concept across cultures, based on the assumption that the same scale (after proper translation procedures have been carried out) can be used to measure it. For example, after back-translation, an IQ test produces a score of 120 for an American woman and a woman in the Philippines. The assumption is that the intelligence scale is measuring exactly the same concept (intelligence) in the two countries and that a score in one country can be directly compared with a score in another.

Of the three equivalences outlined above, conceptual equivalence is the most important criterion for standardized testing because of its theoretical impact. Another way to overcome these issues is to have local test constructors make their own standardized tests but then this approach limits the ability to compare the results with international norms. The issues outlined above raise the question of whether it is even possible to have 'culture fair' standardized tests. According to Frijda and Jahoda (1966) it is possible to have a culture-fair test only if it fulfills the following criteria: (i) if it has a set of items that are equally unfamiliar to all possible persons in all possible cultures, so that everyone would have an equal chance of passing (or failing) the items; or (ii) if it comprises of multiple sets of items, modified for use in each culture to ensure that each version of the test would contain the same degree of familiarity. This would give members of each culture the same chance of being successful with their respective versions of the test. Unfortunately, criteria (i) is virtually impossible and criteria (ii) is possible in theory but in practice very difficult to construct.

Conclusion

It is clear that there are real difficulties faced when using standardized tests developed in Western countries in non-Western cultures. Psychologists need to be aware of them and whenever Western standardized psychological tests are used, equivalence needs to be established first via the rigorous process of test validation. Once the tests are made as equivalent as possible it is crucial that local norms are established so that the results of one group of people are not unfairly compared to a different group that is from a different cultural setting. It is also vitally important to

create more culturally relevant tests in local contexts, which will be of practical use to psychologists in their assessments.

References

- Brislin, R.W. (1993). *Understanding culture's influence on behavior*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Cattell, P. (1973). *Measuring intelligence with culture fair tests*. Metro Manila: Philippine Psychological Corporation.
- Cole, M., & Scribner, S. (1974). *Culture and thought: A psychological introduction*. New York, NY: Wiley.
- Costa, P.T., & McCrae, R.R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology*, 50, 116-123.
- Flanagan, J. C. (1960). *Flanagan industrial tests*. Metro Manila: Science Research Associates.
- Frijda, N., & Jahoda, G. (1966). On the scope and methods of cross-cultural research. *International Journal of Psychology*, 1, 109-127.
- Funder, D. (2010). *The Personality Puzzle*. New York, NY: WW Norton & Company.
- Glick, J. (1975). Cognitive development in cross-cultural perspective. In F.D. Horowitz (Ed.), *Review of child development research* (pp. 595-654). Chicago, IL: University of Chicago Press.
- Guthrie, G. M., & Bennett, A. B., Jr. (1971). Cultural differences in implicit personality theory. *International Journal of Psychology*, 6, 305-312.
- Herskovits, M.J. (1948). *Man and his Works: The Science of Cultural Anthropology*. New York NY: Knopf.
- Hofstede, G. (2001) *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage Publications.
- Kagan, J., Klein, R. E., Finley, G. E., Rogoff, B., & Nolan, E. (1979). A cross-cultural study of cognitive development. *Monographs of the Society for Research in Child Development*, 44, 5.
- Katigbak, M. S., Church, A. T., Guanzon-Lapeña, M. A., Carlota, A. J., & del Pilar, G. H. (2002). Are indigenous personality dimensions culture specific? Philippine inventories and the five-factor model. *Journal of Personality and Social Psychology*, 82, 89-101.
- Lun, V. M. C., Fischer, R., & Ward, C. (2010). Exploring cultural differences in critical thinking: Is it about my thinking style or the language I speak? *Learning and Individual Differences*, 20, 604-616.
- Matsumoto, D. (2003). Cross-cultural research. In S. Davis (Ed.), *The handbook of research methods in experimental psychology* (pp. 189-208). Oxford, UK: Blackwell.
- McCrae, R. R. (2004). Human nature and culture: A trait perspective. *Journal of Research in Personality*, 38, 3-14.

- McCrae, R. R., & Terracciano, A. (2005). Universal features of personality trait terms from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology* 88, 29-44.
- Nisbett, R.E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 291-310.
- Otis, A. S., & Lennon, R. T. (1957). *Otis-Lennon mental ability test*. New York, NY: Harcourt Brace Jovanovich.
- Pike, K. L. (1954). *Language in relation to a unified theory of the structure of human behavior* (2nd. ed.). The Hague: Mouton.
- Popham, W.J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56, 8-15.
- Raven, J. (1936). *Ravens advanced progressive matrices*. Metro Manila: Philippine Psychological Corporation.
- Rogoff, B., & Waddell, K. J. (1982). Memory for information organized in a scene by children from two cultures. *Child Development*, 53, 1224-1228.
- Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1999). *Human behavior in global perspective* (2nd ed.). Boston, MA: Allyn & Bacon.
- Shaules, J. (2007). *Deep culture: The hidden challenges of global living*. Clevedon, UK: Multilingual Matters.
- Smith, P. B., & Bond, M. H. (1998). *Social psychology across cultures* (2nd ed.). Hemel Hempstead, UK: Prentice Hall.
- Triandis, H. C. (1990). Cross-cultural studies of individualism and collectivism. In J. Berman (Ed.), *Nebraska Symposium on Motivation*, 1989 (pp. 41 - 133). Lincoln, NE: University of Press of Nebraska.
- Wober, M. (1974). Towards an understanding of the Kiganda concept of intelligence. In J.W. Berry & P.R. Dasen (Eds.), *Culture and cognition*. London, UK: Methuen.

Correspondence may be sent to:
MarshallDy@ismanila.org



Determining the Construct Validity of a Critical Thinking Test

Marcos Y. Lopez

Centro Escolar University-Malolos, Philippines

Abstract This study deals with verbal reports of thinking in relation to determining the construct validity of each item of *The CEU-Lopez Critical Thinking Test* (Lopez, 2012) which is multi-aspect general knowledge critical thinking test designed for Filipino tertiary students. The procedure is based on the adapted methodology of Norris (1992) in validating his co-authored *Test on Appraising Observations* (Norris & King, 1983). Seven methodological phases were utilized in this study: determining the participants who would be part of sets of interviewees, adapting an interviewing methodology for eliciting trustworthy reports of thinking, collecting verbal reports of thinking, scoring verbal reports of thinking and choice of answer, comparing performance and thinking scores as basis for judging items, modifying suspected items, and retrying the revised or the replaced items in accord with steps 3 to 6. Items were revised, retained, and discarded based on the correlation of thinking and performance scores together with looking into the relevant insights of verbal reports of thinking given by examinees to establish construct validity of the test. It is concluded that verbal reports of thinking are useful in establishing the construct validity of a multi-aspect general knowledge critical thinking test.

