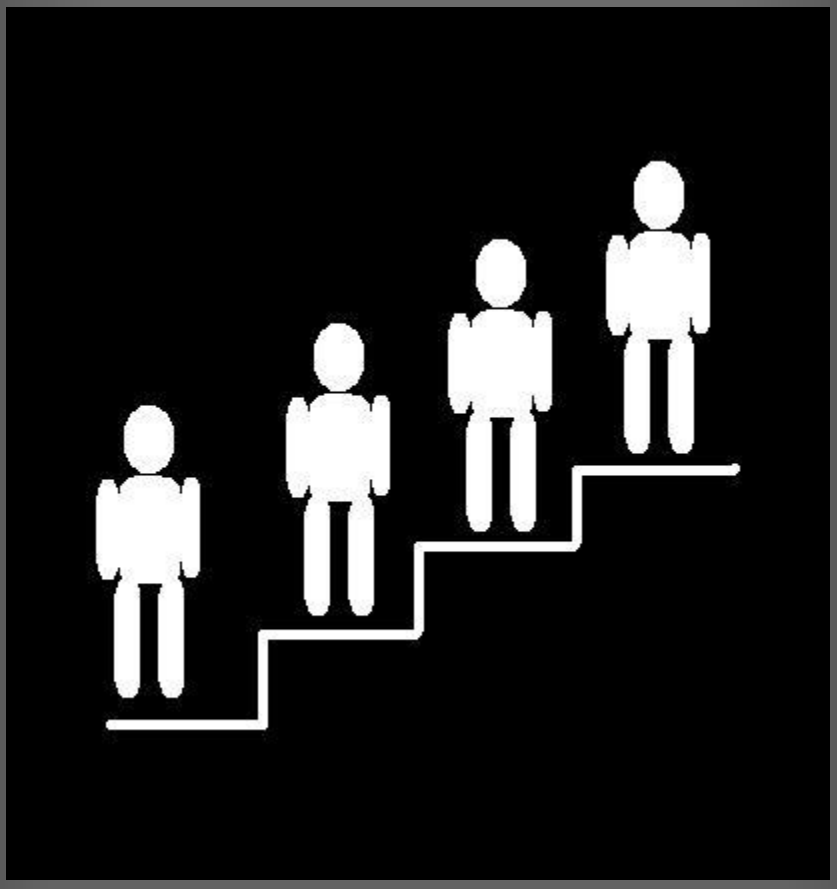
ISSN 2094-1412

The Assessment Handbook

Volume 9, December 2012



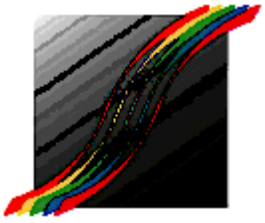
Philippine Educational Measurement and Evaluation Association

The Assessment Handbook

Volume 9,December 2012

The Assessment Handbook is one of the official publications of the Psychometrics and Educational Statistics Division of the Philippine Educational Measurement and Evaluation Association. The journal publishes special articles that are themed related in assessment, evaluation, measurement, psychometrics, psychological testing, and statistics. Each issue of the journal is themed managed by a guest editor. The journal is international, refereed, and abstracted. The journal is presently abstracted/indexed in the Asian Education Index, Social Science Research Network, Google Scholar, Open J-Gate, and NewJour.

The articles of the Assessment Handbook is open access at
http://pemea.club.officelive.com/TheAssessmentHandbook.aspx



Publication Division of PEMEA
Philippine Educational Measurement and Evaluation Association

THE ASSESSMENT HANDBOOK

Volume 9, December 2012
https://sites.google.com/site/theassessmenthandbook/

Articles

Vocabulary and Reading Comprehension as a measure of Reading Skills of Filipino Children

Ryan Francis O. Cayubit
University of Santo Tomas, Manila, Philippines

A Filipino child needs to develop higher order skills and functional literacy. It is given that any Filipino child with sufficient reading skills would have greater chances of success in school compared to a child whose reading skills are poor and more often than not, those with poor reading skills when assessed properly are diagnosed with reading disability. Poor reading skill is manifested with poor comprehension, wrong pronunciations, among others.If no proper intervention is administered early, it could affect the academic, social and psychological development of the child. As such, proper and correct diagnosis of reading disability as early as possible appears to be essential. The purpose of the present study is to develop a new test measuring reading ability or skill (Vocabulary and Reading Comprehension) that could be used for the above-mentioned case. The test that was constructedwas administered to 582 Grades 3 and 4 Filipino pupils. Results showed that the new test has good internal consistency (r = .87 and .74). Using Confirmatory Factor Analysis the model attained an acceptable fit.

Keywords: Vocabulary, Reading Comprehension, Reading, Reading Skills

Academic achievement is at the forefront of any educational institution. Its increase or decrease among pupils or students has always been the concern of experts within or outside an institution as schools and teachers are increasingly held accountable for student achievement (Stipek, 2006; US Department of Education, 2002). Academic achievement in psychology has been loosely defined as a student's previous learning in school. This would often vary from one school to another as it is tied to the school's curriculum. Often, the focus is on common areas of learning or subject like Reading.

Another focus in the research in schools is the proper conduct of assessing student skills. The assessment of one's academic achievement is vital and important in the development of an institution as well as the student's they. When academic achievement of students are assessed, their abilities are always taken into consideration as what they have achieved is largely influenced by their capacity to do it; for example, a good grade in a subject is partly based on the ability of that student for that particular subject. Accurate assessment of student's academic abilities is very important because academic abilities has been identified as one of the most crucial variables related to effective instructional planning and positive student outcomes (Fuchs & Fuchs, 1986; Shinn, 1998; Ysseldyke & Christenson, 1987). It has been argued that without a valid assessment of student's academic skills, instructional decision-making is unlikely to promote academic competence (Martens & Witt, 2004). Given the importance of academic assessment, a variety of measures have been developed that can be used for that purpose. These measures include group-administered achievement batteries, norm-referenced tests of academic achievement and criterion-referenced measures of academic skills (Eckert, Dunn, Coding, Begeny,& Kleinmann, 2006). Although a number of assessment measures are available for measuring the global academic achievement of students, there are also measures that are more specific wherein a particular subject has been targeted. Through the years, several measures of assessment in reading have been developed as discussed by Cain and Oakhill (2006). According to them there are different reasons why practitioners and researchers need to assess a child's reading ability. This is usually done to monitor progress, to detect and diagnose reading difficulties and to test psychological theories of the cognitive skills that underpin reading development and disorders. In addition, Hale et al. (2011) identified reading as one of the greatest areas that assessment is needed as reading skills deficits can interfere with skill development across different academic subject areas, vocational skills and daily living skills. In addition, reading skills have also been linked to students that are commonly referred for special education services (Winn, Skinner, Oliver, Hale, & Ziegler, 2006). For whatever each purpose, what appears to be important is an accurate assessment of reading ability. Hence, the present investigation where the objective is to develop and standardize a tool that could assess the reading ability of Filipino grades 3 and 4 students which is in response to the need for an empirically validated reading interventions and assessment across all grade levels (Hale et al., 2011).

Reading

Reading is said to be one of the most important and complex cognitive skill and such importance has resulted into extensive studies over years (Baddeley, Logie, & Nimmo-Smith, 1985). Reading has been defined as a process of interaction involving one's knowledge of print, vocabulary, and comprehension. Its five essential components include phonemic awareness, phonics, fluency, vocabulary, and comprehension. In addition, Fitzgerald and

Fitzgerald (1965) included word recognition and sentence understanding as components. They further stated that the components involve discovery, comprehension, reflection, reasoning, appreciation, analysis, evaluation, synthesis, organization, and application. This would mean that when one is reading, one is thinking about the meaning conveyed and at the same time integrates his own knowledge to get the meaning of the symbols written by the writer. Though the concept of reading is broad and comprises several components, the focus of this research would only be on the areas of vocabulary and reading comprehension in line with the view that an approach to studying and assessing fluency in reading is to focus in specific reading tasks that will allow individual components of the reading process to be isolated and studied (Baddeley, Logie, & Nimmo-Smith, 1985).

"Reading comprehension is a complex cognitive ability requiring the capacity to integrate text information with the knowledge of the listener or reader and resulting in the elaboration of a mental representation" (Meneghetti, Carretti, & De Beni, 2006, p. 291). As a component of reading, reading comprehension can be best understood if one is adept with the different cognitive processes as current models suggest that such processes play a significant role in comprehension skills (Meneghetti et al., 2006). van den Broek (1994) highlighted that short and long term memory is a factor in the reading comprehension skills of an individual as a reader needs to store and manipulate information in his working memory during text procession and at the same time in order to construct a coherent representation of what he has read, the reader would have to refer to his prior knowledge. Inference also plays a major role in reading comprehension as understanding of the text read goes beyond literal wherein integrated mental representation of what was read is created and processed (Bowyer-Crane & Snowling, 2005; Yuill & Oakhill, 1991). Recent studies on reading comprehension stressed the importance of the concept of individual differences wherein attempts are made to account for how the process and components of reading comprehension differ among those labeled as skilled and less skilled readers (Oakhill, Cain, & Bryant, 2003). Such labels or classifications are results of meaningful assessment of one's reading skills or achievement wherein comparisons are made using tasks that measure either global or specific areas of reading comprehension and making inferences out of its results (Meneghett et al., 2006). The literature also puts emphasis on the effects of being a poor comprehender or being less skilled. Those who are less skilled have problems interrelating successive topics being read (Lorch, Lorch,& Morgan, 1987), integrating information or themes (Palincsar & Brown, 1984), understanding story structure (Cain & Oakhill, 1996), and rarely uses reading strategies (Brown, Armbruster, & Baker, 1986). This further strengthened the need for a tool to serve that particular purpose because of the fact that children vary in their skill as readers.

The other variable that is the focused on this research is vocabulary. Literature for vocabulary are less when compared with reading comprehension, but if one would analyze the nature of both aspects of reading, vocabulary, and comprehension appears supplementary wherein vocabulary focuses on

recognition of words and identification of its meaning while reading comprehension is all about the knowledge or understanding of what has been read the first step of which is recognizing and giving meaning to words. Just like reading comprehension, vocabulary of children differs. Most children acquire vocabulary during the preschool years but the said acquisition is more arduous (Anglin, 1993; Hargrave & Senechal, 2000;) and given its importance wherein children's vocabulary can be a predictor of their overall reading ability (Hargrave & Senechal, 2000), the need for a tool to assess it seems imperative.

Studying reading is synonymous to studying reading disabilities or problems associated with reading. Despite efforts to have a highly literate population among children, there has been a rise in the number of cases of children with reading problems or disabilities. In the United States, many children exhibit reading difficulties as reported by the National Assessment for Educational Progress (2005). In the Philippines, the Department of Education has reported a number of cases of children with learning difficulties. Its impact is not only limited to poor reading achievement because studies have shown that poor readers are at significantly greater risk than good readers for developing attention and behavioral problems (Adams & Snowling, 2001; Maughan & Carroll, 2006). Thus stressing the notion and importance of assessing reading ability or achievement early in order to identify those children that would need intervention.

Aside from using reading tools to assess reading problems or difficulty, it can also be used to gauge reading achievement or accomplishments of children. This is of equal importance as reading achievement have been linked to many research variables. Reading achievement was found to be related to higher levels of self-esteem among students (Kaniuka, 2010), reading achievement was also related to extrinsic motivation (Chiu, Chow, & Mcbride-Chang, 2007), and significant predictors of scholastic achievement (Savolainen, Ahonen, Aro, Tolvanen, & Holopainen, 2008; Meneghetti et al., 2006).

Bilingualism

This study also took into consideration the concept of bilingualism and its possible effects on the reading achievement of Filipinos given that English is not the mother tongue of Filipinos. In the study by Meyer (2000), she stated that bilingual students are having problems using English as a medium of instruction since it brings confusion in learning with their first language. Questions would now arise on the reliability of the results of the different assessment measures of reading achievement that since most of them are written in English in a sense that when these instruments report a child a poor in reading.Is the child really poor in reading or the only reason for his low score is because the test items are written in a language that he cannot fully understand? This is one of the reasons why Lee (2008) conducted his study on the development and validation of a reading-related assessment battery in Malay for the purpose of assessing dyslexia. This is to ensure that the assessment of children in Malaysia for dyslexia would be more accurate and a clearer picture of the current state of

dyslexia in the country would be presented by making use of an instrument in their native language. This may be true also for the Philippines, despite the fact that Filipinos are largely considered to have a good command of the English language, assessment that will make use of the Filipino language would present a clearer picture of the reading ability of the Filipino Grades 3 and 4 students with respect to Vocabularyand Reading Comprehension.Thepresent study developed an instrument written in Filipino that measures the reading skills of Grades 3 and 4 pupils.Particularly, the new test measures Filipino vocabulary and reading comprehension.

Method

Design

The present study used a quantitative study cross-sectional explanatory design. According to Johnson (2001), a cross-sectional explanatory research would entail the gathering of data from the respondents during a single point in time with the objective of developing an instrument that would measure a phenomenon and explaining the nature of the phenomenon. Moreover, this was also a descriptive normative research as it made use of the constructed test and determines its usefulness to explain the current condition as regards to the participants' achievement in vocabulary and reading comprehension.

Participants

The study made use of two sets of participants that were selected through convenience sampling. They are Grades 3 and 4 pupils of selected public and private schools within the Metro Manila area. A total of 582 pupils participated in the research; 238 of who were involved in the development of the preliminary form while the remaining 344 took part in the development of the polished form.

Measure

To gather the needed data for the completion of the project, several instruments were used. The researcher after the consultation with experts developed four different instruments. They are as follows: preliminary and polished form for vocabularyand the preliminary and polished form for reading comprehension. All items were written in Filipino and are based on the Filipino and reading subjects of the Grades 3 and 4 pupils.Table 1 contains the description of all developed or used instruments.

Procedure

The entire test development process was divided into three stages. The first stage involved the writing of the test items in Filipino. This was undertaken after a review of the Filipino and reading subjects in Grades 3 and 4 to ensure

that the items that would be included is reflective of what the pupils are exposed to and is studying. The items were then formatted and submitted to experts for content validation. Experts in both Filipino and reading reviewed and evaluated the items in terms of suitability and appropriateness. From the original 65 items for vocabulary and 60 items for reading comprehension, only 50 items for both subscales were retained. This made up the preliminary form of vocabulary and reading comprehension.

Table 1
Description of test instruments

Type of Test	No. of Items	Description
Preliminary Form Vocabulary	50	Measures the skill of the test taker in identifying Filipino words and matching it with its corresponding meaning.
Polished Form Vocabulary	45	
Preliminary Form Reading Comprehension	50	Measures the skill of the test taker in reading passages and understanding it by answering questions about it.
Polished Form Reading Comprehension	27	

The second stage focused on determining the usability of the preliminary form. It was administered to 238 Grades 3 and 4 pupils from 5 schools that agreed to participate in the research. The result of which was subjected to Cronbach's alpha for reliability testing and item analysis (difficulty and discrimination). After the initial analysis, only 45 items were retained for the Vocabulary subscale 27 for the Reading Comprehension.

Stage three dealt on the use of the polished form. It was administered to 344 Grades 3 and 4 pupils from 3 schools. Data gathering was again subjected to Cronbach's alpha for reliability testing and Confirmatory Factory Analysis for validity.After which, the normative structure was constructed, specifically Stanine and Percentile Rank.

Data Analysis

Several techniques were used to determine the psychometric properties of both the preliminary and polished forms of the vocabulary and reading comprehension subscales. Intra-class analysis method, Cronbach's alpha, item analysis, and Confirmatory Factory Analysis (CFA) were obtained. In addition, the Means, Standard Deviation, Percentiles, and Stanines were also obtained.

Results

Content Validation

Results of the Intra Class Reliability Method of the responses during the expert validation of the two subscales are an indication that the new test has content validity. Obtained value for vocabulary (r = .58) and reading comprehension (r = .81) reveal a consistent rating among the experts, thereby indicating a high degree of agreeability in terms of the usability and appropriateness of the test items.

Descriptive Statistics

Means and Standard Deviations of the participants from the preliminary and polished forms of both the vocabulary and reading comprehension subscales are found in Table 2. The distribution of scores does not appear normal (skewed to the left) with the exception of the preliminary form of reading comprehension. For the confidence interval (95%)the computed means for the preliminary form is both below (Vocabulary; M = 31.54) and above (Reading Comprehension; M = 29.95) the expected ranges. A similar scenario was observed in the polished form with the Vocabulary (M = 35.94) exceeding the upper limit of the interval and Reading Comprehension (M = 21.21) not meeting the lower limit of the interval.