Keywords: validity, test, construct, critical thinking, verbal reports

Norris (1992) said that evidence on the thinking processes of the examinees used to answer critical thinking test using multiple-choice format has a direct bearing on construct validity. Ebel and Frisbie (1991) defined construct validation as the process of gathering evidence to support the notion that a given test item really measures the psychological construct the test developers intend to measure. Alderson, Clapham, & Wall (1995) stated that construct validity is a form of test validation which fundamentally involves assessing to what extent each item in the test is a successful operationalization of the theory. Verbal reports of thinking can be

considered a relevant process to gather direct evidence that a test item is a successful operationalization of critical thinking construct.

The verbal reports of thinking collected for this study helped determine if a trial item measures what it is supposed to measure. If choosing the unkeyed option is associated with thinking well and choosing the keyed option is associated with thinking poorly, then the item is problematic. Conversely, if choosing the keyed option is associated with thinking well and choosing the unkeyed option is associated with poor thinking then the item is considered valid (Norris, 1992). However, these two premises are subject to empirical investigation by collecting a pool of verbal reports from a number of examinees for every test item of *The CEU-Lopez Critical Thinking Test*. Based on previous researches, verbal reports of thinking explicitly show how the test item works as intended by test developer.

This study examined the construct validity of the 87 trial items of *The CEU-Lopez Critical Thinking Test* (2012), a multi-aspect general knowledge critical thinking test measuring five aspects of critical thinking: deduction, credibility judgment, assumption identification, induction, and meaning. Specifically, verbal reports of thinking of the students were looked into to investigate the processes of thinking of examinees that led them to their chosen answers using Norris' (1988; 1989; 1990) argument that verbal reporting does not alter the course of thought and chosen answer of examinees while answering a multiple-choice critical thinking test.

Its main objective is to determine which item needs to be retained, revised, and discarded.

Method

Participants

A total of 2,412 students from all branches of Centro Escolar University (Malolos, Mendiola, Makati) took the experimental version of *The CEU-Lopez Critical Thinking Test* using paper-and-pencil format. Only students enrolled in courses with board examinations in all curricular year levels were considered part of research participants. A stratified random sampling was employed in which one intact class for every curricular year level of every course was randomly selected as samples.

Procedures and Data Analysis

This presents the methodology on how the construct validity of the 87-item test was established using verbal reports of thinking. The seven phases of research methodology were based on adapted procedure of Norris (1992) in determining the construct validity of his co-authored test entitled *Test on Appraising Observations* (Norris & King, 1983). These methodological phases were: determining the participants who would be part of sets of interviewees, adapting an interviewing methodology for eliciting trustworthy reports of thinking, collecting verbal reports of thinking, scoring verbal reports of thinking and choice of answer, comparing performance and thinking scores as basis for judging items, modifying suspected answers, retrying the revised or the replaced items in accord with steps 3-6.

Phase 1: Determining the participants who would be part of sets of interviewees. These students were informed two days before test administration in order to pre-condition them mentally and psychologically that the test is a preparation for their future board examinations. It was done as a motivation for them to take the test seriously so that the validity of test results would not be adversely affected

The entire test can be thematically subdivided into eight parts, wherein, one student in a group which composed of eight students was asked to verbalize his/her thoughts on the assigned number of items allocated to him/her. There are 13 groups included in this study. Hence, a total of 104 students who were chosen randomly from all CEU campuses participated in giving their verbal reports of thinking. Since these students came from different parts of the Philippines with different mental abilities, levels of maturity, interest, and culture, it is believed that they represented individuals who have different background beliefs, critical thinking sophistication, and intellectual ability that might have relevance to determine whether an item be retained or needs revision, modification, or even deletion based on their verbal reports of thinking.

Phase 2: Adapting an interviewing methodology for eliciting trustworthy reports of thinking. For interviewing methodology, the think aloud procedure in eliciting verbal reports of thinking was adopted. Norris (1990) described think aloud as a method for the elicitation of verbalized thoughts in which the subjects were instructed to verbalize all their thoughts as they answer the items assigned to them and to mark their answers on the provided scantron sheet.

Two stages of interview were adapted using standardized procedure espoused by Norris (1992). In the first stage, the interviewer informed the examinees of the main purpose of interviews and the procedure on how the interview would be conducted. The examinees were informed that their verbalized thoughts would be tape-recorded. Each examinee answered all items in the test but there were selected items that need to be answered through verbal reports of thinking.

In second stage, the interviewer asked examinees to say everything that goes on in their mind as they answer the item. The interviewer can only interrupt if there was ambiguous reference of demonstratives or third person pronouns as examinees explain their chosen answer. Interruptions can also be done to probe for obvious reading mistakes of the questions. No additional information provided to examinees if they asked for reasons or facts provided in the question. This was done to determine the clarity of the question and the possibility that the question is problematic.

All research interviewees were aware that their verbalized thoughts were tape-recorded.

Phase 3: Collecting verbal reports of thinking. Thirteen verbal reports of thinking were collected for each of the 87 items of the given test. Hence, a total of 1,131 verbal reports of thinking were collected excluding the other 260 verbal reports of thinking for the items that were revised for further validation.

Each examinee was assigned a certain number of items for him/her to verbalize his/her thoughts as the examinee worked through the items. The number of items

assigned to be answered verbally by each examinee was based on the logical break of the story line. An interview and paper-and-pencil formats were interlaced as a procedure to collect verbal reports of thinking.

Once the first student in a group is done with answering the first part of the test through thinking aloud the student would continue answering the rest of the items through paper-and-pencil format. The next student assigned to answer the second logical break of the story line of the test should finish answering the first part of the test through paper-and-pencil format and continue answering the next set of items through thinking aloud and would continue answering the rest of the parts through paper-and-pencil format. This procedure was done up to the last logical break of the set of items.

The interlacing of thinking aloud and paper-and-pencil formats was done to make it not too tiresome for every randomly selected student to answer the 87-item test through thinking aloud.

Phase 4: Scoring verbal reports of thinking and choice of answer. All tape-recorded verbal reports were transcribed verbatim and analyzed carefully. Two types of scores were assigned for each item: performance and thinking scores. These two types of scores were considered empirically and conceptually independent. Hence, it is possible that the student got the correct answer but gave an unjustifiable reason for the chosen answer. Conversely, student may arrive at the unkeyed option by giving justifiable reason.

The scoring allotted for performance score is 0 and 1. Zero for choosing the unkeyed option and a score of 1 for choosing the keyed option. The same scoring was allotted for thinking score: 0 and 1. Zero for a reason considered unjustifiable and a score of 1 for a justifiable reason. The justifiability of reasons was based on the approximation of the critical thinking principles and criteria from the critical thinking taxonomy of Ennis (1987,1996, 2011a) as used by the students when giving their verbal explanation for their chosen answer.

Performance score is based on the answer chosen from the three options provided in every item of the test whereas the thinking score is based on the justification of the student for his chosen answer.

Phase 5: Comparing performance and thinking scores as basis for judging items. The relationship between performance and thinking scores for each item across subjects sampled was determined as basis for retention, revision or replacement of item. The two scores were correlated through point biserial correlation. The relationship between two scores should be strong for each other as basis to consider that choosing the keyed answer is associated with thinking critically and choosing the unkeyed answer is associated with thinking uncritically. If found that there is a low or no correlation between performance and thinking scores, then it serves as basis for item revision or item replacement. This purports that the student chose the unkeyed answer by thinking well or student thinks uncritically yet arriving at the keyed answer. Conversely, if there is a correlation between performance and thinking scores, then it can serve as a basis that the item be retained. This implies

that the student chose the keyed answer by thinking critically by invoking critical thinking criteria or principles.

Phase 6: Modifying suspected items. Items with low and no correlation between performance and thinking scores were looked into as basis for item revision or item replacement. The contents of transcribed verbal reports served as a guide on how revisions should be done or whether an item is problematic and needs replacement. Revision of items usually includes changes in words used and addition, deletion, and changes of information both in item stem and given options of certain questions. Hence, thinking and performance correlations along with transcribed verbal reports of thinking were used as basis for retention, modification, or replacement of problematic items.

Phase 7: Retrying the revised or the replaced items in accord with steps 3-6. Verbal reports of thinking were collected, transcribed, and analyzed to examine the revised and replaced items. The same scoring procedure was done for performance and verbal reports of thinking for each modified and replaced items. Items were considered valid when there is high correlation between performance and thinking scores. Revised items that did not have correlation between performance and thinking scores were repeatedly modified and underwent comparison between performance and thinking scores till all revised and replaced items have high correlation regarding scores on thinking and performance.

Results

Proportion of Accepted Items

Out of 87 trial items, 59 items were accepted without revision and twenty eight items were accepted but with revision. No item was replaced with new one. The use of verbal reports of thinking in a multiple-choice type of critical thinking test determines whether items be retained, modified, and discarded. Hence, construct validity of the test items was established.

Performance and Thinking Scores of Selected Items

Results of the verbal reports of thinking from selected four items from *The CEU-Lopez Critical Thinking Test* are presented below:

Item 17 shown below is categorized as deduction in which post hoc fallacy is instantiated. The said item is in a context of debate in which the speaker is presenting his argument whether fraternity be banned or allowed in universities and colleges. The examinee decides whether the underlined statement follows necessarily from or contradicts the other statements given or neither.

Item 17: The speaker is Congressman Salisi who argues that: *“A month after the imposition of death penalty, a group of researchers from one neighboring Asian country conducted a study regarding its effectiveness. They found that the crime rate in their country drops by about two percent in a matter of 30 days immediately after its imposition. This only shows that death penalty is effective.”*

The following is a set of pairs of scores on performance and thinking of all students who verbalized their thoughts in item 17 : [(0,0) (1,1) (0,0) (0,0) (0,0) (0,0) (0,0) (0,0) (0,0) (0,0) (0,0) (1,1)]. Noticeably, the thirteen pairs of performance and thinking scores of the said item coincided. The computed point biserial correlation coefficient between performance and thinking scores for item 17 is $r = 1.0$ ($p < .01$). Hence, the item was retained as is. This purports that an item works as desired implying high construct validity.

Item 23 which is given below deals with credibility judgment. Two contradictory statements are presented and examinees should decide which statement is more credible.

Item 23:

- A. *Ace who was curiously looking at the unexplained movements in the nearby rice-field, said, “There is a pair of rats that hastily entered a small hole near that paddy.”*
- B. *Tying his shoelace, Oliver who was intermittently looking at the same thing, which Ace has seen also, said “That is just one rat and not a pair of rats that immediately entered a small hole near that paddy.”*
- C. *Neither statement is more believable*

The performance and thinking scores of students on this item is given below with seven mismatches between performance and thinking scores either combinations (0,1) or (1,0).

[(0,0) (0,1) (0,0) (0,1) (0,0) (0,1) (0,1) (0,1) (0,0) (1,0) (0,0) (1,1) (0,1)]

The computed point biserial correlation between performance and thinking scores for item 23 is $r = -.032$ (n.s.) which suggests that there is a negative and low correlation between performance and thinking scores. This negative and low correlation implies that the item is problematic, thus, needs revision and lacking of construct validity.

Item 67 is categorized as induction item in which each item focuses on explanatory conclusion, specifically causal claims. For every conclusion, examinees decide whether the given information supports the conclusion, goes against the conclusion, and neither.

67. *A group of people from a competing company that also manufactured the same formulation claimed that the ABC pellets dried up rice plants in one of the farms somewhere in the province of Batanes. They concluded that ABC pellets are not safe to rice plants.*

A set of accumulated pairs of performance and thinking scores is given below for item 67.

[(1,1) (0,0) (0,0) (0,0), (0,0) (1,1) (1,1) (1,1) (0,0) (1,1) (0,0) (1,1) (1,1)]

The computed point biserial correlation between performance and thinking scores for item 67 is $r = 1.0$ ($p < .01$) which implies that the item has construct validity. However, it does not mean that the item is free from any problems and does not need any modification. Item 67 exemplifies that even though a correlation is high, it possible that an item may need revision due to the use of some words in test item that lead examinees to easily guess the correct answer.

Item 87 is part of the last aspect of the test that deals with meaning and fallacies. There are three options to choose from after every stem of the item.

87. MARY: If it is not true that Filipinos do not work hard enough then it is false that our country will not become more progressive one. It is not the case that Filipinos do not work hard.

OLIVER: You keep on saying that our country will not become progressive one. Are you not having faith in Filipinos' determination and hard work to make this country a better one? Could you explain why you said so?

Pick the one best reason why some of this thinking is faulty.

A. Mary is offering a proposition she is arguing for as a reason for itself.

B. Oliver misdescribed and challenged the position of Mary.

C. Mary's and Oliver's conclusions do not necessarily follow from their respective reasons.

A set of accumulated pairs of scores between performance and thinking for this item is shown below.