Table 2
Descriptive statistics of the preliminary and polished forms

Subscales	M (SD)	Skewness	Kurtosis	95% Confidence Interval
Preliminary Form				
Vocabulary	31.54 (11.10)	-1.227	1.307	35.25 – 36.62
Reading Comp.	29.95 (6.20)	-.658	-.572	20.82 – 21.59
Polished Form				
Vocabulary	35.94 (6.50)	-1.531	2.094	30.12 – 32.96
Reading Comp.	21.21 (3.58)	-1.247	2.363	29.16 – 30.75

Internal Consistency of the Preliminary and Polished Form

Results showed that the preliminary form of both Vocabulary and Reading Comprehensioncontain items that are internally consistent with alpha values of .93 and .87 reliability respectively, an indication test soundness, stability and dependability. Concerning the polished form, similar interpretation can be made despite the decrease in value for the Vocabulary subscale (r = .87) and reading comprehension (r = .74). The decrease in the value of the coefficient can be explained by the decrease in the number of test items as the reliability

coefficient of a test is affected by the length or the number of items in a particular test (Kaplan & Saccuzzo, 2009). Nonetheless, the polished form is internally consistent and can be dependent on to generate stable scores over a period of time.

Item Analysis

Item analysis was performed to determine the level of difficulty and discriminatory of the contents of the preliminary form. This served as one of the basis to find out which test items should be included in the polished form. Based on the computation the items on vocabularyhave an average difficulty index of 0.59, while reading comprehension'saverage difficulty index is 0.75. In addition, the average discriminatory index or power of the test items is 0.32 and 0.59 for Reading Comprehension and Vocabularyrespectively. Based on the results, all test items in the preliminary form are classified within the range of reasonably good item to very good items (Sevilla, Ochave, Punsalan, Regala, & Uriarte, 1999).

Table 3
Percentage of retained items against discarded items

Subscales	Retain (%)	Discard (%)
Vocabulary	45 (90%)	5 (5%)
Reading Comprehension	27 (54%)	23 (46%)

The results of the item analysis determined the contents of the polished form. Ideally, a good test should have good items as this will aid in enhancing its reliability and validity (Freidenberg, 1999). As such, the composition of the polished form of the instrument developed includes test items that are either reasonably good or very good. Based on the statistical computation the number of test items that composed the polished form was reduced as items classified as poor were discarded. The polished form of Vocabulary and Reading Comprehensionhas 45 and 27 items respectively. The average difficulty level and discriminatory power for Reading Comprehensionpolished form are 0.71 and 0.40. While for Vocabularyit is 0.62 and 0.56 for difficulty and discrimination.

Confirmatory Factor Analysis

The polished forms of the instruments were also subjected to Confirmatory Factor Analysis (see Figure 1). It was tested for goodness of fit using the chi-square (χ^2), RMSEA, Akaike Information Criterion (*AIC*), Comparative Fit Index (*CFI*), Population Comparative Fit Index (*PCFI*), among others.Table 3 contains the results of the Confirmatory Factor Analysis for the polished form. To extensively evaluate the factor structure of the instrument, multiple indices were used. The traditional chi-square was evaluated along with the RMSEA, PGI, APGI, Joreskog GFI and AGFI.

Table 3
Goodness of Fit indices

Goodness of Fit Index	Values
RMSEA	0.043
PGI	0.885
APGI	0.879
Joreskog GFI	0.752
Joreskog AGFI	0.737
Chi square	4131.038

Results show that the chi square value is considered low and is significant at 0.05 alpha level. This is an indication of the departure of the data from the model as the significant value suggests. Of all the indices that were reported, it is only the chi-square that generated such an indication. As discussed by Anderson and Gerbing (1988), and Huang and Michael (2000), cited by Ganotice (2010, p.66) that the value of the chi-square likelihood ratio statistic is directly dependent on sample size whereas large sample size may generate significant values even if the discrepancies are trivial.

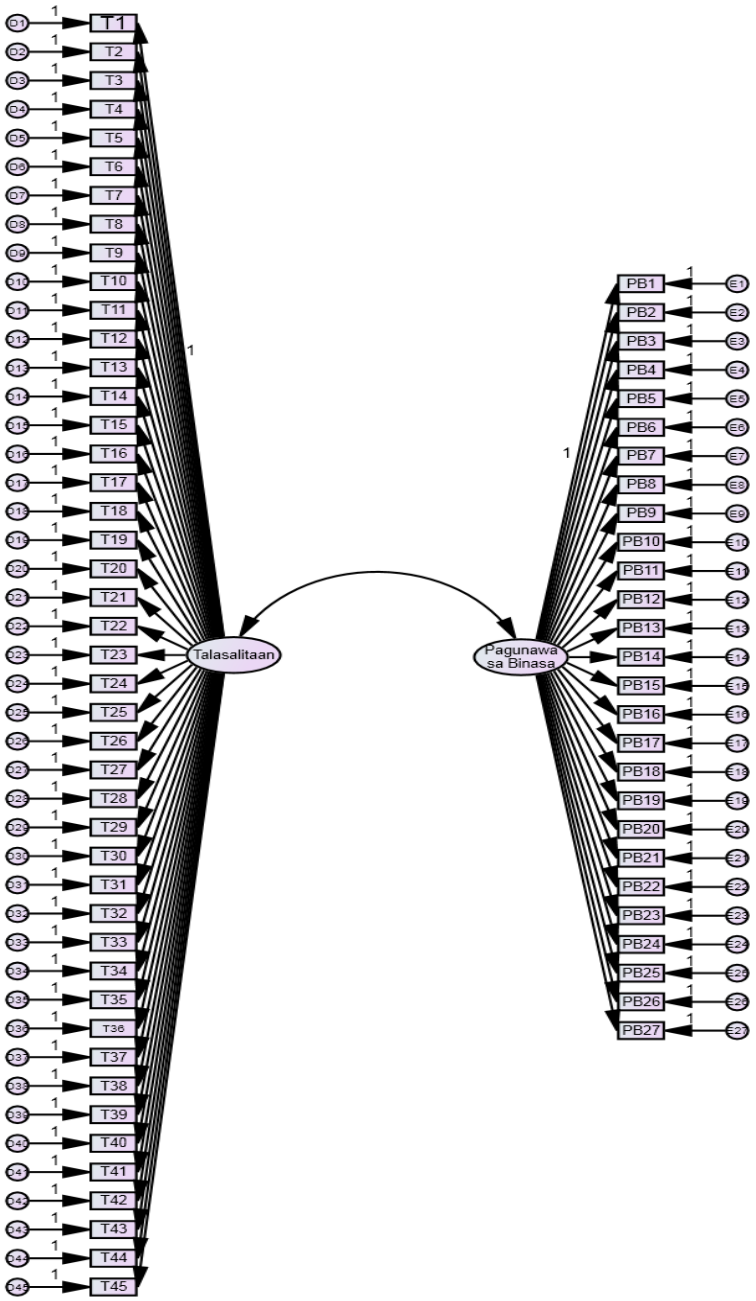
Despite the chi-square results, other indices reported a fit that is along the adequate and good level. The RMSEA value of 0.043 is an indication of a good fit as discussed by Hu and Bentler (1999) wherein they stated that values that are less than .06 indicate a good fit. Other indices computed through CFA indicate an adequate fit as they are near the cut off-point or rule of thumb in determining whether the fit is good or not. The above results indicate that both Vocabulary and Reading Comprehensionis a valid instrument to measure reading skill.

Normative Structure

The normative structure of the developed test is based on its polished form and derived from 344 respondents within the Metro Manila area. Though it is ideal to have included samples outside the Metro Manila area, due to time constraints and practical reasons, the researcher decided to confine its investigation within the declared area. This move is supported by Gregory (1996) wherein he stated that very few test developers would go into using and fully employing large samples taken from different areas in selecting their norm group. He further explained that what is more typical is a good faith effort to pick a diverse and representative sample.

In relation to the constructive normative structure, the raw scores were converted to percentile and the percentiles were converted to STANINE (Standard Nine). Percentile ranks ranges from a low of 1 to a high of 99; while the STANINE is from 1 to 9. In terms of qualitative interpretation, the normative structure was divided into three categories, namely: Below Average, Average, and Above Average. STANINE scores that ranges from 1 to 3 would indicate Below Average achievement while STANIE scores that ranges from 4 to 6

translates into an Average achievement and the STANINE scores of 7 to 9 means an Above Average achievement.



Note: Talasalitaan is the Vocabulary and Pagunawa sa Binasa is Reading Comprehension

Figure 1
Model tested for CFA.

Discussion

The aim of the present study was to construct and validate a new test that will measure Filipino children's reading skills with respect to the facets of vocabulary and reading comprehension. The choice of vocabulary and reading comprehension as subscales is based on the notion that these are the two basic components of reading. According to Hargrave and Senechal (2000) Vocabulary is a component of reading that provides an indication of the overall reading ability of the participants and participants who scored high on this subscale demonstrates high skills in terms of understanding and deducing the meaning of Filipino words common among Grades 3 and 4 pupils. Reading Comprehension was included because it has been identified as one of the basic components of reading ability (Baddeley et al., 1985). This measures the level of understanding of the participants of the passages that they have read by answering its succeeding questions. Participants who score high on this subscale exhibit the ability to understand and make inferences about what they have read.

The researcher also choose to write the test in Filipino in order to accurately assess Filipino children's reading ability and avoid instances wherein one would be judged as poor in reading simply because he or she cannot understand the English language but appears to do well in reading when items are in Filipino.

The pertinent data that has been gathered indicates that Vocabulary and Reading Comprehension is a valid and reliable instrument that can measure reading skills of Grades 3 and 4 pupils. It's content validity has been established via expert validation and is deemed important because according to Freidenberg (1995), one of the basic requirement for a test that make inferences about the broader domain of knowledge and/or skills is a valid content. This has been supplemented by the results of the confirmatory factor analysis where several fit indices have adjudged the factor structure of the new test as fitting. Such empirical evidences aids in the realization of the objective of the instrument to provide accurate information about the reading skills of Filipino students.

In addition to being valid, the new test is also reliable. In terms of application, a test that is high in reliability is highly favored over one that is not because a reliable test can be depended on to generate scores that are realistic estimates of the test taker's actual knowledge or characteristics (Freidenberg, 1995). Thus the Vocabulary and Reading Comprehension subscales can produce data that is reflective of the skill of Filipino children with respect to the domains measured by each of them

The above psychometric properties of the new test is particularly important because an accurate assessment of student's academic abilities has been identified as one of the most crucial variables related to effective instructional planning and positive student outcomes (Fuchs & Fuchs, 1986; Shinn, 1998; Ysseldyke & Christenson, 1987). Likewise, without a valid assessment of student's academic skills, instructional decision-making is unlikely to promote academic competence (Martens & Witt, 2004). Tests of such nature accurately yields information on student's abilities that can be used by teachers,

administrators and educational managers in designing their lessons, curriculum, study plan and other similar programs wherein all of the above can be tailored fit in order to meet the needs of the students that they serve thereby increasing chances for academic success. And in a more practical sense, the use of the said instrument can aid in the Philippines implementation of the K to 12 Basic Education Program where one of its features is building proficiency through language as the use of language has been identified as a factor in reading development and vice versa.

References

Adams, J.,& Snowling, M. (2001). Executive functioning and reading Impairments in children reported by their teachers as hyperactive. *British Journal of Developmental Psychology,19*(2), 293-306.

Angling, J.M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10).

Baddeley, A., Logie, R.,& Nimmo-Smith, I. (1985). Component of fluent reading. *Journal ofMemory and Language, 24,* 119-131.

Bowyer-Crane, C.,& Snowling, M.J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology,75,* 189-201.

Brown, A.L., Ambruster, B.B.,& Baker, L. (1986). The role of metacognition in reading and studying. In J. Orasanu (Ed.), *Reading comprehension: From research to practice.* Hillsdale, NJ: Erlbaum.

Cain, K.,& Oakhill, J. (1996). The nature of the relation between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology, 14*(2), 187-201.

Cain, K.,& Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology,76*(4), 697-708.

Chatterji, M. (2006). Reading achievement gaps, correlates and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology,98*(3), 489-507.

Chiu, M.M., Chow, B.W.,& Mcbride-Chang, C. (2007). Universals and specifics in learning strategies: Explaining adolescent mathematics, science, and reading achievement across 34 countries. *Learning and Individual Differences, 17,* 344-365.

Eckert, T., Dunn, E., Codding, R., Begeny, J., & Kleinmann, A. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools,* (43), 247-265.

Fitzgerald, J.,& Fitzgerald, P. (1965). Teaching reading and the language arts. USA:Bruce Publishing Company.

Friedenberg, L. (1995). Psychological testing: Design, analysis and use. Massachusetts:Allyn & Bacon.

Fuchs, L.S.,& Fuchs, D. (1986). Effects of systematic formative evaluation on studentachievement: A meta-analysis. *Exceptional Children,*(53), 199-208.

Ganotice, F. (2010). A confirmatory factor analysis of scores on Inventory of School Motivation (ISM), Sense of Self Scale, and Facilitating Conditions Questionnaire(FCQ): A study using a Philippine sample.

Ghelani, K., Sidhu, R., Jain, U., & Tannock, R. (2004). Reading comprehension and reading related abilities in adolescents with reading disabilities and attention-deficit/hyperactivity disorder. *Dyslexia,*(10), 364-384.

Gregory, R. (1996). Psychological Testing: History, principles and applications (2nd ed.) Needham Heights, MA: Allyn & Bacon

Hale, A. et al. (2011). Reading assessment methods for middle-school students: An investigation of reading comprehension rate and maze accurate response rate. *Psychology in the Schools, 48*(1), 28-36.

Hargrave, A.C.,& Senechal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1), 75-90.

Hu, L.T.,& Bentler, P.M. (1999). Cutoff criteria for fit indixes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* (6), 1-55.

Huang, C.,& Michael W. B. (2000). A confirmatory factor analysis of scores on a Chinese version of an academic self-conceptscale and its invariance across groups. *Educational and Psychological Measurement,*(60), 772-786.

Kaniuka, T. (2010). Reading achievement, attitude toward reading, and reading self-esteem of historically low achieving students. *Journal of Instructional Psychology, 37*(2), 184-188.

Kaplan, S.,& Saccuzzo, D. (2009). Psychological testing: Principles, applications and Issues (7th ed.). Belmont CA: Thomson Wadsworth

Keller-Margulis, M., Shapiro, E.,& Hintze, J. (2008). Long-term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review,* (3), 374-390.

Lee, L. W. (2007). Development and validation of a reading-related assessment battery in Malay for the purpose of dyslexia assessment. *Ann of Dyslexia,*(58), 37-57.

Lorch Jr., K.G., Lorch, E.P., & Morgan, A.M. (1987). The task effects and individual differences in an on-line processing of the topic structure of a text. *Discourse Processes, 10,* 63-80.

Martens, B. K., & Witt, J.C. (2004). Competence, persistence and success: The positive psychology of behavioral skill instruction. *Psychology in the Schools,*(41), 19-30.

Maughan, B.,& Carroll, J. (2006). Literacy and mental disorders. *Current opinion in Psychiatry,* (19), 350-354.

Meneghetti, C., Carretti, B.,& De Beni, R. (2006). Components of reading comprehension and scholastic achievement. *Learning and Individual Differences, 16,* 291-301.

Meyer, L. (2000). Barriers to meaningful instruction for English learners. *Theory into practice*. http://www.jstor.org/stable/1477342

Oakhill, J., Cain, K.,& Bryant, P.E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18,*443-468.

Palincsar, A.S.,& Brown, A.L. (1984). Reciprocal teaching of comprehension fosteringand comprehension monitoring activities. *Cognitive and Instruction, 1,* 117-175.

Ponitz, C., Rimm-Kaufman, S., Grimm, K.,& Curby, T. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review,38*(1), 102-120.

Sevilla, C., Ochave, J., Punsalan, T., Regala, B.,& Uriarte, G. (1992). Research Methods (Revised ed.). Quezon City: Rex Bookstore

Stipek, J.D. (2006). No child left behind comes to preschool. *Elementary School Journal,*(106), 455-465.

Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A.,& Holopainen, L. (2008). Reading comprehension, word reading and spelling as predictors of school achievementand choice of secondary education. *Learning and Instruction, 18,* 201-210.

U.S. Department of Education. (2002). No child left behind: A desktop reference. Office of the Elementary and Secondary Education. Washington D.C.

van den Broek, P. (1994). Comprehension and memory for narrative texts: Inferences and coherence. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA:Academic Press.

Winn, B.D., Skinner, C.H., Oliver, R., Hale, A.D.,& Ziegler, M. (2006). The effects of listening while reading and repeated reading on the reading fluency of adult learners. *Journal of Adolescent and Adult Literacy, 50,* 196-205.

Yuill, N.,& Oakhill, J. (1991). *Children's problem in text comprehension: An experimental investigation.* New York: Cambridge University Press

Establishing the Construct Validity and Reliability of an Urdu Translation of the Test Anxiety Inventory

Muhammad Shabbir Ali
Government College University, Faisalabad, Pakistan

The aim of this study was to explore the psychometric properties of an Urdu translation of the Test Anxiety Inventory (U-TAI) and replicate findings ongender differences and relations with performance. A sample of 1885 secondary school students from the Punjab province of Pakistan completed the U-TAI approximately three months before taking the Secondary School Certificate examinations (examinations in Pakistan required to leave Secondary School) and data collected for performance in math and science subjects. A two-factor structure consisting of worry and emotionality components of test anxiety showed acceptable construct validity and internal reliability. Female students reported higher emotionality scores, and inverse relations with performance were stronger for the worry component. The U-TAI has showed sufficient validity and reliability to be used in subsequent research with Urdu speaking people.

Keywords: Construct Validity, Reliability, Test Anxiety Inventory (TAI)

Ullah, Richardson and Hafeez (2011) noted the paucity of research into the experiences of university students in Pakistan. This research has inadvertently discovered that the same applies to students at school level in Pakistan. One of the main barriers to facilitate work into experiences of students of all ages in Pakistan is the lack of measures typically found in contemporary psycho-educational research (self-efficacy, achievement goals, self-concept, assessment-related emotions, and so forth) which have been translated into Urdu, the main language spoken in Pakistan, or one of the other regional dialects. Although cultural variations in the definitions and experience of emotions may exist, anxiety remains one of the basic universal markers of psychological well-being (Spielberger, 2006)

Defining the Test Anxiety Construct

Test anxiety is defined as a situational-specific anxiety trait in which individuals have a greater or lesser tendency to appraise performance-evaluative situations as threatening (Spielberger & Vagg, 1995). Transactional process models of test anxiety (Loweet et al., 2008; Zeidner & Mathews, 2005) position trait test anxiety as one of several possible personal (e.g., competence beliefs) and situational (e.g., importance of test) antecedents which may combine to determine the actual degree of (state) anxiety experienced in a specific performance-evaluative situation. Test anxiety has usually been investigated in an educational context, concerning the tests, examinations and other assessments taken by students in school, college and universities (Putwain, 2008a). In principle, test anxiety could, however, apply to any situation in which one's performance is judged or evaluated by others (e.g., a driving test) although examples in the literature are relatively rare (Fairclough, Tattersall,& Houston, 2006).

The Multidimensionality of Test Anxiety

Test anxiety has long been considered as multidimensional and a fundamental distinction is made between the cognitive and affective-physiological components of anxiety (Spielberger, Gonzalez, Taylor, Algaze,& Anton, 1978). The cognitive component, typically labelled as worry, represents thoughts and other self-depreciating statements regarding failure and the consequences of failure (e.g., not attaining cherished goals, being judged negatively by others and so forth). The affective-physiological dimension, usually labelled as emotionality, represents the person's perception of their autonomic arousal. The worry/emotionality distinction has proved extremely robust and has also been replicated in many studies (Benson, Moulin-Julian, Schwarzer, Seipp,& El-Zahhar, 1992). Although this distinction has been elaborated on in subsequent work (Benson et al., 1992; Sarason, 1984) and other components of test anxiety have been proposed (Friedman & Bendas-Jacob, 1997; Lowe et al., 2008), the distinction between cognitive and affective-physiological components remains central to the test anxiety construct definition and domain.

Although worry and emotionality are related, one of the ways in which the distinction is useful, both theoretically and substantively, is in relations with educational performance or achievement. A robust and well replicated finding is that small inverse relations are reported between educational performance and test anxiety, which tend to be larger for the worry component than for the emotionality component (Chapell et al., 2005; Hembree, 1988; Seip, 1991). Explanations usually focus on the role played by worry cognitions in occupying working memory resources, making it difficult to organise one's thought and recall material which has been previously learned, particularly when examination or test questions require the student to conduct several sequential steps and hold the answer to one step in mind, all while thinking about the next step (Derakshan & Eysenck, 2009; Owens, Stevenson, Norgate,& Hadwin, 2008).

One other notable and well-replicated finding regarding the different components of test anxiety is that female students report higher scores on the emotionality component whereas gender differences on the worry component are either smaller or not present (Zeidner, 1990; Zeidner & Nevo, 1992; Zeidner & Schleyer, 1999). Explanations focus on presentation bias and socialization processes although there has been no convincing evidence for either. Gender differences do not, however, appear to moderate the test anxiety and educational performance relationship (Putwain, 2008b).

Although test anxiety has been investigated in many different counties (Seipp & Schwarzer, 1996) and measures such as the Test Anxiety Inventory (TAI) have been translated into many different languages (O'Neil & Fukumura, 1992), sometimes for use in cross-cultural studies and sometimes for use in host cultures, there has been no other measures to-date available for use in Pakistan. To facilitate future work into test anxiety using samples of Pakistani students, we report work in which we have translated the most well-known measure of test anxiety with arguably the most widespread use, the TAI, into Urdu and checked the psychometric features of this measure via its construct validity, reliability and discriminative validity.

Aim of the Study

The aim of this study was to translate the TAI into Urdu and then check the properties of this measure to establish its reliability and validity for use in future research. First, the factorial validity and internal reliability of the translated TAI was examined, expecting that the two-factor structure of worry and emotionality components would be demonstrated in a Pakistani sample of students with acceptable internal reliability. Second, the gender differences (including factorial invariance across male and female students) were measured, expecting to find female students reporting higher emotionality scores and no (or smaller) gender differences in worry scores. Last, the correlations between test anxiety and examination performance was examined, expecting to find inverse relations, which were stronger for the worry component than the emotionality component (a test of discriminative validity). Although these theoretical predictions are replications of existing research, the research described here offers an extension to the extant literature by establishing the reliability and validity of the TAI in a new culture; an important step in preliminary research.

Method

Participants

Data was collected from 1885 secondary school students drawn from sixty-four schools located in four districts from the Punjab province of Pakistan. The sample was stratified so that data was collected from equal numbers of schools in urban (n = 1197) and rural (n = 688) locations, single sex girls' (n = 887) and

boys' schools (*n* = 998) in each of the four districts in Punjab province. Participants were in the 10th grade of school (the final year of compulsory education in Pakistan), aged 15-16 years in which students take public Secondary School Certificate (SSC) examinations in math, physics, chemistry and biology.

Measure

Test anxiety was measured using an Urdu translation of the Test Anxiety Inventory (U-TAI: Spielberger, 1980). Although more recent measures are available, this classic measure was selected for several reasons: (1) it is the most widely used measure of test anxiety (Benson et al., 1992) in which factorial validity has been demonstrated in versions translated for use in other cultures (Benson et al., 1992; Seipp & Schwarzer, 1996), (2) other measures all incorporate the fundamental distinction between cognitive and affective-physiological components which are included on the TAI, (3) there is no consensus in the literature over which additional components of test anxiety should be included in the construct (cf. Lowe et al., 2008) and equivocal findings regarding additional components (cf. Putwain, Connors,& Symes, 2010). The TAI would therefore seem an appropriate measure with which to start preliminary research. The TAI consists of twenty statements regarding the worries and anxieties that students experience in tests and examinations. Students respond on a scale of 1 = almost never, 4 = almost always. Eight statements correspond to the worry subscale (e.g., 'Thoughts of doing poorly interfere with my concentration in tests'), while another eight focus on emotionality (e.g., 'I feel very jittery when taking an important test') scale, with the remaining four statements included in a total TAI score. Measures of educational performance were taken from board certified SSC examination results in math, physics, chemistry and biology.

Procedure

The TAI was independently translated and back-translated from English to Urdu. The researcher as part of an on-going project collected data about test anxiety in Pakistan in the usual classroom environment, approximately three months before students appeared in their SSC examinations. Also, prior to data collection, the aims of the project were explained to students.

Results

Factorial Validity of the U-TAI

Using confirmatory factor analyses, five different models of the U-TAI were tested: (1) a unidimensional model, (2) a model based on the original TAI with eight items loading separately on each of the worry and emotionality components as first-order factors and covariance specified between worry and emotionality, (3) an alternative model also based on the original TAI with eight

items loading separately on each of the worry and emotionality components as first-order factors, four items loading on both factors and covariance specified between worry and emotionality, (4) a model which specified worry and emotionality as lower order factors and test anxiety as a higher order factor, based on model 2 with covariance removed, and (5) a model which specified worry and emotionality as lower order factors and test anxiety as a higher order factor, based on model 3 with covariance removed. In line with recommendations for assessing model fit (Marsh, Hau,& Wen, 1999; Marsh, Hau,& Grayson, 2005), used several criteria including the χ^2 statistic, Root Mean Square Error Approximation (RMSEA), Confirmatory Fit Index (CFI) and Tucker-Lewis Index (TLI). RMSEA values of ≤.05 and CFI/TLI values of ≥.95 are considered as evidence of a good fitting model and RMSEA values of ≤.08 and CFI/TLI values of ≥.90 are considered as evidence of a reasonable fitting model. Confirmatory factor analyses are reported here in Table 1.

Table 1
Confirmatory Factor Analysis

Model	x2	df	RMSEA	CFI	TFI
Model 1: Unidimensional	1128.35***	170	.055	.880	.866
Model 2: 16-item first order	473.92***	103	.046	.949	.940
Model 3: 20-item first order	737.75***	165	.043	.928	.918
Model 4: 16-item higher order	473.92***	103	.046	.949	.940
Model 5: 20-item higher order	737.75***	165	.043	.928	.918

The analyses reported in Table 1 indicate that models 2 and 4 offered the best fit, but there was no particular advantage to a model with a higher order factor (model 4) of general test anxiety that comprised only of two lower level factors (model 2), worry and emotionality, which covaried. Therefore model 2 was accepted. Factor loadings are reported in Table 2.

Factorial Invariance for Male and Female Subsamples

To establish whether this factor structure was equivalent for male and female students, this model separately tested for each subsample. Confirmatory factor analyses are reported in Table 3, which suggested a good to reasonable fit for both male and female students when tested separately.I then proceeded to test a configurable model in which the factor structure is fitted to both groups simultaneously. The reasonable model fit here indicates that items are indicators of the same factors in both males and female subsamples. I then tested a model in which factor loadings were constrained to be equivalent across both groups (metric invariance) was then tested.

Table 2
Factor loadings and reliability coefficients for the whole model and for gender subsamples

	Total Sample		Female subsample		Male subsample	
	W	E	W	E	W	E
3. Thinking about my grade in a course interferes with my work on tests	.55		.49		.65	
4. I freeze up on important exams	.46		.56		.44	
5. During exams I find myself thinking about whether I'll ever get through school	.13		.12		.18	
6. The harder I work at a test, the more confused I get	.90		.73		.89	
7. Thoughts of doing poorly interfere with my concentration of tests	.39		.42		.43	
14. I seem to defeat myself while working on important tests	.68		.63		.75	
17. During tests I find myself thinking about the consequences of failing	.53		.55		.58	
20. During examinations I get so nervous that I forget facts I know	.85		.81		.89	
2. While taking exams I have an uneasy upset feeling		.34		.30		.36
8. I feel very jittery while taking an important test		.71		.70		.67
9. Even when I'm well prepared for a test, I feel very nervous about it		.42		.42		.40
10. I start feeling very uneasy just before getting a test paper back		.44		.42		.44
11. During tests I feel very tense		.33		.38		.29
15. I feel panicky when I take an important test		.90		.81		.94
16. I worry a great deal before taking an important examination		.82		.71		.88
18. I feel my heart beating very fast during important tests		.43		.43		.39
Cronbach's α	.68	.81	.70	.82	.67	.78

Although tested models may be compared by examining $\Delta\chi^2$, as this statistic is sensitive to sample size and sample was relatively large, I used ΔCFI an alternative, where a ΔCFI ≤ .01 indicates invariance (Cheung & Rensvold, 2002). The ΔCFI=.003 between the configurable and metric invariance models indicates that the factor loadings are equivalent in the male and female subsamples. Lastly, a model was tested in which the variances and covariance were constrained to be equivalent across both groups. The ΔCFI <.001 between the latter two models indicates that variances and covariance are equivalent between in the male and female subsamples.

Table 3
Tests of factorial invariance

Model	χ^2	df	RMSEA	CFI	TFI
Female students	258.13***	103	.041	.955	.947
Male students	346.45***	103	.050	.922	.909
Configural Model	622.63***	206	.033	.939	.929
Metric invariance	646.07***	220	.032	.937	.932
Construct variance and invariance	649.80***	223	.032	.937	.932

In summary, it was demonstrated that the two-factor first order model of worry and emotionality with eight items each is equivalent for male and female students, and between group differences, can be examined. Factor loadings for the male and female subsamples are reported in Table 2. Low factor loadings (<.4) are again reported for one worry item (item 5) and two emotionality items (items 2 and 11) in both male and female subsamples and also for an additional emotionality item (item 18) in the male subsample. Reliability coefficients are acceptable (α >.7) for the emotionality factor and marginally (α ≥ .67) under for the total sample and male subsample.

Gender Differences

A one-way between-participants multivariate analysis of variance was conducted with gender as between-participants factor and worry and emotionality scores, and a total TAI score (comprised of all 20 questions: Total sample α = .85; female α = .87; male α = .84) as the dependent variables. The omnibus test indicated significant gender differences: Λ = .94, $F(3,1181)$ = 41.62, p <.001 and so univariate analyses were followed up separately for each dependent variable (descriptive statistics are reported in Table 4). Female students reported small but significantly higher TAI total (F = 44.03, p<.001, η_p^2= .02) and emotionality scores (F = 71.58, p<.001, η_p^2= .04) but not worry scores (F = 3.42, p =.07, η_p^2< .01).

Table 4
Descriptive statistics for TAI scores by gender

	TAI Total Scores		Worry		Emotionality	
	M	*SD*	*M*	*SD*	*M*	*SD*
Whole Sample	42.58	10.07	16.11	4.53	17.67	5.09
Female students	44.20	10.33	16.32	4.64	18.71	5.11
Male students	41.14	9.62	15.93	4.41	16.76	4.90

Bivariate Correlations with Educational Performance

Reported in Table 5, test anxiety shows significant inverse relations with academic performance which are significantly stronger in the worry than the emotionality component for aggregated performance (z = -4.23, p<.001), math's (z = -3.50, p<.001), physics (z = -4.46, p<.001) and biology (z = -3.19, p<.001). Significant intercorrelations are reported for the worry and emotionality components of test anxiety, which also correlate strongly with the total score, and academic performance in math, physics, chemistry and biology.

Table 5
Bivariate correlations between TAI scores and educational performance

	2.	3.	4.	5.	6.	7.	8.
1. TAI Total	.89	.92	-.22	-.20	-.20	-.20	-.25
2. Worry	--	.67	-.26	-.24	-.26	-.24	-.30
3. Emotionality		--	-.15	-.13	-.12	-.14	-.17
4. Maths			--	.61	.63	.57	.81
5. Physics				--	.69	.58	.75
6. Chemistry					--	.70	.83
7. Biology						--	.78
8.Aggregated Grade							--

All relations significant at p<.01

Discussion

The aim of this study was to translate the TAI into Urdu and then examine the factorial validity, reliability, discriminant validity in a sample of Pakistani students, along with gender differences in TAI and component scores. A two-factor model of the U-TAI, based on the worry and emotionality components showed an acceptable model fit and internal reliability. Furthermore, this factor structure was shown to be equivalent for male and female students. As expected, female students reported significantly higher test anxiety scores,

which are attributable to differences on the emotionality component only. Also, consistent with our prediction, a small but significant, inverse relationship was reported between test anxiety and performance in math and science school leaving examinations. Evidence of discriminative validity was also shown through the significantly stronger relations with performance reported for the U-TAI worry scale. Thus, the results were satisfactory, showing sufficient validity and reliability to be used in future research with confidence.

It was, however, considered that the validation process was incomplete. Low factor loadings were reported for several items, suggesting that these items may not be as relevant to the Pakistani context. Further work may wish to examine the usability of these items and whether they could be replaced with more appropriate items. Having established that the fundamental cognitive and affective-physiological factors have been demonstrated with our sample of Pakistani students, future work may also wish to examine whether the test anxiety construct and domain should be expanded to include other components. A fear of being judged negatively by others (such as peers, parents and teachers) has been included in more recent test anxiety measures (Bodas, Ollendick,& Sovani, 2008; Friedman & Bendas-Jacob, 1997), often labelled as social derogation. Before such additional components are added, at the risk of imposing an inappropriate construct from one culture to another, preliminary work is required to establish which constructs are relevant to the host culture and what those domains might consist of (Bodas et al., 2008).

The findings for gender differences are consistent with those previously reported in the literature, but do not add to the weight of evidence for the presentation of socialization explanations. Future work may then wish to explore the possibility of measuring test anxiety via an implicit association task, used to examine gender differences in trait anxiety (Egloff & Scmuckle, 2004), which are less prone to presentation bias. If gender differences in test anxiety remained as an implicit association task, presentational effects could be ruled out. The researcher's findings for the test anxiety and examination performance relationship were also consistent with previous work. The interfering role of worry has been long established, however recent advances afforded by attentional control theory have allowed a much more specific understanding of how anxiety influences working memory processes (Derakshan, Ansari, Shoker, Hansard,& Eysenck, 2009; Eysenck, Santos, Derekeshan,& Calvo, 2007). Building on Owens et al. (2008) this work could be usefully extended to investigate specific hypotheses about the influence of test anxiety on educational performance and achievement, through diminished working memory capacity and functioning. The possibility is also highlighted that training students to improve working memory capacity (Gathercole & Packiam-Alloway, 2008) might prove effective in ameliorating the negative impact of anxiety on performance, and therefore become a useful intervention for students with high test anxiety.