[(1,0) (1,1) (1,1) (0,0) (0,0) (0,0) (1,0) (0,0) (1,1) (1,0) (1,1) (1,1) (1,0)]

The computed point biserial correlation between performance and thinking scores for this item is $r = .53$ ($p < .01$) which shows moderate correlation. As shown above, there are four pairs of scores that are mismatched.

Discussion

The construct validity of an item was established using verbal reports of thinking as espoused by Norris (1988,1990,1992). An item that was found to have construct validity through verbal reports of thinking is given below.

The following is the verbatim transcription of verbal reports of thinking of Student A with its English translation in parentheses originally given in Filipino or Taglish (combination of Tagalog and English) on item 17.

Student A: *Letter A kasi based on what the speaker said ahhh... yun talaga yung justification kung bakit yung ano po death penalty is effective. Kasi after nung inimpose yung death penalty in a matter of thirty days daw bumaba yung crime rate ng bansa. So clearly parang yun din yung, death penalty nga ang dahilan kung bakit ahhh... bumababa yung crime rate. I mean effective siya. Kasi nga after nung inimpose 'yon dun sa bansa, bumaba yung percentage ng crime rate nila. So letter A.*

(Letter A because based on what the speaker said that is the real justification why death penalty is effective. Because after the imposition of death penalty in a matter of thirty days crime rate of the country goes down. So clearly, death penalty is the reason of the decrease of crime rate, I mean that is effective. Because after its imposition the percentage of crime rate goes down. So letter A.)

Student A chose the unkeyed response by thinking poorly. He received zero score for both thinking and performance scores for he attributed the two percent decrease of crime rate to the imposition of death penalty. He did not think that there may be other plausible explanations that can be justifiably attributed to the decrease of crime rate.

Another student verbalized his thoughts in answering the said item. His answer is given below:

Student B: *Letter C kasi sa second statement ano parang inassume na nilayung two % na nag drop yung crime rate ay dahil yon sa imposition of death penalty. Walang sinabing statement dito para sa first statement na pag bumaba yung crime rate, ibig sabihin death penalty is effective. Saka pwede din na may ibang dahilan kung bakit bumaba yung death penalty in that month.*

(Letter C because in second statement they somewhat assumed that the 2 % decrease in crime rate is due to the imposition of death penalty. There is nothing given in the first statement that when crime rate goes down that means death penalty is effective. Besides, there may be other explanations why death penalty goes down in that month.)

Student B received a score for both thinking and performance. He arrived at the keyed response by thinking critically. His justification by saying that there may be other explanations as to the decrease of 2 % to the crime rate of death penalty shows that he could recognize post hoc fallacy in an argument though he did not use that Latin phrase during his verbalization of thoughts but his explanations captured the essence of the said fallacy. Hence, its construct validity was established.

Concerning item that was accepted but with revision is exemplified by item 23.

Student C verbalized his thoughts as he worked on item 23. His chosen answer and its explanations are given below:

Letter B. Parang ito ang mas believable. Kasi mas mababa yung pwesto ni Oliver so mas believable siya kesa kay Ace. Mas naniniwala ako sa kanya kasi nung nagta tie siya ng shoelace syempre parang nakayuko siya so parang mas kita niya yung rat na pumasok don sa small hole na sinasabi kesa kay Ace na nakatayo na di masyadong malinaw ang pagkakakita.

(Letter B. This one seems more believable. Because the body position of Oliver is lower than that of Ace. Oliver is more believable because he is tying his shoelace and his body position is stooped that makes him clearly see the rat that entered the small hole. Ace is less believable for he is in standing position that makes it difficult for him to see clearly the rats that entered the small hole.)

Based on the test manual, the keyed answer is A because the attention of Oliver is distracted due to tying his shoelace. Hence, student C received a zero score for performance but one score for thinking. His given reason is justified because he is implicitly invoking other critical thinking criterion on judging observation statement which is the condition of the observer must be conducive to observation. He is assuming that the one who is in a sitting position can better observe the rat in a paddy than a person who is in standing position. These verbalized thoughts call for revision of the item.

Furthermore, student D has the same chosen answer and very much the same explanation as regards item 23. His verbal report of thinking is given below.

Mas pinaniniwalaan ko yung letter B. Sabi niya isa lang daw yung rat na nakita niya. Mas naniniwala ako sa kanya kasi nung nagta-tie siya ng shoelace syempre parang nakayuko siya so parang mas kita niya yung rat na pumasok doon sa small hole na sinasabi kesa kay Ace na nakatayo siya. Though, nagta tie ng shoes si Oliver at divided ang attention niya, para sa akin mas na view niya ng maayos ang rat dahil halos ka level ng eyes niya ang paddy.

(I tend to consider letter B as more believable. He said that he just spotted one rat. He is more believable because when he is tying his shoelace his body is somewhat bent downward and this makes him view more clearly that there is just one rat that entered the small hole than Ace who is in standing position. Though, Oliver is tying his shoelace and has a divided attention, for me he really views the rat more clearly because his eyesight level is nearer the paddy than that of Ace.)

Similarly, student D received a zero score for performance and one score for thinking. His justification for his chosen answer is the same with that of student C. Though their chosen option does not agree with the keyed response, but their reason justifies their selected answer.

Hence, item 23 was revised by making Ace in sitting position which makes his body position similar to that of Oliver who is tying his shoelace. This makes Ace's and Oliver's body position quite similar and somewhat differ only in which one of them has a divided attention which makes Oliver statement less credible.

The revised version of this item makes option A the conclusive answer with justification that the attention of Oliver is distracted due to tying his shoelace whereas Ace attention is more focused than that of Oliver. Thus, Ace is considered more believable than Oliver.

High correlation between performance and thinking scores through verbal reports of thinking is not without a problem. It is possible to have a high correlation between performance and thinking scores in an item but a revision or modification of item is needed. An illustrative example is item 67 which is part of induction item of the said test. Each item in this section of the test focuses on explanatory conclusion, specifically, causal claims.

Student G chose letter C as her answer in item 67 which means that the information does not help us decide that ABC pellets lethally poison the rats. His detailed answer is given below.

Take note sabi dito “competing.” Sa pagkabasa ko pa lang sa word na competing hint nasa akin yon para letter C sagot ko. Kasi opinion yun ng competing company na manufacturer din ng same product. Syempre may bias don kasi may conflict of interest.

(Take note it says here “competing.” Just merely reading the word “competing” is enough hint for me to choose letter C as my answer. Because that is the opinion of the competing company that is a manufacturer also of the same product. For sure there is bias there because of conflict of interest.)