As already highlighted, the principal weakness of this study is that as a replication study concerned with validation of the U-TAI, it does not advance theory. The research is useful however, in providing an instrument to measure test anxiety, which can be used with Urdu speaking persons. In summary, our

study has translated the TAI into Urdu and established the factorial validity and internal reliability of that measure. We have also demonstrated how this measure shows the expected gender differences and relations with educational performance (this demonstrating divergent validity), and identified ways in which the U-TAI could be used to extend the extant literature.

References

Bodas, J., Ollendick, T. H., & Sovani, A.V. (2008). Test anxiety in Indian children: a cross-cultural perspective. *Anxiety, Stress and Coping, 21*(4), 387-404. Doi 10.1080/10615800701849902

Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology, 97*, 268-274. DOI: 10.1037/0022-0663.97.2.268

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of fit indices for testing measurement invariance. *Structural Equation Modelling, 9*(2), 233-255. Doi: 10.1207/S15328007SEM0902_5

Derakshan, N., & Eysenck, M. W. (2009). Anxiety, processing efficiency and cognitive performance: New developments from attentional control theory. *European Psychologist*, *14*(2), 168−176. doi:10.1027/1016-9040.14.2.168

Derakshan, N., Ansari, T. L., Shoker, L., Hansard, M. E., & Eysenck, M. W. (2009). Anxiety, inhibition, efficiency, and effectiveness: An investigation using the antisaccade task. *Experimental Psychology, 56*(1), 48-55. doi:10.1080/02699930903412120.

Eysenck, M. W., Santos, R., Derekeshan, N., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. doi:10.1037/1528-3542.7.2.336

Egloff, B., & Scmuckle, S.C. (2004). Gender differences in explicit and implicit anxiety measures. *Personality and Individual Differences, 36*(8), 1807-1815. Doi: 10.1016/j.paid.2003.07.002

Fairclough, S. H., Tattersall, A.J., & Houston, K. (2006). Anxiety and performance in the British driving test. *Transportation Research, Part F, 9*(1), 43-52. Doi: doi:10.1016/j.trf.2005.08.004

Friedman, I. A., & Bendas-Jacob, O. (1997). Measuring perceived test anxiety in adolescents: A self-report scale. *Educational and Psychological Measurement*, *57*(6), 1035–1046. Doi:10.1177/0013164497057006012

Gathercole, S., & Packiam-Alloway, T. (2008). *Working Memory and Learning: A Practical Guide for Teachers*. London: Sage.

Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research*, *58*(1), 47–77. doi:10.3102/00346543058001047

Lowe, P. A., Lee, S. W., Witteborg, K. M., Pritchard, K.W., Luhr, M.E., Cullinan, C.M., . . .,Janik, M. (2008). The Test Anxiety Inventory for Children and Adolescent. *Journal of Psycheducational Assessment*, *26*(3), 215-230. Doi: 10.1177/0734282907303760

Marsh, H. W., Hau, K. T.,& Wen, Z., (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cut off values for fit indexes and dangers in over generalising Hu & Bentler's (1999) findings. *Structural Equation Modelling, 11*(3), 320-341. Doi: 10.1207/s15328007sem1103_2

Marsh, H. W., Hau, K.T.,& Grayson, D. (2005). Goodness of Fit Evaluation in Structural Equation Modelling. In A. Maydeu-Olivares and J. McArdle (Eds.) *Contemporary Psychometrics. A Festschrift for Roderick P. McDonald* (pp. 275-340). Mahwah NJ: Erlbaum.

O'Neil, H. F., & Fukumura, T. (1992). Relationship of worry and emotionality to test performance in a juku environment. *Anxiety, Stress and Coping, 5*(3), 241-151.Doi: 10.1080/10615809208249525

Owens, M., Stevenson, J., Norgate, R., & Hadwin, J. A. (2008). Processing efficiency theory in children: Working memory as a mediator between test anxiety and academic performance. *Anxiety, Stress and Coping*, *21*(4), 417–430. doi:10.1080/10615800701847823

Putwain, D. W. (2008a). Deconstructing test anxiety. *Emotional and Behavioural Difficulties*, *13*(2), 141-155. Doi: 10.1080/13632750802027713

Putwain, D. W. (2008b). Test anxiety and academic performance in KS4. *Educational Psychology in Practice*, *24*(4), 319-334. Doi: 10.1080/02667360802488765

Putwain, D.W., Connors, E., & Symes, W. (2010). Do cognitive distortions mediate the test anxiety and examination performance relationship? *Educational Psychology*, *30*(1), 11-26. Doi: 10.1080/01443410903328866

Sarason, I. G. (1984). Stress, anxiety and cognitive interferences: Reactions to tests. *Journal of Abnormal and Social Psychology, 46,* 929-938.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of recent findings. *Anxiety, Stress and Coping 4*(1), 27–41.Doi: 10.1080/08917779108248762 [when this article was originally published, the journal title was *Anxiety Research*]

Seipp, B., & Schwarzer, C. (1996). Cross-cultural anxiety research: A review. In C. Schwarzer & M. Zeidner (Eds.), *Stress, anxiety and coping in academic settings* (pp. 13-68). Tubingen, Germany: Francke-Verlag.

Spielberger, C. D. (1980). *Test Anxiety Inventory: Preliminary Professional Manual*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D. (2006). Cross-cultural assessment of emotion states and personality traits. *European Psychologist, 11*(4), 297-303. Doi: 10.1027/1016-9040.11.4.297

Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: a transactional process model. In C. D. Speilberger & P. R. Vagg (Eds.) *Test anxiety: Theory, Assessment and Treatment* (pp. 3-14). Bristol: Taylor & Frances.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Algaze, B., & Anton, W. D. (1978). Examination stress and test anxiety. In C. D. Spielberger & I. G. Sarason (Eds.). *Stress and Anxiety* (Vol. 5). New York: Hemisphere/Wiley.

Ullah, R., Richardson, J. T .E., & Hafeez, M. (2011). Approaches to studying and perceptions of the academic environment among university students in

Pakistan. Compare: A Journal of Comparative and International Education, 41:1, 11 127. Doi: 10.1080/03057921003647065

Zeidner, M. (1990). Does test anxiety bias scholastic aptitude test performance by gender and social group? *Journal of Personality Assessment, 55*(1-2), 145–160.Doi: 10.1080/00223891.1990.9674054

Zeidner, M., & Nevo, B. (1992). Test anxiety in examinees in a college admission: Incidence, dimensionality and cognitive correlates. In K. A. Hagvet & B. T. Johnsen (Eds.), Advances in test anxiety research, Vol. 7 (pp. 288-303). Lisse, The Netherlands: Swets and Zeitlinger.

Zeidner, M., & Schleyer, E. J. (1999). Test anxiety in intellectually gifted students. *Anxiety, Stress and Coping, 12*(2), 163–189.Doi: 10.1080/10615809908248328

Zeidner, M., & Mathews, G. (2005).Evaluation anxiety. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141-163). London: Guildford Press.

A Needs Assessment in the Conduct of Program Evaluation in Schools

Carlo Magno
De La Salle University, Manila, Philippines

The present report surveyed selected schools in one province in the Philippines to determine their needs in evaluation and how evaluation is currently being practiced. Very few schools engage in the real concept of evaluation, which brought the attention to assess the practices and needs of schools in the area of evaluation. There is a need to study how schools undertake their evaluation of programs in order to identify the problems encountered by schools and to target specific ways to improve them. Based on the results, the needs of the schools include: (1) to improve teacher performance and teacher training evaluation, (2) to consider alternative ways of evaluation such as focus group discussions and surveys, (3) training on instrumentation as a technical skill, (4) seminars and workshops that provide venues for learning about the process of evaluation, (5) improve the personnel, practice, relational, and study aspects of evaluation in schools.

Keywords: Needs Assessment, Needs of Schools, Program Evaluation

Schools offer different programs to enhance the teaching and the learning process. One way to determine whether certain educational programs needs to be improved or is effective is through proper evaluation. The information generated through proper evaluation is important since it can aid practitioners in the ministry or department of education for undertaking various decisions and plans to make improvements and decisions regarding educational programs.Although very few schools engage in the real concept of evaluation. This claim is justified by assessing the practices and needs of schools in line with proper evaluation. Personnel involved in running educational programs

Personnel involved in running educational programs have different viewpoints and misconceptions of program evaluation. According to Fitzpatrick, Worthen, and Sanders (2004) some authors opt for a systems approach, while others view evaluation as a process of identifying and collecting information to assist decision makers. Others view evaluation as synonymous with professional judgment, where judgment of a program's quality is based on opinion of experts. In one school of thought, evaluation is viewed as the process of comparing performance data with clearly specified objectives, while in another evaluation is seen as synonymous with carefully controlled experimental research on programs. Others urge the importance of naturalistic inquiry or urge that value pluralism be recognized where the individuals evaluated play a prime role in determining what direction the evaluation study takes. These various points of view on evaluation bring about differences in opinion on how evaluation is suppose to be done but the worst are misconceptions on handling evaluations. Even if there are various ways in the conduct of evaluation, a key result area that needs to be obtained is the end goal of evaluation, which is to come up with a judgment and an overall value of a program and its relative value. The American Evaluation Association made effort to generate the "Guiding Principles for Evaluation" to properly guide evaluators. These guide entails (a) that inquiries should be data-based whatever is evaluated, (b) evaluators needs to be competent, (c) honesty and integrity is needed in the process, (d) respect the security, dignity, and self-worth of the respondent, and (e) the articulation of responsibilities for the general and public welfare. Schools conducting the process of evaluation must be aware of these guiding principles needs to be placed into proper perspectives.

There is a need to study how schools undertake their evaluation and their needs in order to assess the problems faced by schools and to generate specific ways to improve them especially in provincial areas of the Philippines where students achievement are low. There is also a need to investigate how the practice of evaluation is being conducted in the provincial area to see how evaluation is conceptualized in different contexts. This report surveyed selected schools in the one provincial area in the Philippines to determine their needs in evaluation and how evaluation is being practiced currently.

The Need for Initial Assessment

Before a program is started, institutions would conduct needs assessment. An evaluation of a need seeks to identify and measure the level of unmet needs within an organization (Posavac & Carey, 2003). The planning would then be guided by answering the needs that arises. Astin (1993) expresses the need for schools to evaluate regularly since the quality of education provided to students as well as services are based on it. The quality of education likewise is measured through evaluation. Conducting needs assessment, is concerned whether a problem or a need exist and to describe the problem of a program and then make recommendations for ways to reduce the problem. Basically needs assessment determines whether there is a sufficient need existing to initiate a

program and if there is, then there needs to be assistance in program planning by identifying potential program models and activities that might be conducted to achieve goals. McKillip (1998) identified five processes on needs analysis, which includes: (1) identification of users and uses, (2) description of the target population and service environment, (3) need identification, (4) needs assessment, and (5) communication. McKillip (1998) explains that needs assessment is to produce recommendations for action. Different studies may focus on the processing and monitoring a component of a program. Such studies focuses on whether the program is being delivered according to some delineated plan or model or may be more open–ended, simply describing the nature of delivery of the successes and problems encountered (Fitzpatrick, Worthen, & Sanders, 2004).

School wide Assessment

Educators need to pay attention to some alarms that might be sounding in their school and districts as they seek accountability through assessment. The alarms are associated with three extremely important areas of schooling; namely, the quality of instruction taking place in the classroom, the professional development required of teachers and provided by schools and districts, and the ethical standards expected of students and teachers (Ferrera, 2005). Vigilant administrators who are sincerely seeking effective assessment methods to be more accountable should take note of the following:

(1) Quality of instruction - Linda Darling-Hammond's (2005) research suggests that broader assessments actually raise standards and achievement. "The ability to make effective oral arguments and conduct significant research projects are considered essential skills by both employers and postsecondary educators ... these skills are very difficult to measure on a paper-and-pencil test" (Darling-Hammond, 2005). The issue is obviously the amount of instruction that seems one-sided. Balance of instructional strategies seems to be necessary for the quality of instruction to be high.

(2) Professional development - A recent report sponsored by the Spencer Foundation revealed that professional development activities that focused on accountability systems appeared to help teachers focus more on standards, but often resulted in misunderstood test data and rankings. Even more alarming were the "few organizational mechanisms that linked usable student performance data to teacher learning opportunities" (Berry, Turchi, Johnson, Hare, Owens, 2003).

(3) Ethical behavior - The Education Commission of the States has reported an increasing number of children suffering from sleep disorders and other stress-related maladies as a result of high-stakes testing systems connected with rigorous accountability efforts (Dounay, 2000). There is concern that increased cheating is the inevitable result of this pressure.

Assessment and Accountability

In the study reported by Morre, Dexter, Berube, and Beck (2005), they reviewed the literature on accountability and assessment in order to design a questionnaire to survey superintendents across Wyoming on their existing and needed knowledge about student assessment. Results on the presence or absence of gaps in knowledge that is deemed important by respondents and/or the literature can be used by colleges and universities in Wyoming, and perhaps elsewhere, to improve superintendent certification programs. Their report explained that the ability to plan assessment systems, to implement data-based decision making, to improve the classroom assessment used by teachers, and to communicate student assessment data requires technical knowledge in the area of student assessment. Arter, Stiggins, Duke, and Sagor (1993) articulated a set of 12 assessment competencies for principals and, by extension, superintendents. According to Arter et al. (1993), these instructional leaders should:

1. Know the attributes of sound student assessment and how to apply them to the assessments used in the school building;
2. Know the attributes of a sound student assessment system and how to apply them to the assessment systems used in the building;
3. Know issues related to ethical and inappropriate use of assessment information and how to protect students and staff from misuses;
4. Know the importance and features of assessment policies and regulations that contribute to the development and use of sound assessments at all levels of use;
5. Know the importance of and be able to work with staff members to set specific goals for integration of assessment into instruction, and to assist teachers in reaching those goals;
6. Know the importance of and be able to evaluate teachers' classroom assessment competencies and build such evaluations into the supervision process;
7. Know the importance of and be able to plan and present, or secure the presentation of, staff development experiences that contribute to the development and use of sound assessment at all levels of decision making;
8. Know the importance of and how to use assessment results for instructional improvement at the building level;
9. Know how to accurately analyze and interpret building-level assessment information;
10. Be able to act effectively upon assessment information;
11. Know and create the conditions necessary for the appropriate use of achievement information; and
12. Be able to communicate effectively with all interested members of the school community about assessment results and their relationship to instruction, (p. 5)

Warna (1995) stressed the importance of assessment to improve school performance. School administrators need to work with stakeholders in the evaluation process. Stakeholders are defined as individuals whose lives are affected by the program and whose decisions affect the future of the program (Greene, 1988). For educators this provides the broadest possible definition of school community, including as it does students, teachers, administrators, parents, school board members, central office personnel, and county residents. Stakeholder participation in the formal evaluation process may be fostered through individual interviews, group meetings, questionnaires, and open-ended feedback.

In a report by Duran (2005) an overview of factors to consider when evaluating the validity and reliability of interpretations and uses of results used for the purpose of complying with the No Child Left Behind (NCLB) Act.1 A number of factors are identified and used to examine current interpretations and uses of assessment results for purposes of accountability. A concern is that sanctions and consequences may be imposed on schools through the use of invalid and unreliable results. Specific NCLB Act (NCLB) requirements are identified and used to examine this claim. The requirements include: The development and implementation of content and performance standards and standards-based assessments; Adequate Yearly Progress (AYP) with a focus on potential negative impact based on immediate implementation and vague operational definitions; and sanctions and consequences. Clarification of interpretations and uses of results is provided to develop a better understanding for stakeholders who are responsible for making policy and educational decisions. In conclusion, the author suggests that the NCLB Act creates an opportunity for all states to develop and implement valid and reliable accountability systems that clearly and accurately identify effective schools and also provide adequate support to schools in need of improvement so that all students are able to receive quality and effective instruction that improves academic achievement and ultimately allows for students to reach their full potential.

Harada (2005) reports that evaluation is necessary in schools because it targets four goals of improvement. The function of assessment includes empowering student learning, informing instructional effectiveness, and communicating evidence of learning to parents, and winning support from administrators.

(1) Empowering Student Learning - By engaging students in assessment, we invite students to reflect on their own progress. Students more clearly understand what is expected. They connect new ideas to prior knowledge and strengthen their ownership over making the learning happen. Assessment also provides them with critical opportunities to give descriptive feedback as they are learning (Davies, 2000).

(2) Informing Instructional Effectiveness - Assessment provides the instructional team-classroom teachers, teacher-librarians, and additional school partners-with crucial information on what students are learning and how teaching might be shaped to help students do even better. Assessment provides

a map for planning curriculum and instructional activities (Harada & Yoshina, 2005). The result is more opportunities for peer learning and collaboration, more choices for students in the learning environment, and more integrated and interdisciplinary teaching (Falk, 2000).