Student G chose the keyed answer by thinking critically because she invoked one critical thinking criterion in judging the credibility of information which is *conflict of interest*. His justification is tantamount to saying that suspension of judgment is needed for there is a presence of conflict of interest. However, he also implied that the item can be answered easily without reading the entire question because of the hint of the word “competing” as part of the question. With his reasonable justification, he received a score for both performance and thinking.

In the same way, student H has similar answer in the said item. His brief explanation and chosen answer is given below.

Obviously, the answer is letter C because of the word “competing”....A group of people from a competing company, from there you could say automatically that there is bias in it. So my answer is letter C which is neither.

In a like manner, student H chose the keyed answer by thinking well. He was able to justify his answer by stating the word bias in which the competing company tends to be biased because it also manufactures similar product. Like student G, this student received a score for both performance and thinking.

Despite the high correlation between performance and thinking scores of item 67, it was revised by changing the word *competing* into other term for the said word led examinees easily to determine the keyed correct answer.

Furthermore, it is possible that despite the moderate correlation between performance and thinking scores the item was retained as is because there was no clear indication from verbal reports of thinking of examinees that the item needs revision or replacement. This is instantiated by item 87 in which verbal reports of thinking are shown below.

In this item, student I chose letter B as his answer. He explained his justification with the following statements: *Letter B sagot ko. Kasi parang si Oliver di naniniwala sa sinabi ni Mary na hindi hardworking ang mga Filipino. Yun ang naisip ko.)*

(Letter B is my answer. It seems that Oliver does not believe in what Mary said that Filipinos are not hardworking. That is what I thought about.)

Giving a vague explanation to his chosen answer, student I was asked by the interviewer to take note and analyze the negative words in the propositions of Mary. After which, a follow-up question was asked to student I: *Why did you choose letter B which states that Oliver misdescribed and challenged the position of Mary?* Student I

simply replied, *Basta letter B sagot ko. Mahirap ang tanong di ko din alam explain. (I just simply answered letter B. The question is difficult. I don't know how to explain.)*

Student I was able to get the keyed correct answer but was not able to justify his chosen answer. Thus, he received a score of 1 in the performance but no score on thinking. His answer just shows that it is highly probable for an examinee to choose the keyed correct answer without any valid reason at all or by mere sheer guessing in a multiple-choice type of test which is one of the big drawbacks of this type of examination.

Student J answered the same number of test item. He said that: *Ang hirap nito. Ang sagot ko is letter A. Kaya hindi B ang sagot ko kasi si Oliver ang sinasabi niya hindi naman na misdescribe sa pagkakaintindi ko. Yung letter C di ko pinili because siguro yung conclusion ni Mary hindi nag match sa reason niya pero yung kay Oliver parang malinaw naman ang statements niya. So ang parang sagot ko dito, yung kay Mary lang ang parang mali, parang ganon. Yung statement dito sa question na to, napakadaming negative words. Kaya lalo nagiging complex.*

(This is a tough question. My answer is letter A. I did not choose B as my answer because Oliver does not misdescribe anything. That is how I understood it. I did not choose letter C because the conclusion of Mary does not match her given reason whereas the given statements of Oliver are seemingly clear. So it looks that it is only the given statements of Mary that are wrong. The statements in this question are loaded with negative words. That makes the question more complex.)

Student J chose the unkeyed option by thinking poorly. He arrived at a conclusion that the answer is letter A without thorough analysis of the propositions given by two characters in the test. Students I and J must have explained that Oliver incorrectly attributed a conclusion which can serve as a position to the argument of Mary. The stated conclusion of Mary is "our country will become more progressive one" which is contrary to attributed conclusion of Oliver to argument of Mary. The wrongly attributed conclusion of the former is challenged by himself. This is a case of straw-person fallacy. Furthermore, the double negation in the given argument of Mary makes a positive. Both A and C are false.

In spite of the comment of student J that item 87 is difficult due to compounding of negative words, the said item was not revised for Ennis (1987) said that some sophistication in dealing with negation is necessary in the teaching and testing of critical thinking.

It is interesting to note that though correlation is moderate between performance and thinking scores does not necessarily mean that item 87 needs revision. If the mismatch of scores is due to the performance scores outweigh thinking scores then it may mean that the item can be retained as is because there is no basis for the revision of the item that can be inferred from the verbal reports of thinking. However, if the mismatch of scores is due to thinking scores outweigh the performance scores, then, the item needs revision or may even be basis for item rejection. In case of the latter, revisions can be made in the item stem or the keyed correct answer. The insights from the verbalized thoughts of the examinees are strong basis for item revision for they may lead test developers how the items should be revised.

Verbal reports of thinking are relevant in establishing the construct validity of critical thinking test items. Norris (1989) used this procedure in validating *Test on Appraising Observations* and recommended that verbal reports could be explored in determining the construct validity other than single-aspect critical thinking test in which a case in point is the *Test on Appraising Observations* which deals only with credibility judgment. Hence, his procedure was adapted in establishing the construct validity of all items of *The CEU-Lopez Critical Thinking Test* which is a multi-aspect general-knowledge critical thinking test that deals with five aspects of critical thinking: deduction, credibility judgment, assumption identification, induction, and meaning. These verbal reports provide direct evidence on the processes of students' thinking on how they arrived at their answers in a trial version of a multi-aspect general-knowledge critical thinking test using multiple-choice format. If students choose the unkeyed option but think critically then an item needs modification or replacement. If the students choose the keyed answer by mere guessing then the item can be retained for there is no clear basis from verbal reports that an item is defective. This is one notable disadvantage of using multiple-choice test format for a critical thinking test that an examinee may possibly select the answer keyed correct out of sheer guessing.

Furthermore, it is interesting to note that the correlation between performance and thinking scores together with the verbal reports provide clear guidance on whether the test item is valid or invalid. This is affirmed by Norris (1992) who said that direct evidence on the thinking processes that students used to answer items of multiple-choice critical thinking test is relevant to establish their construct validity.

Undoubtedly, there may be other relevant ways in determining the construct validity of a multiple-choice critical thinking test but verbal reports of thinking gathered from this study show empirically whether an item needs to be revised, retained, or discarded. The high correlation between performance and thinking scores may not mean that the item is free from error as exemplified in item 67. Alternatively, moderate correlation may not mean that the item is problematic and needs to be revised or discarded as shown in item 87. A test developer who considers to adapt this methodology should also look into the insights of verbal reports of thinking than just merely rely on the correlation of performance and thinking scores to establish construct validity of critical thinking test items.