(3) Communicating Evidence of Learning To Parents - While parents are interested in their children's scores on norm-referenced, standardized tests, they are also grateful for more personalized information that shows specific examples of what their children are actually learning. If students are creating their own learning portfolios, they include samples of their work; assessments of the samples; and reflections about what they learned, how they learned it, and what future directions they wish to pursue (Harada & Yoshina, 2005).

(4) Winning Support from Administrators - School leaders are besieged with much to do and limited resources and little time with which to do it. When they have to make decisions about allocating funds and staffing positions, they want evidence built on systematically collected data to make their determinations. They also need the evidence in capsulated formats. Providing this type of documentation builds a compelling case for the value of the library program. In short, communicating evidence of what is being learned through library instruction is a vital tool for library advocacy.

Supervision and Evaluation

Supervision is supposed to improve classroom teaching by enhancing teacher thinking, reflection, and understanding of teaching. Evaluation systems are supposed to increase effective teaching behaviors and enhance teacher professionalism. In a context of increased accountability-based, evaluation systems of all kinds, principals and teachers need to practice supervision and evaluation that facilitates meaningful adult learning (Ponticell & Zepeda, 2004). The result of their study indicates that for both teachers and principals, supervision was evaluation. Principals conducted evaluations, generally once a year, following the steps required by law: some kind of pre-observation conference, an observation based on an observation checklist, a post-observation conference, and signing and filing an official form. Teachers put on the required show, trying to display all the items on the observation checklist. As a result, neither supervision nor evaluation is done particularly well. Emphasis on the legal requirements of supervision and evaluation places principals in a hierarchical position over teachers. The principal determines the time, nature, and extent of supervision and evaluation. The principal observes, monitors, and checks the teacher. The principal informs the teacher regarding what needs to be changed or improved. The principal determines if improvement has occurred or progress has been made. The principal judges the effectiveness of a teacher's performance and assigns a performance rating. The teacher is subordinate to the principal. The teacher performs the show when required to do so by the principal, listens to the principal's judgment of the teacher's effectiveness, accepts the principal's judgment, and implements the principal's recommendations for change.

In the same way Bernstein (2004) expresses the same sentiments on supervision and evaluation. The explained that evaluation should be intended to support teacher growth, and evaluation should enhance teacher professionalism. The National Board for Professional Teaching Standards (NBPTS) started the initiative propositions on policy of what teachers should know and be able to do. The first three propositions are very much related to evaluation. These propositions include "supervision and evaluation procedures are committed to teacher growth." The proposition further explains that supervision and evaluation procedures that respect individual differences in teachers and can adjust accordingly must be based on an understanding of adult learning. Practices such as peer coaching, action research, and mentoring have that capacity. Using those practices, supervision is removed from the sole purview of the school administrator and shared among professional peers. By using a variety of such programs, teachers can be treated equitably without necessarily being treated equally. Stages of teacher development can be taken into account as trust with peers is established, feedback is relevant, and collaborative time is focused on the work of teaching. The second proposition states "supervision and evaluation supports teachers to learn content and employ a wide variety of pedagogical techniques." Supervision and evaluation procedures must, acknowledge that there is no one way to teach a subject to an individual child. Schools and teachers dependent on prescriptive programs should be encouraged to branch out. Schools and districts must create easy access for teachers to work with and observe their peers, to learn new curricular knowledge, and to enlarge their instructional repertoire. Teachers must have access to various forms of support from district and school administrators to take advantage of multiple paths of developing teacher knowledge. The third proposition states "supervision and evaluation procedures are responsible for managing and monitoring teacher growth." As teachers are engaged with peers and principals in supervision and evaluation, their learning must be relevant and link to past experiences. There must be support and teachers must be able to enlist their colleagues' knowledge and expertise. Accomplished peers and supervisors must command a range of techniques and programs and provide teachers with choices about when each is appropriate.

The Present Study

The body of literature in the conduct of proper evaluation and best practices are rich. Educators in other countries are also conscious on the use of evaluation and assessment results especially school administrators. They see the value of assessment to improve their practices and to aid them in making better decisions. There are no information available in the Philippine setting as to how educators and school personnel conducts evaluation. There are several reports in the evaluation of programs from external funding agencies but the process on how schools go about in evaluating their programs are not yet available in literature. The present study aims to conduct a needs assessment of schools in

the area of program evaluation in a province in the Southern Luzon region in the Philippines. The present study seeks to answer the following research questions:

1. Who commonly conducts the evaluation on the schools?
2. What activities are commonly evaluated on the schools?
3. What activities need evaluation improvement in the schools?
4. Who commonly handles the evaluation on the schools?
5. What are the different ways of gathering data for the purpose of evaluation among the schools?
6. What are the technical needs of schools in evaluation?
7. Is there a significant difference between private and public schools on their technical needs in evaluation?
8. What are the sources of information that schools use in coming up with an evaluation?
9. What are the ways to improve the practice of evaluation on schools?

Method

Participants

There are 37 respondents who participated in the study coming form 33 different institutions. There are 21 participants from private schools and 16 from the public schools. The position of the respondents varied that includes principals, directors, administrators, assistant principal, coordinators, and guidance counselors. The participants belong to the upper and middle level management positions.

Instrument

A needs assessment inventory focusing on the needs of schools in the area of evaluation of curricular programs was constructed. The inventory is composed of eight items that reflects how evaluation is practiced. It also identified the needs on various aspects of evaluation (See Appendix). The items identifies (1) whether evaluation is being conducted in the school, (2) activities evaluated, (3) activities the needs improvement, (4) personnel who handles the evaluation, (5) how data is gathered during evaluation, (6) areas in evaluation that needs technical assistance, (7) information where evaluation is taken, and (8) suggestions how to improve the evaluation.

A professor handling an evaluation course reviewed the items of the needs assessment inventory and it was revised for improvement. The inventory was pretested with few respondents to determine if the items are comprehensible.

Procedure

The instrument was constructed for the purpose of surveying the needs of schools on evaluation. The survey was reviewed and revised. It was administered to principals, assistant principals, coordinators, and other school heads from different schools in two provinces in the Southern Luzon Region of the Philippines. The participants were given a letter and the purpose of the needs assessment was explained to them. The administrators were asked to accomplish the survey form. Some surveys forms were answered immediately and others needed time to answer and requested to be retrieved in another time. After accomplishing the forms, the school heads were again debriefed about the purpose of the assessment.

Data Analysis

The responses were coded as 1 if the item is selected and 0 for those items that were not selected. In analyzing the data, the items with multiple responses were converted into percentage. The number of responses was counted and it was divided with the total number of responses and multiplied with 100 to approximate its proportion. For the item on technical needs on evaluation, the means and standard deviations were obtained. To interpret the weighted mean for each item the following range was used: 1.0–1.74 (Don't Know), 1.75–2.49 (Low Need), 2.5–3.24 (Moderate Need), 3.5–4.00 (High Need).

The public and private schools were compared on their technical needs on evaluation using the t-test for two independent samples. The alpha level of significance was set at .05.

The comments and suggestions on how to improve the evaluation process in schools were analyzed qualitatively using cluster analysis. The responses were grouped into emerging clusters.

Results

The responses for each item in the survey were counted by frequencies and converted into percentage for each item for conducting school evaluation, activities that schools evaluate, activities that needs evaluation improvement, persons who conduct the evaluation, data gathering procedures for evaluation, technical needs, sources of information, and suggestions to improve the practices of evaluation.

Most schools reported that they conduct evaluation (89.2%) and all of them are from the private schools. Four public schools (10.8%) indicated not conducting evaluations.

Table 1
Percentage of Conducting Evaluation in Schools

	Public		Private		Total	
	f	%	f	%	f	%
Evaluation is Conducted	12	75%	21	100%	33	89.2%
Evaluation is not conducted	4	25%	0	0%	4	10.8%
N	16	100%	21	100%	37	100.0%

Note. N=37

The activity that is mostly evaluated in schools is the teachers' performance (81.19%) since the teaching is one of the primary functions of a school. Then it is followed by teacher training (81.08%) which again refers to teacher performance. A large percentage is also evaluated on selecting students for academic and special awards (72.97%). The service feedback (43.24%) and student publications (48.64%) have lowest percentage being evaluated since these are not common in most schools.

Table 2
Activities that Schools Evaluate

Activities	Frequency	Percent
Teacher Performance	33	81.19%
Administrative Performance *(e.g., coordinator, principal, director, etc.)*	20	54.05%
Support Staff Performance	22	59.45%
Implementation of New Academic Programs	22	59.45%
Teacher Training Programs *(e.g., seminars, symposia, etc.)*	30	81.08%
Selecting Students for Academic and Special Awards	27	72.97%
Guidance and Counseling Programs	22	59.45%
Homeroom Guidance Program	21	56.76%
Administrative Services *(e.g., maintenance, engineering, accounting, etc.)*	19	51.35%
Student Organizations	25	67.57%
Student Publications	18	48.64%
Sports Development Program	22	59.45%
Cultural Activities	23	62.16%
Community Service	19	51.35%
Retreat, Recollection and other Formation Programs	21	56.75%
Service feedback	16	43.24%
Canteen/Cafeteria Evaluation	26	70.27%
Others, *please specify:*	1	2.70%
Total number of responses	387	

Note. N=37

Most schools reported that the activity selected that needs to be improved is on teacher training (37.83%) followed by teacher performance (32.43%). These two activities are the ones evaluated the most but the way they are being implemented needs improvement. There is a low percentage on service feedback (10.81%) since it is not mostly done in schools and still needs evaluation improvement.

Table 3
Activities that Needs Evaluation Improvement

Activities	Frequency	Percentage
Teacher Performance	12	32.43%
Administrative Performance *(e.g., coordinator, principal, director, etc.)*	9	24.32%
Support Staff Performance	7	18.92%
Implementation of New Academic Programs	10	27.02%
Teacher Training Programs *(e.g., seminars, symposia, etc.)*	14	37.83%
Selecting Students for Academic and Special Awards	5	13.51%
Guidance and Counseling Programs	9	24.32%
Homeroom Guidance Program	6	16.22%
Administrative Services *(e.g., maintenance, engineering, accounting, etc.)*	8	21.62%
Student Organizations	5	13.51%
Student Publications	5	13.51%
Sports Development Program	8	21.62%
Cultural Activities	6	16.22%
Community Service	8	21.62%
Retreat, Recollection and other Formation Programs	7	18.92%
Service feedback	4	10.81%
Canteen/Cafeteria Evaluation	10	27.03%
Others	0	0%
Total number of responses	133	

Note: N=37

The principal is mostly the one conducting the evaluation in schools (64.86%). This can be attributed to the role of principal as maintaining and monitoring the quality of schools, which makes evaluation part of the responsibility. High percentage is also obtained for the coordinators (40.54%) and guidance counselors conducting the evaluation (48.14%). The school president (2.70%), human resource (2.70%) and evaluation committee (2.70%) were reported of not conducting the evaluation in most schools.

Table 4
Persons who Conduct Evaluations

Persons	Frequency	Percentage
Guidance Counselor	13	48.14%
Teachers	11	29.72%
Principal	24	64.86%
Coordinators	15	40.54%
Assistant Principal	4	10.81%
Division Supervisor	4	10.81%
School President	1	2.70%
Director	4	10.81%
Human Resources	1	2.70%
Evaluation Committee	1	2.70%
Total number of responses	82	

Note. N=37

Observation is the most common data gathering procedure for obtaining evaluation data (83.78%) in most schools. There are several schools that use inventories and questionnaires (78.37%) when conducting evaluation. A large percentage also uses interviews (56.76%) and tests (56.76%) when evaluating. Very few of the schools use experiments (16.22%) and surveys (24.32%) when gathering evaluation data.

Table 5
Data Gathering Procedures for Evaluations

Procedures	Frequency	Percentage
Inventory/Questionnaires	29	78.37%
Focus Group Discussion	11	29.73%
Surveys	9	24.32%
Personal Interview	21	56.76%
Observation	31	83.78%
Tests	21	56.76%
Experiments	6	16.22%
Total number of responses	137	

Note. N=37

The selected schools reported a high need on the instrumentation as part of the process of evaluation (*M*=3.05) since not anyone has the skill to construct items and use framework guided by the domain assessed. The schools selected report that they have a moderate need in the process of data analysis (*M*=2.97), report writing (*M*=2.76), utilization of results (*M*=2.70), and dissemination of results (*M*=2.70). There is also a moderate need on the planning (*M*=2.84) and conceptualizing (*M*=2.89) of evaluation. The total mean score of the participants is 2.84 with a standard deviation of 1.11, which is moderate.

Table 6
Technical Needs on Evaluation

	N	*M*	*SD*	Interpretation
Planning the evaluation	37	2.84	1.191	Moderate Need
Conceptualization	37	2.89	1.100	Moderate Need
Instrumentation *(constructing assessment forms, etc.)*	37	3.05	1.026	Moderate Need
Data Analysis	37	2.97	1.013	Moderate Need
Report Writing	37	2.76	1.065	Moderate Need
Utilization of results	37	2.70	1.222	Moderate Need
Dissemination of results	37	2.70	1.175	Moderate Need
Others	0			
Total	37	2.84	1.11	Moderate Need

Note. N=37, 1.0 – 1.74: Don't Know, 1.75 – 2.49: Low Need, 2.5 – 3.24: Moderate Needs, 3.5 – 4.00:High Need

The global mean scores of the public and private schools were obtained and compared using the t-test for two independent samples. The *p* value obtained is 0.549 is greater than the alpha level of significance, there is no significant difference between the private and public schools in their technical needs in evaluation. This shows that both the public (*M*=20.63) has the same level of technical need in evaluation with the private school (*M*=19.38).

Table 7
Comparison of Public and Private Schools in their Technical Needs in Evaluation

Type of school	*N*	*M*	*SD*	*SE*	*t*	*df*	p value
Public	16	20.63	5.94	1.49			
Private	21	19.38	6.40	1.40	0.604	35	0.549

Note: *p<.05

Most of the participants indicated that they learn information on evaluation through seminars and workshops (78.38%). A large percentage also indicated that discussion groups (67.56%) and the Internet (62.16%) are sources of obtaining information in conducting evaluation. Few schools rely on colloquia (8.11%) as a source of information on evaluation.

Table 8
Source of Information when conducting Evaluation

Source of Information	Frequency	Percentage
Books	21	56.75%
Journals	16	43.24%
Internet	23	62.16%
Discussion Groups	25	67.56%
School-Based Experts	21	56.76%
External Experts	12	32.43%
Seminars/Workshops	29	78.38%
Colloquia	3	8.11%
Inter-school Collaboration/Consortium	15	40.54%
Accreditation/Certification Documents/Manuals	14	37.83%
Total number of responses	179	

Note. N=37

The responses from the comments and suggestions on improving the evaluation process in schools were listed then clustered into emerging themes where each strand would fall under. The strands were grouped into four clusters: Personnel, practice, relational, and faculty development.

The cluster on personnel refers to the characteristics of people who will perform the evaluation which includes qualification, training, expertise, and open mindedness. In this category, the respondents mentioned that "evaluation must be done by a qualified personnel," "evaluators in schools must be trained how to do evaluation," "leaders should know if their performance is effective," "people who evaluate must be experts," and "the administrators needs to be open minded"

The practice refers to strands on consistency of evaluation, areas that needs to be improved and specific courses of actions. The respondents gave comments like "conduct needs assessment," "evaluation must be consistent in a regular basis," "evaluation results should be well utilized," "evaluation should be done with sincerity," "evaluation should be implemented properly," "evaluation should be planned highly," "hire extra personnel for such function," "lessen the load of guidance counselors to do evaluations," "minimize school contests to concentrate on evaluation," "needs conceptualization," "Provide materials for improvement," "should be objective rather than subjective," and to "use data as basis for improvement."

The relational cluster emerged where respondents indicated having a "discussion" and "involvement" would improve the evaluation practice.

The faculty development cluster is based on suggestions such as training, workshops and observations on how evaluation is done. This basically refers to learning the rigors of evaluation. The suggestions of the respondents include to "have seminars to make more teaching effective," "observation," "seminar

about question technique," "seminar on Table of Specifications." Most of the responses focus on giving seminars in this category.