In this fast-changing and highly-connected world, the challenge for educators is to come up with a test in which the focus is on the processes of thinking than just a type of test that calls for students to regurgitate information they just memorized from text. Verbal reports of thinking from this study show that students are capable to think well and it could be enhanced further by challenging our students to give them test that would stretch their thinking ability. Furthermore, although *The CEU-Lopez Critical Thinking Test* is a multiple-choice type of test, the verbal reports of thinking provide direct evidence that students think critically as they answer the questions of the said test that have direct bearing on the processes of critical thinking.

References

- Alderson, J., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Anastasi, A. (1988). *Psychological testing*. New: Macmillan.
- Ebel, R. L., & Frisbie, D.A. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: W. H. Freeman.
- Ennis, R. H. (1996). *Critical thinking*. Upper Saddle River, NJ: Prentice-Hall.
- Ennis, R. H. (2003). Critical thinking assessment. In D. Fasko, Jr (Ed.), *Critical thinking and reasoning: Current research theory and practice* (pp.293-313).USA: Hampton Press, Inc.
- Ennis, R. H. (2009). Investigating and assessing multiple-choice critical thinking tests. In J. Sobocan & L. Groarke (Eds.), *Critical thinking education and assessment: Can higher order thinking be tested?* (pp. 75-97). Canada: The University of Western Ontario.
- Ennis, R. H. (2011a). Critical thinking: Reflection and perspective Part I. *INQUIRY: Critical Thinking Across the Disciplines*, 26(1), 4-18.
- Ennis, R. H. (2011b). Critical thinking: Reflection and perspective Part II. *INQUIRY: Critical Thinking Across the Disciplines*, 26(2), 5-19.
- Fisher, A., & Scriven, M.(1997). *Critical thinking: Its definition and assessment*. Edgepress and Centre for Research in Critical Thinking, University of East Anglia.
- Lopez, M. Y. (2004). *Development and validation of critical thinking infusion lessons in communication skills for freshman college students*. Unpublished doctoral dissertation, Philippine Normal University, Manila.
- Lopez, M. Y. (2012). *The CEU-Lopez Critical Thinking Test*. Research and Evaluation Office, Centro Escolar University, Manila.
- Norris, S.P. (1988). Controlling for background beliefs when developing multiple-choice critical thinking tests. *Educational Measurement*, 7(3), 5-11.
- Norris, S.P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. *Journal of Educational Measurement*, 27, 41-58.
- Norris, S. P. (1991). Informal reasoning assessment: Using verbal reports of thinking to improve multiple-choice test validity. In J. F. Voss, D. N. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 451-472). Hillsdale, NJ: Erlbaum.
- Norris, S. P. (1992). A demonstration of the use of verbal reports of thinking in multiple-choice critical thinking test design. *Alberta Journal of Educational Research*, 38, 155-176.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- Norris, S. P., & King, R. (1983). *Test on appraising observations*. St. John's, NF: Institute for Educational Research and Development, Memorial University of Newfoundland.



Scale on Attitude Toward Mathematics (SATM)

Ivee Guce
Joy Talens
De La Salle Lipa

Abstract This study developed an instrument suited for Filipinos to measure students' attitude towards mathematics. The underlying dimensions of the instrument were determined by examining the responses of 249 students using a factor analysis. The Scale on Attitudes Towards Mathematics among Filipino college students can be recommended for use in the investigation of students' attitudes towards mathematics.

Keywords: attitudes, mathematics, Filipino college students

Filipino students commonly see mathematics as a difficult subject in school. Results of national achievement tests for several years support these findings that Filipino students are poor in mathematics. In the 1999 and 2003 results of Trends in International Mathematics and Science Study (TIMSS) Philippines ranked third from the bottom of the participating countries in terms of mathematics achievement. Learners as social beings with personal beliefs, emotions and views, students' attitudes should be considered in the learning process. Their motivation to learn plays an important role for them to achieve. Thus, attitude cannot be easily separated from learning because they are acquired through the process of learning (Akinsola & Olowojaiye, 2008).

Numerous definitions of attitudes towards mathematics had been proposed which can be classified as "simple" definition which is just a positive or negative emotional disposition toward mathematics (McLeod, 1992; Haladyna, Shaughnessy, & Shaughnessy, 1983; Zan & Martino, 2007). Another classification is multidimensional definition of attitudes towards mathematics which refers to individual's attitude toward mathematics (which, however, have a positive or negative value) by the individual's beliefs towards mathematics, and how he or she behaves (Hart, 1989). Another classification is the bi-dimensional definition as the pattern of beliefs and emotions associated with mathematics. In relation to these definitions, students' attitudes towards mathematics can be measured through instruments. Dutton Scale (Dutton, 1954; Dutton & Blum, 1968) was the first instruments which measured

“feelings” toward arithmetic. Unidimensional scales were developed by Gladstone, Deal and Drevdahl (1960) and Aiken and Dreger (1961). Later Aiken (1974) constructed scales designed to measure enjoyment of mathematics and the value of mathematics. Michaels and Forsyth (1977) and by Sandman (1980) developed multidimensional attitude scales. Examples of scales dealing exclusively with math anxiety are the Mathematics Anxiety Rating Scale (Richardson & Suinn, 1972), the Mathematics Anxiety Rating Scale - Revised (Plake & Parker, 1982) and the Mathematics Anxiety Questionnaire (Wigfield & Meece, 1988). One of the most popular used in research for the last three decades is The Fennema - Sherman Mathematics Attitude Scales (1976) developed in 1976 which consist of a group of nine instruments: (1) Attitude Toward Success in Mathematics Scale, (2) Mathematics as a Male Domain Scale, (3) and (4) Mother/Father Scale, (5) Teacher Scale, (6) Confidence in Learning Mathematics Scale, (7) Mathematics Anxiety Scale, (8) Effectance Motivation Scale in Mathematics, and (9) Mathematics Usefulness Scale.

On the other hand, Ma and Kishor (1997) proposed a definition of attitude towards mathematics as an aggregated measure of liking or disliking of Mathematics, a tendency to engage in or avoid mathematical activities, a belief that one is good or bad at Mathematics and a belief that Mathematics is useful or useless. This definition became the basis for the researchers to develop an instrument on scale on attitudes toward Mathematics which numerous researchers in the Philippines found to play crucial role in students’ achievement in mathematics. To cite an example, for liking and disliking mathematics - I like working on word problems in mathematics; and I get anxious when it’s time for me to study mathematics. Another factor included in the new instrument is tendency to engage in or avoid mathematics activities - “I look forward to our next class meetings in mathematics; and I get overjoyed when my mathematics class is off. For the factor a belief that one is good or bad in mathematics - I can easily solve complex problems; and I do not feel confident about my answers in mathematics seatwork. Another factor is a belief that mathematics is useful or useless - My knowledge in mathematics can help me solve real - life problems and Studying advanced topics in mathematics is just a waste of time.