Table 9
Suggestions to Improve the Practices of Evaluation

Cluster	Strands	f	%
Personnel	• Evaluation must be done by a qualified personnel	3	8.11
	• Evaluators in schools must be trained how to do eval.	1	0.27
	• Leaders should know if their performance is effective	1	0.27
	• People who evaluate must be experts	3	8.11
	• The administrators needs to be open minded	1	0.27
Practice	• Always do assessment	1	0.27
	• Conduct needs assessment	1	0.27
	• do it on a regular basis	1	0.27
	• evaluation must be consistent	1	0.27
	• evaluation results should be well utilized	1	0.27
	• evaluation should be done with sincerity	1	0.27
	• evaluation should be implemented properly	1	0.27
	• evaluation should be planned highly	1	0.27
	• hire extra personnel for such function	1	0.27
	• lessen the load of guidance counselors to do evaluations	1	0.27
	• minimize school contests to concentrate on evaluation	1	0.27
	• needs conceptualization	1	0.27
	• Provide materials for improvement	1	0.27
	• should be objective rather than subjective	1	0.27
	• use data as basis for improvement	1	0.27
Relational	• Group discussion	1	0.27
	• Involvement	1	0.27
Faculty	• Have seminars to make more teaching effective	1	0.27
Development	• observation	1	0.27
	• seminar about question technique	1	0.27
	• seminar on Table of Specifications	1	0.27
	• seminars and workshops	1	0.27

Discussion

The needs of schools include: (1) to improve teacher performance and teacher training evaluation, (2) to consider alternative ways of evaluation such as focus group discussions and surveys, (3) high need on instrumentation as a technical skill, (4) seminars and workshops are rich venues for learning about the process of evaluation, (5) need to improve the personnel, practice, relational, and study aspects of evaluation in schools.

Principals who Conduct Evaluation in Schools

In most of the schools, the principal is reported who conducts the evaluation processes for different activities. The evaluation process goes hand in hand with supervision and thus they are accountable to the performance and monitoring in their school. Given their accountability, evaluation is a tool for them to monitor and maintain their performance standards. Given the

accountability of administrators like the principal, Ferrera (2005) noted that administrators should be vigilant and needs to seek effective assessment methods to be more accountable. Morre, Dexter, Berube, and Beck (2005) further explained that school administrators should have the ability to plan assessment systems, to implement data-based decision making, to improve the classroom assessment used by teachers, and to communicate student assessment data requires technical knowledge in the area of student assessment. Arter, Stiggins, Duke, and Sagor (1993) in their study emphasized the assessment competencies set of school principals. The findings show that principals play a key role in the conduct of evaluation on schools.

Schools Mostly Evaluate Teacher's Performance

The teaching process as implemented by teachers is one of the most important factors in education. The effectiveness of schools and development of students' competencies depends on the teachers' ability to effectively carry out teaching. Given the importance of teaching in schools, it is the foremost factor to be evaluated. Most of the studies on school evaluation primarily emphasize teaching as the one of the components that is evaluated. According to Darling-Hammond (2005) that balance of instructional strategies is deemed necessary for the quality of instruction to be high in evaluation. In the standards set by Arter, Stiggins, Duke, and Sagor (1993) emphasized that evaluators should know the importance of integrating assessment into instruction and to assist teachers in reaching their goals.

Improving the Evaluation for Teacher Performance and Teacher Training

The teaching is one of the primary functions that need to be evaluated in schools because the effectiveness in the facilitation of learning depends on this. The quality of teaching as performed by teachers are usually assessed by superiors such as the principal, assistant principal, and coordinators to ensure the quality of teaching being delivered to the students. Although this activity is widely assessed, there is a need for schools to improve the way they evaluate and assess their teachers. O'Donnel (1996) identified three dilemmas when evaluating teachers. The first is about the confusion about the supervisory judgment based on process rather than outcome where teachers find it easy to observe the processes. Second is the tension and potential contradiction between evaluating for development and arriving with summative performance, which is a misconception. And lastly, the disagreement between supervisors and teachers as to what constitutes ideal teaching behavior. Evaluating teacher training goes hand in hand in evaluating teacher performance. Most commonly, teacher trainings are given but the practices are not sustained continuously by supervisors and teachers. The process of evaluating the continuity of teacher training entails a long process and should be done before putting other teacher training programs. The report of Harada (2005) and Ferrera (2005) explains the benefits of assessing teacher performance and teacher training. What is lacking

perhaps in Philippine schools is the setting of standards on evaluations and education. The practice of evaluation in schools would most likely be within the proper standards if they follow and adhere to the set of standards like the No child left behind, NBPTS, and AEA evaluation guiding principles.

Alternative Ways of Conducting Evaluation

Most commonly, teachers are observed inside the classroom through pop-in visits and some teachers do not feel comfortable especially the ones not prepared. The worst part is that a pop-in visit is used to serve as a summative evaluation for a teachers' performance. Aside from the crude ways of observation, there is a call to use alternative ways of evaluation such as focus group discussions which generated qualitative data and provides a rich source of information in determining what areas does the teaching and learning needs to be improved or maintained. Some schools are not accustomed to qualitative methods where they remain positivistic in their approach on evaluation. Hughes-Hassell and Bishop (2004) used focus group discussion to determine the needs of librarians in schools. They explained that focus group interviews offer several benefits. In one hour, a teacher can gather the ideas, views or opinions of six to eight people instead of only one person. The social interaction involved in the interview typically helps participants focus on the most important aspects of a topic or issue, thus getting to the core of an issue in less time. Plus, the format allows the interviewer to probe for clarification or solicit greater detail throughout the interview, thus enhancing the completeness of the data collected.

Instrumentation as a Technical Skill

Constructing scales and items that measure a construct needs a highly trained psychometrician who is not usually available in schools. The findings indicate this to be highly needed and a call on skilled personnel on test construction is needed. The undergraduate degree in psychological testing in some universities like De La Salle University, Philippine Normal University, Mirriam College and University of the Philippines are offering courses where students are trained to construct their own measures (Magno, 2010). The students need to continue to enhance their skills in the field of measurement and evaluation where it is mostly needed. Part of the standards of the guiding principle in evaluation is the competence of the evaluator. The competence entails appropriate education, ability, skill, and experience in the area of measurement and evaluation.

Seminars and Workshops as Source to Learn the Process of Evaluation

Most teachers in the basic education are sent to seminars and forums in order to update and gain new insights on the teaching and learning process. This is a good opportunity for them to learn new things as indicated in the survey.

Seminars can focus on the area on measurement and evaluation to train teachers and administrators the rigors of evaluation in schools and programs. It was reported in the Philippine Human Development (2000) that the capacities of colleges and universities to offer quality teacher-training programs are strained but the approach is still popular. As a result, teacher training has expanded rapidly only at great sacrifice of quality. It was even emphasized that even today there are only a handful of teacher training institutions that offer relatively high quality programs. There are only reported feeble attempts in later years to remedy the poor quality of teacher training as a result the quality is still poor.

The Need to Improve Personnel, Practice, Relational, and the Study of Evaluation

It is expressed in the qualitative data that there is need for trained personnel and practices of evaluation that should be improved in schools. It is also helpful if the staff engaged with each other to carefully plan how the evaluation is going to take place (relational). The faculty development aspect of evaluation opens the opportunity of teachers to be trained in the conduct of evaluation. These findings support the report of Berry, Turchi, Johnson, Hare, and Owens (2003) that there is few organizational mechanism that link usable students performance data to teacher learning opportunities. Because of this scenario, professional development activities that focused on accountability systems appeared to help teachers focus more on standards. Even professional development activities are given like seminars, these often resulted in misunderstood test data and rankings.

Implications

The major findings in the study show that most schools conduct their evaluation concentrating on teacher performance, however, the process needs to be improved. Most often, the principal is the one conducting the evaluation using inventories and questionnaires to gather evaluation information. Although, expertise in the use of instrumentation appears to be a great need. Schools commonly rely on seminars as intervention to update them on the trends in evaluation. The implication of the needs assessment conducted is that schools have not realized yet their need to improve their evaluation practices. This is primarily because of the low exposure of the school administrators and teachers on empirical studies and reports that provides updated information on the field of educational evaluation. Although, their practice of evaluation is concentrated on teachers' performance since it is a primary concern, their major belief is that faculty development will provide for whatever gaps in knowledge they have.

Based on the results of the needs assessment, it is concluded that schools have a need to improve the manner in conducting their evaluation in terms of the following:

1. The schools a need to improve the evaluation of their curricular programs in terms of teacher performance and teacher training.

2. There is a need to include all staff and personnel in schools especially the teachers to perform evaluation in various educational programs of the school.
3. The schools need to use other means of gathering qualitative data such as focus group discussions since they are only limited on quantitative approaches.
4. The schools need to source out experts and technical reports on the proper way of conducting evaluation.
5. The schools need further training to equip them in the conduct of proper evaluation especially in terms of instrumentation.

Given the results of the needs assessment, the following are recommended for the administrators and principals in the selected schools:

1. Conceptualize a set of standards in conducting evaluation in schools by collaborating with other schools in the same district.
2. Share resources in evaluation such as reports in evaluation studies and methodologies so that the procedures employed are not always traditional.
3. Encourage other school personnel to engage in evaluation activities like teachers and guidance counselors.
4. Establish a separate office that is concentrated on performing evaluation and assessment tasks for the school.
5. Be open to new methodologies, procedures, and principles in administration that would include a system of feedback and assessment.

The following are recommended for teachers and school staff in the schools:

1. Engage in evaluation activities for various school programs such as student organizations and curriculum. This will provide accurate information on which area needs to be improved.
2. Participate willingly and be open to recommendations of experts in line with evaluation and assessment but still maintaining a critical mindset.
3. Ask for school support when conducting evaluation studies such as deloading, grant, and other incentives.
4. Enroll in courses such as measurement and evaluation offered by reputable schools in order to gain the appropriate perspective in conducting evaluation in schools.

The following are recommended for the next researcher in conducting needs assessment:

1. Include a mechanism in the assessment whether schools are truly conducting the real nature of evaluation and not just assessment by exhibiting evaluation reports.
2. Use other methodologies to get in-depth information on the status of evaluation in schools such as interview and focus group discussions.
3. Perform a research synthesis on the evaluation content of one area of schools in order to gain a more detailed insight on the status of evaluation and the rigors on how it is conducted.

References

American Evaluation Association, "*Guiding Principles for Evaluators: A report from the AEA Task Force on Guiding Principles for Evaluators*," by D. Newman, M. A. Schreir, W. Shadish, & C. Wye. Available on-line: eval.org/EvaluationDocuments/

Arter, J. A., Stiggins, R. J., Duke, D., & Sagor, R. (1993). Promoting assessment literacy among principals. *NASSP Bulletin, 77*(556), 1-7.

Astin, A. W. (1993). *Assessment of excellence: The philosophy and practice of assessment and evaluation in higher education*. Arizona: The Oryx press.

Bernstein, E. (2004). What teacher evaluation should know and be able to do: A commentary. *NASSP Bulletin, 88*, 80-89.

Berry, B. et al. (November 2003). "The Impact of High-Stakes Accountability on Teachers' Professional Development: Evidence from the South." A Final Report to the Spencer Foundation. Southeast Center for Teaching Quality, Inc.

Berube, W. G, Gaston, J., & Stepans, J. L. (2002, December). The role of the principal in teacher professional development. Paper presented at the annual meeting of the Northern Rocky Mountain Educational Research Association, Estes Park, CO.

Darling-Hammond, L. (n. d.). Multiple measures approaches to high school graduation: A Review of state student assessment policies. [on-line available] vvww. Schoolredesign.net/srn/mm/mm/ php

Davies, A. (2000). *Making classroom assessment work*. Merville, BC: Connections Publishing.

Dounay, J., (n. d.) High-stakes testing, Systems. [on-line available] www.ecs.org/clearinghouse/14/56/1456.htm

Duran, A. (2005). Factors to Consider When Evaluating School Accountability Results. *Journal of Law and Education, 34*, 73-101.

Falk, B. (2000). *The heart of the matter: Using standards and assessment to learn*. Portsmouth, NH: Heinemann.

Ferrera, R. J. (2005). Accountability ALARMS. *Leadership, 35*, 24-26.

Greene, J. (1988).Stakeholder participation and utilization in program evaluation. *Evaluation Review, 2*, 91-116.

Harada, V. H. (2005). Working smarter: Being strategic about assessment and accountability. *Teacher Librarian, 33*, 8-16.

Harada, V. H., & Yoshina, J. M. (2005). *Assessing learning: Librarians and teachers as partners*. Westport, CT: Libraries Unlimited.

Hughes-Hassell, S., & Bishop, K. (2004). Using focus group interviews to improve library services for youth. *Teacher Librarian, 32*, 8-13.

Magno, C. (2010). A brief history of educational assessment in the Philippines. *Educational Measurement and Evaluation Review, 1*, 140-149.

McKillip, J. (1998). Need analysis: process and techniques. In L. Bickman and D. J. Rog (eds.), *Handbook of Applied Social Research Methods*. Thousand Oakes, CA: Sage.

Moore, A. D., Dexter, R. R., Berube, W. G., & Beck, C. H. (2005). Student assessment: What do superintendents need to know?*Planning and Changing, 36*, 68-70.

O'Donnell, J. (1996). *For the chosen few: A guide to classroom supervision*. Manila: Cacho Publishing House.

Ponticell, J. A. & Zepeda, S. J. (2004). Confronting Well-Learned Lessons in Supervision and Evaluation. National Association of Secondary School Principals. *NASSP Bulletin, 88*, 43-50.

Posavac, E. J. (2003). *Program evaluation: Methods and case studies*. New Jersey: Prentice Hall.

Philippine Human Development Report (2000). *Quality, access, and relevance in basic education*. Human Development Network and United Nations Development Programme.

Warna, G. D. (1995). Program evaluation for school improvement: Guidelines for school administrators. National Association of Secondary School Principals. *NASSP Bulletin, 79*, 76.

Appendix

Survey of Needs Assessment

1. Is evaluation done in your school? ___ Yes ___ No
   If NO, please proceed to item #5.

2. Which of the following activities do you evaluate?
   *(Indicate your response in **Column I**. You may check as many as applicable.)*

Column I	**Column II**		
_____	_____	a.	Teacher Performance
_____	_____	b.	Administrative Performance *(e.g., coordinator, principal, director, etc.)*
_____	_____	c.	Support Staff Performance
_____	_____	d.	Implementation of New Academic Programs
_____	_____	e.	Teacher Training Programs *(e.g., seminars, symposia, etc.)*
_____	_____	f.	Selecting Students for Academic and Special Awards
_____	_____	g.	Guidance and Counseling Programs
_____	_____	h.	Homeroom Guidance Program
_____	_____	i.	Administrative Services *(e.g., maintenance, engineering, accounting, etc.)*
_____	_____	j.	Student Organizations
_____	_____	k.	Student Publications
_____	_____	l.	Sports Development Program
_____	_____	m.	Cultural Activities
_____	_____	n.	Community Service
_____	_____	o.	Retreat, Recollection and other Formation Programs
_____	_____	p.	Service feedback
_____	_____	q.	Canteen/Cafeteria Evaluation
_____	_____	r.	Others, *please specify:* _____ _____

3. Of the activities you checked in question no. 2, which needs to be improved?
   *(Indicate your response in **Column II**. You may check as many as applicable.)*

4. Who handles the evaluation in your school?

   ___ Guidance Counselor ___ Teachers ___ Coordinators ___ Assistant Principal
   ___ Principal ___ Others, *please specify:* _____

5. How do you gather data for evaluation purposes? *(You may check as many as applicable.)*

   ___ Inventory/Questionnaires ___ Surveys ___ Tests
   ___ Focus Group Discussions ___ Observations ___ Experiment
   ___ Personal Interview ___ Others*, please specify:*
   _____

6. In which of the following areas in evaluation would you need technical assistance?

	High Need	Moderate Need	Low Need	Don't Know
a. Planning the evaluation				
b. Conceptualization				
c. Instrumentation *(constructing assessment forms, etc.)*				
d. Data Analysis				
e. Report Writing				
f. Utilization of results				
g. Dissemination of results				
h. Others, please specify: _____ _____ _____				

7. Where do you get information regarding evaluation? *(You may check as many as applicable.)*

_____ a. Books
_____ b. Journals
_____ c. Internet
_____ d. Discussion Groups
_____ e. School-Based Experts
_____ f. External Experts
_____ g. Seminars/Workshops
_____ h. Colloquia
_____ i. Inter-school Collaboration/Consortium
_____ j. Accreditation/Certification Documents/Manuals
_____ k. Others, *please specify:*
_____
_____

8. What are your suggestions to improve the practice of evaluation in schools?

Background Information About the Respondent

Name *(optional)*: _____

Current job position: _____

Number of years in this institution: _____

Year of establishment of school: _____

School population:
- Elementary _____
- High School _____
- College _____
- Graduate School _____

Use of the Rasch Model in the Abnormal Psychology Achievement Test

Ma. Joanna Tolentino-Anonuevo
La Salle College Antipolo Rizal, Philippines

The purpose of this study was to determine the unidimensionality and item characteristic of Abnormal Psychology Achievement Test (APAT) for psychology undergraduates using Rasch based analysis. Abnormal psychology was identified as one of the content area for the licensure examination for psychometrician, for which, psychology graduates shall be qualified to take. This study was an initial attempt to construct items for all the content areas indicated in the Philippine Psychology Act of 2009, which will eventually be administered to psychology undergraduates as Psychometrician Licensure Readiness Test. APAT was administered to 39 psychology students from two different schools. Result supported the unidimensionality of the test, which obtained a unidimensionality value of .98. Thirty-three out of the 35 items achieved an acceptable goodness of fit indices, with MNSQ INFIT and OUTFIT values <1.5. The item map revealed that psychology students' ability was higher than the items of APAT. This result suggests that more items on abnormal psychology need to be formulated since there were observed gaps on the item map.

Keywords: Abnormal Psychology, Achievement Test, Rasch Analysis

Psychology as a profession has the ethical obligation to do no less than its best to ensure that its members are competent and to offer evidence of competence through proper assessment procedures such as licensures(Roberts, Borden, Christiansen, & Lopez, 2005).With the enactment of the Philippine Psychology Act of 2009 to regulate and ensure qualified delivery of psychological services, psychology students need to be prepared and should possess the competencies needed to qualify as practitioners. Licensure examinations are considered as a summative assessment of knowledge and skills learned in schooling and training. However, in assessing competencies, both

formative and summative assessment are valuable in providing information as to the progress made by a student and in providing intervention programs to implement to ensure that necessary competencies are learned and demonstrated. To prepare psychology graduates for the licensure examination, student competencies in the various areas that will be covered may be assessed through an achievement test designed specifically for this purpose. Achievement test results are being utilized for various purposes and decisions related to learning. However, test scores may not provide an accurate information if interpretation is limited only to one's total score relative to how many items and the population for which it was used. This does not give accurate information about the person's ability relative to the items of a given test. Thus, statistical techniques that would provide information on both the items of the test and the person's ability would be valuable indeed if information obtained will be used as a basis to assess readiness of examinees in crucial competencies expected of the students, such as in licensure examinations.

RA 10029

The Philippine Psychology Act of 2009 was approved in the Senate and House of Representatives on December 2009 on and was enacted on March 16, 2010. It acknowledges the role of psychologist and psychometrician in nation building and development. It is seeks to ensure that the practice of psychology in the Philippines is regulated and that specific services inherent in the profession are delivered by qualified, trained and globally competitive individuals through administration of credible licensure examinations (Republic Act 10029, 2009).

Parts of the exam are four of the major subjects in psychology, which are taken between second to fourth year college. The subject descriptions based on CHED Memorandum Order are as follows: (a) Theories of Personality-a survey of the major theories or personality and the theoretical and practical issues involved in the scientific study and understanding of personality formation and dynamics; (b) Abnormal Psychology-an introduction to the nature, causes, and possible interventions of psychological disorders. The students are expected to be familiar with the nomenclature and classifications of mental disorders. Indigenous concepts of abnormality and abnormal behavior will also be discussed. Ethical considerations in abnormal psychology/clinical psychology are also discussed; (c) Industrial Psychology-a course providing an overview of psychological concepts, theories and research findings for effective human interactions and performance in the workplace. Topics include organizational structures and systems, organizational communication processes, leadership, motivation, conflict resolution, problem solving and decision making, team dynamics, efforts in human resource development and management, and organizational change and development; and, (d) Psychological Assessment-orientation into the rudiments of psychological testing. The principles, methods and uses of psychological testing are tackled. Emphasis is placed on issues of

item analysis, reliability, and validity in test construction. The administration, scoring, and interpretation of objective cognitive and affective tests used in various applied fields of psychology, particularly the educational, industrial, and government settings are covered. Ethical considerations as well as current trends and issues in psychological testing in the Philippine setting are discussed. Since the competency areas were identified for the upcoming licensure examination for psychology graduates, particularly, the licensure examination for psychometrician, an achievement test that would adequately represent the content domains of each area would allow for an accurate assessment of students' capabilities once the test's strength is established.

Assessment of Competence

There has been difficulty encountered in the definition and measurement of competencies in professional psychology (Roberts, Borden, Christiansen, & Lopez, 2005). Professional competency has been defined as "the habitual and judicious use of communication, knowledge, technical skills...in daily practice for the benefit of the individual and the community being served" (Epstein & Hundert, 2002, p.227). Roberts and colleagues (2005) proposed that competencies related to psychology, as a profession needs to be assessed utilizing methods that would ensure crucial psychometric properties of instruments and acceptable and maximum levels of skills and knowledge should also be established. In the assessment of competencies in professional psychology, formative and summative assessment techniques are needed (Lamb, 2010; Scriven, 1967). Formative assessment involves a continuous process of assessing by providing feedback for progress made toward a specified learning goal.

Licensure Exam in Psychology

The core assumption and dominant argument for the implementation of licensure is that it serves to protect the public (Hess, 1977). In the United States, licensure laws were enacted to ensure that psychological services might only be administered and conducted by qualified psychology practitioners (Hess, 1977; Danish & Smyer, 1981; Reaves, 1995; Bickman, 1999). Rehm & DeMers' (2006) article briefly discussed the history of psychology licensure in the United States. Licensure examinations for psychology may be traced back during the period after the World War II. It was during this time that national efforts were directed toward increasing the availability of mental health services in the United States. In 1965, the Examination for the Professional Practice of Psychology (EPPP) was formulated and administered by the American Association of State Psychology Boards (AASPPB), now called Association of State and Provincial Psychology Boards (ASPPB). There were 150 multiple-choice items from the original version of EPPP and every two years, a new test is formulated. Since 2003, EPPP is administered via computer, and it now consists of 225 items

(25 items not scored; used for test development), which have undergone calibration using item response theory statistics (p. 250).

The EPPS assesses if candidates have "a sufficient base of knowledge regarding psychology ... the exam is more analogous to an achievement test, measuring knowledge acquired..." (Erikson, Cornish & Smith, 2009, p.341). Similarly, Sharpless & Barber (2009) identified EPPP as a "high-stakes examination", meaning, it is a consequence of performance, and even likened to the Scholastic Aptitude Test. It is a summative assessment, which not only determines accumulated knowledge about psychology, but also how to apply knowledge in several content areas (ASPPB, 2009). Moreover, EPPS was not designed to predict professional performance (ASPPB, 2009) but was created to "establish, nationwide, a minimum standard for certification or licensure" (Reeves, 2006, p. 24). Sharpless and Barber (2009) and Erikson Cornish & Smith (2009) examined the strengths and weaknesses of the EPPS. Content validity is referred to as the degree to which a test represents to content domain it seeks to cover. This type of validity is considered as the primary consideration in licensure examinations. The content validity of EPPP was established through information gained about crucial competencies from thousands of psychology practitioner. The criterion validity of the EPPP was difficult to establish since there was no standardized measure against which it may be assessed (Erickson, Cornish, & Smith, 2009). Reasons for failure rates were also looked into by the authors. Sharper and Barber (2009) attributed low passing rates of applicants to problems in EPPP's validity and examiners' lack of preparation for the examination. Though Erickson, Cornish, and Smith (2009). Agreed on the aspect of unprepared examinees, they argued that failure in the licensure exams might be due to examiners inability to successfully integrate and apply all knowledge they learned and test anxiety. Regardless of some controversies encountered by the EPPS, it still remained as the "best standardized measures of the broad knowledge base needed to ensure minimal competence for entry-level practice and licensure" (Erickson, Cornish, & Smith, 2009).

Rasch Analysis

The study sought to construct and validate initial items for Abnormal Psychology using the Rasch model of analysis. Item Response Theory (IRT) is also known as Latent Trait Theory, Strong True Score Theory, or Model Mental Test Theory. IRT provides an estimate of a latent trait—referred to as a characteristic or ability of an individual that is not directly observed nor absolutely determined yet may be inferred from an aspect of an individual's performance or presentation (Baylor et. al., 2011). One-Parameter Logistic or the Rasch Model is considered as applicable for dichotomous test items, or those items with either right or wrong answer. In this model, each item has its own characteristic curve, which describes the probability of answering the items given the ability of the examinee (Kaplan & Sacuzzo, 1997). Due to the inherent characteristic of latent traits, measurements used in estimation need to be carefully examined as regards its validity and reliability in capturing the

construct of interest. As compared with Classical Test Theory, which considers the person's total score as the true score or assumed to represent the actual latent trait including the corresponding measurement error (Crocker & Algina, 1986), IRT models considers both model and item-based measurement. Model based measurement provides information about the relationship between the latent trait measured and the person's responses to the items in a test and item-based measurement because relationship between the instrument (and corresponding per item parameters or characteristics) and the latent trait are also determined (Baylor et. al., 2011). Moreover, one of the advantage of IRT is the use of logit scales which approximates an equal-interval scale, that allows for determining a person's ability or trait level independent of any normative or comparison group (Baylor et. al., 2011). One-item parameter logistics uses item difficulty to determine the relationship between the item, the latent trait, and the person response (Baylor et. al., 2011). One advantage of the IRT models is that it has parameters for both items and person ability while Classical Test Theory is sample dependent (Hambleton, 2000).

Unidimensionality Coefficient

One of the key assumptions in most IRT models is unidimensionality, which means that all of the items in an instrument represent a single underlying construct or latent trait (Baylor et. al., 2011). It is imperative that unidimensionality be quantitatively determined in the early stages of test development and to use result as basis for removal or modification of items which represent a different trait or construct (Baylor et. al., 2011). To determine unidimensionality coefficient in a Rasch model, the ratio of model standard error to real standard error for person separation reliability was determined. According to Wright (1999), model standard error considers model misfit as random variation while real standard error regards misfit as a true departure from the unidimensional model. The closer the coefficient value to 1.00, the closer the data approximates unidimensionality.

The information on both the person ability and test characteristic on the Abnormal Psychology Achievement Test through Rasch analysis would be valuable in assessing licensure readiness not only to ensure that chances of passing the licensure would eventually be determined but also in enhancing classroom instruction and student achievement.

It is the goal of every educational institution to develop the crucial and relevant competencies of students in their respective field of specialization. With the approval of RA 10029, necessary knowledge and competencies among psychology graduates need to be ensured. The present study is an initial attempt to come up with an instrument that would measure psychology students' readiness for the licensure examination on the abnormal psychology content area.

Purpose of the Study

1. To construct a psychometrically sound abnormal psychology achievement test for college students.
2. To determine psychology student's ability and item performance on the Abnormal Psychology Achievement Test
3. To determine the unidimensionality and item fit statistics of the Abnormal Psychology Achievement Test using Rasch analysis.

Method

Participants

There were 39 psychology students from the Philippines who completed the Abnormal Psychology Achievement Test. Only those students who finished taking up the subject were included in the study. There was a limited access to psychology students, thus convenience sampling was used.

Measure

The Abnormal Psychology Achievement Test is comprised of 35 items in multiple-choice format, which was reviewed by an expert in test construction. Items were constructed based on the major topics, specifically, on definition of abnormality and basic assessment of psychopathology, different psychological disorders, and interventions. Moreover, items were constructed to measure levels of comprehension and analysis of relevant concepts and principles relevant to abnormal psychology. Table 1 shows the table of specification for APAT.

Table 1
Table of Specification

Content Areas		Comprehension	Analysis	Items
Definition of abnormality	13%	1-5	0	5
Major Psychological Disorders	74%	6-26	27-30	26
Interventions	13%		31-35	5
Weight	100%	74%	26%	35

Procedure

A request was made to the Department Chairperson of the two schools regarding the administration of the 35-items Abnormal Psychology Achievement Test. There was a total of 39-psychology student who took the test. APAT items was checked and encoded "1" for correct answers and "0" for wrong answers. Data analysis was performed using Winsteps Statistical Package (Linacre, 1991).

Descriptive statistics was computed using the Statistical Package for the Social Science version 11.0.

Data Analysis

Means, standard deviations, maximum and minimum statistic, kurtosis, skewness and reliability estimates were obtained using SPSS version 11. Unidimensionality, item fit, test characteristic curves, item-scale correlations, item difficulty, person ability mean and item mean were also computed using Winsteps (Linacre, 1991). Unidimensionality index determines if APAT is measures a single dominant construct, as hypothesized.

Results

APAT was pilot-tested to 39 psychology students who finished the subject in abnormal psychology. The mean scores of psychology students on APAT is 19.82 (SD=6.77). Minimum score is 7 while maximum score on the test is 33. Cronbach's alpha is .86. The skewness value of .12 means that the distribution of scores approximates the normal curve; it is within the acceptable range of +1.0 to -1.0. Moreover, kurtosis value of -.54 signifies that it is platykurtic.

Table 2
Separation Table

	n	*M*	*SD*	Model Error	S.E.	Real RMSE				Model RMSE			
						RMSE	True *SD*	Separation	Reliability	RMSE	True *SD*	Separation	Reliability
Person	39	19.8	6.7	.43	.19	.46	1.09	2.37	.85	.45	1.10	2.45	.86
Item	34	22.7	6.2	.38	.15	.39	.75	1.92	.79	.38	.76	1.98	.80

The person ability estimate mean of +.50 indicates that the psychology students who took APAT finds the test as generally average in difficulty since a value closer to 0 signifies a well matched test (Bond & Fox, 2001). The reliability and separation index provide information about the hierarchy of items and persons on a particular test; the higher the values denote the replicability of item placement across other samples and order of persons on similar test (Bond & Fox, 2001). As shown on table 2, reliability of person ability estimate value of .85 and person separation index value of 2.37 as compared with item reliability value of .79 and item separation value of 1.92, signify that in the current analysis, better information about person ability is derived.

Item fit statistics were derived to determine if observed item characteristics is consistent with the Rasch model. For a multiple-choice test,

items with mean square indices greater than 1.2 or less than 0.8 are considered misfit items (Linacre & Wright, 1994).

All the items of APAT were included in the analysis since no item was correctly answered by all psychology students nor was there an item that was not correctly answered by all. This signifies that all the 35 items may provide information as to the ability of the psychology students who took the test. Table 2 shows the index of reliability of test scores and the ability of psychology students in the Abnormal Psychology Achievement Test (APAT), as follows: psychology student (person) reliability was .85 with an RMSE of .46, and APAT reliability was .79 with an RMSE of .39. Standard error was also satisfactory since it is near the value of "0" (.15 for items and .19 for person). The more the standard error is near "0", the better.

It can be noted on the item polarity result (See Appendix) that item-scale correlation (pt.- correlation column) has positive values ranging from .20 to .68. This means that items are performing well together because all values are positive. As shown in the item map (figure 1) and item polarity (table 2), most of the items are within easy to average difficulty level with few difficult items. This is based on logit values ranging from -.01 to 1.48. The positive and higher the logit value, the more difficult the item and the negative and lower the logit value, the easier the item. Item map (figure 1) also shows satisfactory clustering of all the items, which are within the -1.96 to + 1.96 logit values.

Information gleaned using the Rasch analysis on the Abnormal Psychology Achievement Test showed that majority of the items formulated was good. However, since most of the items range from easy to average, this may explain why the mean for person ability was much higher and not equal to the item mean. This signifies that the students who took the test have high ability levels but the items generated did not include many items, which are high in difficulty level to match ability. Item map provides an overview on how the items and the persons are performing on a particular measure. Ideally, the mean of both items and ability should be equal or near in value and that the items should be spread out in such a way that ability levels are also matched to the items, with no gaps as much as possible. With this finding, new items may be generated to enhance the Abnormal Psychology Achievement Test characteristics

Examination of Fit

The average INFIT MNSQ value was 1.00 (SD=.15), and the mean OUTFIT MNSQ value was .99 (SD=.29), which indicated that data for the items showed goodness of fit with values <1.5 except for items 3 and 28 (OUTFIT MNSQ value of 1.89 and 1.56, respectively).

For APAT, the ratio of model standard error is .45 and the real standard error is .46, yielding a coefficient value of .98. This implies that the test is unidimensional. The person reliability is .85 and the item reliability is .79. The sample produced a person separation of 2.37. To determine the number of distinct ability strata, the strata formula was used (HP=[4GP+1]/3), which yielded a value of 3.49. This means that ability may be separated into three

distinct groups. Using the same formula for the item separation value of 1.92, the result is 2.89, which denotes that the test items of APAT may also be categorized into three subgroups. This result is consistent with the sub categories of APAT, which are: (a) definition of abnormality and basic assessment of psychopathology, (b) different psychological disorders, and (c) interventions.

Rasch analysis has two assumptions (a) the higher the ability of a person, the higher the probability that difficult items will be answered correctly, and (b) the easier the items, the greater the chance that it will be answered correctly. Rasch analysis reflects the matching of person's ability with the difficulty of the item. As seen in Figure 1, the mean of the items is in the 0.0 logit value while the person mean is .5 higher than the item mean. This means that the ability of the psychology students was higher than the difficult items of APAT. As observed in the figure, items within the mean are item numbers: 12 (*"Which alter is the most common manifestation for Dissociative Identity Disorder?"*), 30 (*"Arnel knows that every time he sneezes, an earthquake happens in other places. What type of delusion is being referred?*), and 31 (*The therapist helps Therese to recall her childhood experiences and uncover unconscious motives by encouraging her to talk spontaneously about past recollections, dreams, and other related experiences. Which psychological therapy is described?*). As shown in figure 1, items and person's ability were plotted against each other. The higher the items are from the item mean (logit=0.0) the more difficult the item becomes and the higher the ability that is required for a person to answer the item correctly. When there is a match between item and person ability, there is a 50% chance that the person may answer the item correctly or incorrectly.