Kulh (1980, Zan & Martino, 2007.) suggested that “it is probably not possible to offer a definition of attitude toward mathematics that would be suitable for all situations, and even if one were agreed on, it would probably be too general to be useful. Thus, an instrument scale on attitudes toward mathematics among Filipinos has been developed based on the definition proposed by Ma and Kishor (1997). Likewise, not even one of the existing instruments on mathematics scale emphasized the four factors such as liking and disliking mathematics, tendency to engage in or avoid mathematics activities, belief that one is good or bad in mathematics and belief that mathematics is useful or useless.

Method

Content Domain

The instrument was anchored from the definition of attitudes towards mathematics given by Ma and Kishor (1997). The four main clauses in the definition namely - liking and disliking; tendency to engage in or avoid mathematical activities; a belief that one is good or bad in mathematics; and a belief that mathematics is useful or useless were the latent constructs in the study. They are the proposed factors composing the attitudes of Filipino students toward mathematics. Eight statements were composed for each of the four factors. For each factor, there were four positively-stated and four negatively-stated questions to test the consistency of the respondent's answers. That is, for instance, in the first latent construct, four statements address to the liking of the student toward mathematics and the other four are statements about dislike in mathematics. The scale was presented before a panel of test development specialist for comments and suggestions. Some statements were deleted, some were rephrased. All the suggestions given were noted and incorporated before finalizing the instrument. Table 1 shows sample items from each of the factors. All thirty-two statements were jumbled in the final scale, with the four proposed factors not indicated.

Table 1
Sample Items by Factors

Item Number and Content by Factor
Liking or Disliking Mathematics
7. I like working on word problems in Mathematics.
26. I get anxious when its time for me to study Mathematics.
Tendency to Engage in or Avoid in Mathematics Activities
11. I look forward to our next class meetings in Mathematics.
14. I get overjoyed when my Mathematics class is called off.
A Belief That One is Good or Bad in Mathematics
5. I do not feel confident about my answers in Mathematics seatworks.
15. I can easily solve complex problems.
A Belief that Mathematics is Useful or Useless
12. Studying advanced topics in Mathematics is just a waste of time.
20. My knowledge in Mathematics can help me solve real-life problems.

The SATM is 32-item scale. It is constructed using a Likert-scale format with the following anchors: 1 strongly disagree, 2 disagree, 3 moderately disagree, 4 moderately agree, 5 agree, and 6 strongly agree.

Pretesting and Participants

The participants of the study were 249 freshmen engineering department enrolled in College Algebra for the first semester of school year 2012-2013. The test was administered by the participants' respective College Algebra instructor during

their period. They were given 15 minutes to answer the questionnaire. Retrieval of the questionnaires followed a day after for the encoding of data. During the tabulation, the scores for the negatively-stated items were reversed for the analysis of data. Sixteen items were reversed and were given appropriate values. The score was the average of the ratings of the respondents.

Results

To check the internal consistency of the items, Cronbach's alpha coefficients were calculated for all the 32 items and for each set of eight questions for the four latent constructs. The result is given on the table below.

Table 2
Reliability: Internal Consistency Measures Results

Factors	Cronbach's Alpha	Average Inter-Item Correlation
Liking or Disliking Mathematics	0.82	0.37
Tendency to Engage in or Avoid Mathematics Activities	0.72	0.25
A Belief That One is Good or Bad in Mathematics	0.82	0.38
A Belief that Mathematics is Useful or Useless	0.74	0.27

The table above reveals that the Cronbach's alphas and the average inter-item correlations for the four factors are adequate especially the first and third factors. For the overall internal consistency of the 32 items, a Cronbach's alpha of 0.92 was obtained. This indicates a high degree of internal consistency in the statements of the scale.

Table 3
Descriptive Statistics for the Factors of Attitudes Toward Mathematics Scale

Factors	M	Min	Max	SD
Liking or Disliking Mathematics	34.02	14	45	6.08
Tendency to Engage in or Avoid Mathematics Activities	30.82	8	44	5.24
A Belief That One is Good or Bad in Mathematics	27.80	8	45	6.04
A Belief that Mathematics is Useful or Useless	35.90	23	48	5.13

The highest mean computed is for the fourth factor which also gave the maximum sum of responses out of the four. The minimum sum of responses falling under the second and the third factors are also reported. The standard deviations computed are acceptable, especially with the second and fourth factors, indicating that the responses are not much spread or dispersed.