In terms of the item map (See Appendix), the person mean is .5 logit which is .5 higher than the item mean. This signifies that psychology students' ability exceeds the difficulty of the items. There were potential item gaps observed, which were not significant since it was < 1.00 logit value. As the person means value (.5 logit) exceeds item mean value (0.0 logit), it may be an indication that psychology students are more likely to give correct answers than incorrect ones.

There are 14 items that fall below item mean (logit value of 0.0) as follows: 3, 5, 8, 9, 15, 17, 18, 20, 22, 23, 24, 26, 33, and 34. There are 17 items above the item mean, these are: 1, 2, 4, 6, 10, 11 13, 14, 16, 19, 21, 25, 27, 28, 29, 32, and 35. As regards person ability, the figure shows that psychology students' ability exceeded the difficult items of APAT.

Discussions

The present study generally provides support for the construct and content validity of the Abnormal Psychology Achievement Test. All the items were included in the analysis since no item was answered correctly nor was not answered by all psychology students who took the test. Acceptable fit of 33 out of 35 of the APAT items was evident on the INFIT and OUTFIT MNSQ values <1.5.

Moreover, Rasch unidimensionality coefficient of .98 supported the hypothesis that APAT measures a unidimensional construct. The clustering of items were satisfactory since all were within the logit value of -1.96 to + 1.96, which means that these were good items.

The present study provides valuable information on how to construct test items that would fully capture the target construct under study. The use of Rasch model in analysis was deemed appropriate and provided valuable information. Results revealed that more items are needed to be able to represent the range of abilities of examinees. Moreover, majority of the items to be added may to be formulated to tap into higher-level cognitive skills such as evaluation and synthesis.

Aforementioned findings of the study revealed the advantages of using the Rasch model in analyzing test questions, particularly, Abnormal Psychology Achievement Test (APAT). The purpose for which the test was constructed is to be able to formulate a psychometrically sound assessment measure that will provide useful information regarding the readiness of psychology students in a specific content area, which is abnormal psychology. Though Classical Test Theory has its own strengths, information derived would just be limited to item characteristics such as discrimination and difficulty and said characteristics would be sample dependent. For an assessment measure that would be comparable to high stakes testing such as licensure examinations, more information about the test needs to be determined and established. The use of Rasch analysis in determining the psychometric characteristic of APAT has its advantages. Firstly, the characteristics of the items are independent of the group to which it was administered, thus, similar trend may be observed if APAT will be administered to a different sample. Second, the scores which describe the abilities of psychology students who took the test is not dependent on APAT, which signifies that students' capabilities were best identified and it may also be expected that student's level of performance may also be observed in other assessment measures. And lastly, there is a correct estimation in determining how a person with a particular ability will answer a test of varied difficulty level. Use of Rasch analysis results in estimating ability and test characteristic for the Abnormal Psychology Achievement Test would allow for an informed decision in various areas such as psychology students' school achievement and curriculum content. Moreover, information gleaned on this initial attempt in constructing a test for psychometrician licensure readiness, particularly in abnormal psychology, would also direct efforts on how to enhance the existing APAT items and in formulating items for the other 3 content areas to best approximate the domains to be measured.

References

Association of State and Provincial Psychology Boards (2009). *Guidelines on Practicum Experience for Licensure*. Peachtree City: GA.

Baylor, C., Hula, W., Donovan, N., Doyle, P., Kendall, D., & Yorkston, K. (2011). An introduction to item response theory and rasch models for speech-language pathologists. *American Journal of Speech-Language Pathology, 20*, 243-259.

Bickman, L. (1999). Practice makes perfect and other myths about mental health services. *American Psychologist, 54*, 965-978.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences (2nd ed.)*. IOM.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Belmont, CA: Wadsworth/Thomson Learning.

Danish, S., & Smyer, M. (1981). Unintended consequences of requiring a license to help. *American Psychologist, 32,* 365-368.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association, 287*(2), 226-235.

Erickson Cornish, J. A., & Smith, R. D. (2009) Reflections on the EPPP: A commentary on Sharpless and Barber. *Professional Psychology:* Research and Practice, 40 (4), 341–344.

Hambleton, R. K. (2000). Emergence of item response modeling in instrument development and data analysis. *Medical Care, 38*, 60-65.

Hess, H. (1977). Entry requirements for professional practice of psychology. *American Psychologist, 32,* 365-368.

Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological Testing: Principles, Applications, & Issues*. Belmont, CA: Thomson and Wadsworth.

Lamb, J. H. (2010). Reading grade levels and mathematics assessment: An analysis of Texas mathematics assessment items. *The Mathematics Educator, 20*, 22-34.

Linacre, J.M. (1991, April). *Structured rating scales*. Presented at the Sixth International Objective Measurement Workshop, Chicago.

Magno, C. P., & Ouano, J. A. (2010). Designing written assessment for student learning. Quezon City. Phoenix Publishing House.

Reaves, R. (1995, April). *The history of licensure and certification of psychologists in the United States and Canada*. Paper presented at the First International Congress on Licensure, Certification, and Credentializing of Psychologists, New Orleans, LA.

Rehm, L. P., & DeMers, S. T. (2006). Licensure. *Clinical Psychology and Science Practice, 13*(3), 249-253.

Roberts, M. C., Borden, K. A., Christiansen, M. D., & Lopez, S. J. (2005). Fostering a culture shift: Assessment of competence in the education and careers of professional psychologists. *Professional Psychology: Research and Practice,* 36, 355–361.

Scriven, M. (1967). The methodology of evaluation. In R.W. Tyler, R.M. Gagne, & M. Scriven (Eds.)., *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand-McNally.

Sharpless, B. A., & Barber, J. P. (2009). The examination for professional practice in psychology (EPPP) in the era of evidence-based practice. *Professional Psychology: Research and Practice*, 40, 333–340.

The Philippine Psychology Act of 2009, Republic Act 10029 (2009).

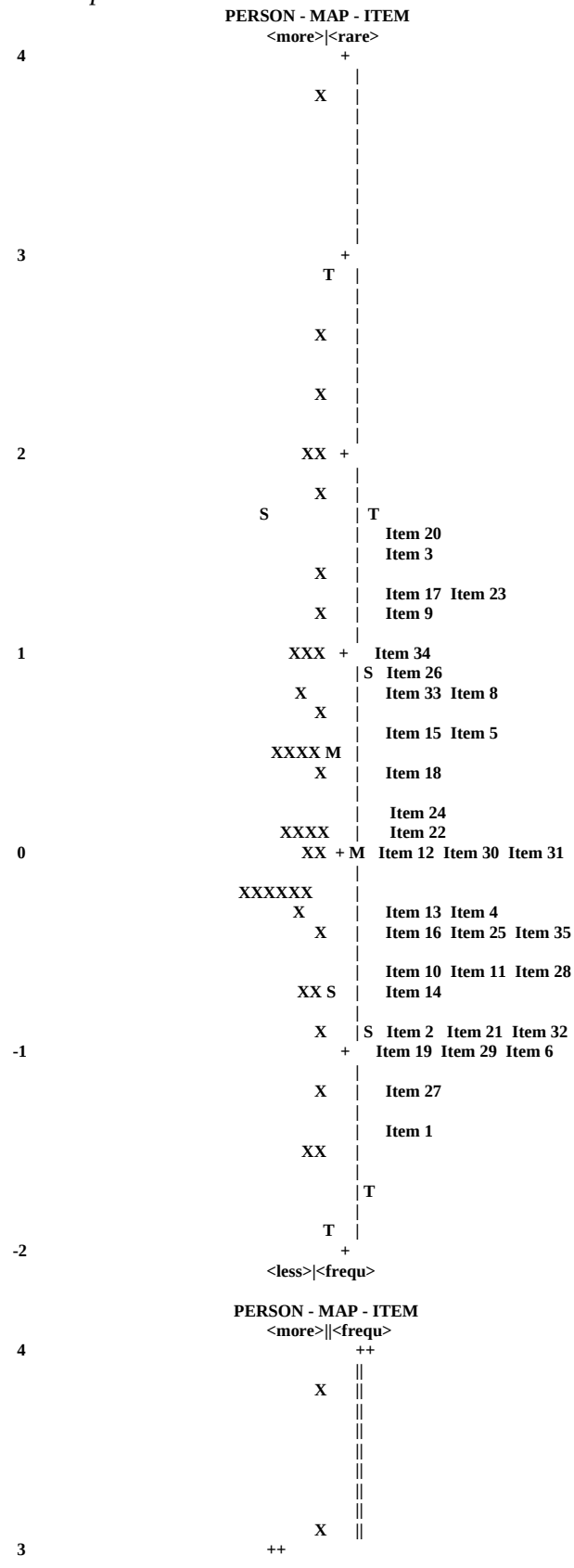
Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know* (pp. 65-104. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

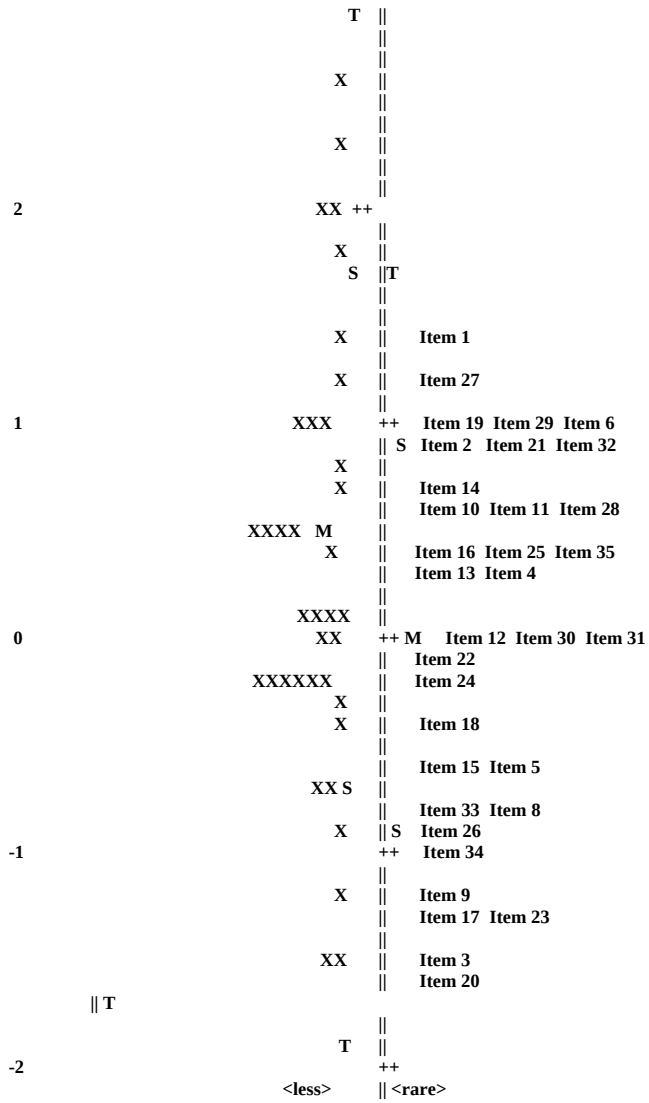
Appendix

Item Polarity

ENTRY NUMBER	MATCH| TOTAL SCORE	COUNT	OBS%	EXP%|	MEASURE	MODEL S.E.	ITEM	INFIT MNSQ	ZSTD|	OUTFIT MNSQ	ZSTD|	|PT-MEASURE		|EXACT CORR.	EXP.	
3	12	39		1.48	.40|	1.27	1.3|	1.89	2.5|	.20	.47|	76.9	77.0		Item 3
15	26	39		-.41	.37|	1.26	1.7|	1.18	.6|	.21	.39|	59.0	71.6		Item 16
27	27	39		-.56	.38|	1.17	1.1|	1.56	1.4|	.21	.38|	69.2	73.1		Item 28
17	20	39		.37	.36|	1.26	1.9|	1.37	1.5|	.23	.44|	56.4	68.8		Item 18
2	29	39		-.86	.40|	1.06	.4|	1.50	1.1|	.26	.36|	74.4	76.5		Item 2
9	27	39		-.56	.38|	1.14	.9|	1.32	.9|	.26	.38|	69.2	73.1		Item 10
24	26	39		-.41	.37|	1.13	.9|	.99	.1|	.32	.39|	64.1	71.6		Item 25
11	23	39		-.01	.36|	1.14	1.1	1.04	.2|	.34	.42|	53.8	68.3		Item 12
29	23	39		-.01	.36|	1.10	.8|	1.21	.8|	.34	.42|	64.1	68.3		Item 30
10	27	39		-.56	.38|	1.04	.3|	1.11	.4|	.34	.38|	69.2	73.1		Item 11
1	32	39		-1.38	.44|	.98	.0|	.76	-.2	.35	.31|	82.1	82.4		Item 1
26	31	39		-1.19	.43|	.95	-.1|	.86	-.1|	.37	.33|	84.6	80.4		Item 27
22	13	39		1.32	.39|	1.11	.6|	1.33	1.2|	.37	.47|	71.8	75.5		Item 23
8	14	39		1.18	.38|	1.16	.9|	1.03	.2|	.38	.47|	69.2	74.0		Item 9
23	21	39		.25	.36|	1.05	.5|	1.12	.6|	.38	.44|	69.2	68.5		Item 24
7	17	39		.77	.36|	1.09	.6|	1.16	.7|	.39	.46|	64.1	70.6		Item 8
28	30	39		-1.02	.41|	.93	-.3|	.79	-.3|	.40	.35|	79.5	78.4		Item 29
4	25	39		-.28	.37|	1.01	.2|	.90	-.2|	.41	.40|	69.2	70.1		Item 4
31	29	39		-.86	.40|	.94	-.3|	.82	-.3|	.41	.36|	79.5	76.5		Item 32
20	29	39		-.86	.40|	.92	-.4|	.74	-.5|	.44	.36|	74.4	76.5		Item 21
30	23	39		-.01	.36|	.99	.0|	.88	-.3|	.44	.42|	64.1	68.3		Item 31
34	26	39		-.41	.37|	.94	-.4|	.81	-.5|	.45	.39|	69.2	71.6		Item 35
5	18	39		.63	.36|	.99	.0|	.99	.0|	.46	.46|	69.2	69.7		Item 5
18	30	39		-1.02	.41|	.80	-1.0|	.69	-.5|	.50	.35|	84.6	78.4		Item 19
12	25	39		-.28	.37|	.89	-.8|	.76	-.7|	.50	.40|	74.4	70.		Item 13
19	11	39		1.64	.41|	.94	-.2|	.86	-.3|	.52	.47|	79.5	78.4		Item 20
6	30	39		-1.02	.41|	.77	-1.1|	.60	-.8|	.53	.35|	84.6	78.4		Item 6
14	18	39		.63	.36|	.91	-.6|	.81	-.8|	.54	.46|	69.2	69.7		Item 15
25	16	39		.90	.37|	.91	-.5|	.82	-.7|	.54	.46|	74.4	71.6		Item 26
16	13	39		1.32	.39|	.88	-.6|	.84	-.5|	.56	.47|	76.9	75.5		Item 17
32	17	39		.77	.36|	.85	-1.0|	.89	-.4|	.56	.46|	74.4	70.6		Item 33
13	28	39		-.70	.39|	.75	-1.5|	.60	-1.0|	.57	.37|	82.1	74.7		Item 14
33	15	39		1.04	.37|	.83	-1.0|	.81	-.7|	.59	.47|	79.5	72.		Item 34
21	22	39		.12	.36|	.68	-2.8|	.59	-1.8|	.68	.43|	87.2	68.1		Item 22
MEAN	22.7	39.0	.00	.38	1.00	.0		.99	.0			72.6	73.3			
S.D.	6.2	.0	.85	.02	.15	1.0	.29 .8			8.1	3.9					

Item Map


PERSON - MAP - ITEM
<more>|<rare>
4
3
2
1
0
-1
-2
<less>|<frequ>
PERSON - MAP - ITEM
<more>||<frequ>
4
3
Item 20
Item 3
Item 17 Item 23
Item 9
Item 34
Item 26
Item 33 Item 8
Item 15 Item 5
Item 18
Item 24
Item 22
Item 12 Item 30 Item 31
Item 13 Item 4
Item 16 Item 25 Item 35
Item 10 Item 11 Item 28
Item 14
Item 2 Item 21 Item 32
Item 19 Item 29 Item 6
Item 27
Item 1


T
X
X
2
XX
X
S
X
X
1
XXX
X
X
XXXX M
X
XXXX
0
XX
XXXXXX
X
X
XX S
X
-1
X
XX
|| T
T
-2
<less>
||
++
Item 1
Item 27
Item 19 Item 29 Item 6
S Item 2 Item 21 Item 32
Item 14
Item 10 Item 11 Item 28
Item 16 Item 25 Item 35
Item 13 Item 4
M Item 12 Item 30 Item 31
Item 22
Item 24
Item 18
Item 15 Item 5
Item 33 Item 8
S Item 26
Item 34
Item 9
Item 17 Item 23
Item 3
Item 20
<rare>