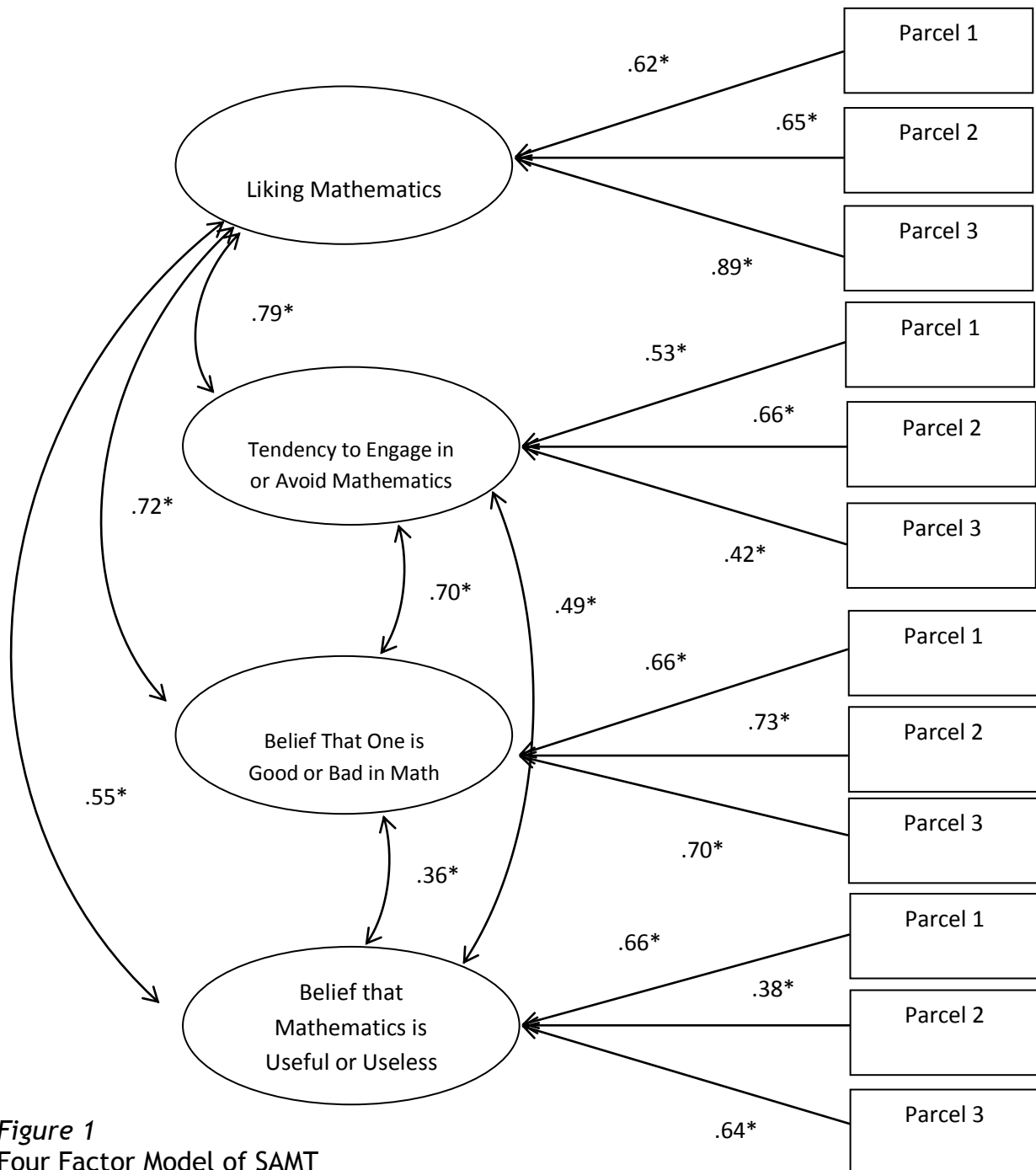


Figure 1
Four Factor Model of SAMT

The factors of SAMT were tested using a Confirmatory Factor Analysis (CFA). It is composed of four latent constructs with eight items each. When these eight items were loaded under one latent construct, it resulted to a single covariance matrix. It was then reduced to three parcels for each factor. Items 1 to 3 made up the first parcel, items 4 to 6 for the second parcel, and the remaining two for the third parcel. With these, the parameter estimates resulted to be significant. This reveals that the four are indeed factors affecting the students' attitudes towards mathematics. The fit of the four-factor model was also obtained.

Table 4

Goodness of fit Indices

RMS Standard Residual	0.05
Steiger-Lind RMSEA Index	0.04
McDonald Noncentrality Index	0.97
Population Gamma Index	0.92
Adjusted Population Gamma Index	0.87
GFI	0.94
AGFI	0.92
Bentler - Bonett Normal Fit Index	0.90
Bollen's Rho	0.90
Bollen's Delta	0.92

It can be seen from the results that different fit indices had acceptable values indicating goodness of fit. This supports the factorial validity of the questionnaire and thus it can be used to assess the attitudes of the students towards mathematics under the mentioned factors.

Discussion

Four subscales were identified as liking and disliking mathematics, tendency to engage in or avoid mathematics activities, belief that one is good or bad in mathematics, and belief that mathematics is useful or useless. Scores on the 32-item scale developed good internal reliability. The estimated time to complete the instrument ranges from 15 to 25 minutes.

The SATM for Filipino students was analyzed using factor analysis and may be an efficient and effective research tool to assess factors that influence expectations and performance in math because of its content validity, reliable factor scores and brevity.

Research on attitude has a long history in mathematics education (Zan & Di Martino, 2007). Factor analysis had matured as a method to examine interrelationships among a number of variables with minimal loss of information. The scale on attitudes towards mathematics for Filipino students was constructed using these standards. Based on a number of researches conducted in the Philippines, to improve math instruction over the last decade focused on teaching methodologies, kind of instructional materials to use, strategies to improve teacher training, and

appropriate sequencing of math courses in the curriculum. Investigation on student attitude towards mathematics among college students in the Philippines has been given less attention. Although there is a body research about attitudes toward mathematics, most of them are concerned only with anxiety (Ma & Kishor, 1997). Most of this research is also based on results derived from instruments that predated modern statistical standards for factor analysis that currently guide the examination of interrelationships among variables. With the development of this scale, teachers and researchers will realize that success or failure in math performance is greatly determined by personal beliefs. It is also anticipated that Filipino college students will exert effort to improve their performance in mathematics influence by personal beliefs such as liking mathematics, tendency to engage in mathematics activity, belief that mathematics is good and belief that mathematics is useful.

References

- Akinsola, M. K., & Olowojaiye, F. B. (2008). Teacher instructional methods and student attitude towards mathematics. *International Electronic Journal of Mathematics Education*, 3, 61-73.
- Goodykoontz, E. (2009). Factors that affect college students' attitudes toward mathematics. Retrieved from http://sigmaa.maa.org/rume/crume2009/Goodykoontz_LONG.pdf.
- Hart, L. (1989). Describing the affective domain: saying what we mean. In R. McLeod & D. Adams (Eds.). *Affect and Mathematical Problem Solving* (pp. 37-45). New York: Springer Verlag.
- Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: a meta analysis. *Journal for Research in Mathematics Education*, 28(1), 26-47.
- McLeod, B. D. (1994). Research on Affect and Mathematics Learning in the JRME: 1970 to the Present. *Journal for Research in Mathematics Education*, 25(6), 637-647.
- Michaels, L.A. & Forysth, R. A. (1977). Construction and validation of an instrument measuring certain attitudes toward mathematics. *Educational and Psychological Measurement*, 37(4), 1043-1049.
- Nicolaidou, M., & Philippou, G. (2012). Attitudes toward mathematics, self-efficacy and achievement in problem solving. Retrieved from www.dm.unipi.it/~didattica/CERME3/
- Olatunde, P. (2009). Students attitude towards mathematics and academic achievement in some selected secondary schools in Southwestern Nigeria. *European Journal of Scientific Research*, 36(3), 336-341.
- Plake, B. S., & Parker, C.S. (1982). The development and validation of a revised version of the Mathematics Anxiety Rating Scale. *Educational and Psychological Measurement*, 42, 551-557.
- Richardson, F.C., & Suinn, R.M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19, 551-554.

- Sandman, R. S. (1980). The mathematics attitude inventory: Instrument and user's manual. *Journal for Research in Mathematics Education*, 11(2), 148 - 149.
- Wigfield, A., & Meece, J.L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, 80, 210-216.
- Zan, R., & Di Martino, P. (2007). Attitude toward mathematics: overcoming the positive/negative dichotomy. *The Montana Mathematics Enthusiast, Monograph*, 3, 157- 168.