



ISSN 20194-5876

# **Educational Measurement and Evaluation Review**

VOLUME 8 ISSUE 1  
JULY 2017

THE PHILIPPINE EDUCATIONAL  
MEASUREMENT AND EVALUATION  
ASSOCIATION (PEMEA)

---

The Educational Measurement and Evaluation Review (EMEReview) is the official publication of the Philippine Educational Measurement and Evaluation Association (PEMEA). It is international, refereed, and abstracted/indexed. The EMEReview publishes scholarly reports about contemporary theories and practices in the field of education and social science that highlights measurement, assessment, and evaluation. It welcomes articles that are about test and scale development, quantitative models of a construct, evaluation studies, best practices in evaluation, issues and policies on assessment, contemporary approaches in educational and psychological measurement, and other studies with direct implication to assessment in education, social science, and related fields. EMEReview is indexed/abstracted in the Open J-Gate, JournalTOCs, Google Scholar, InfoBase Index, Social Science Research Network, Open Academic Journals Index, Scientific Indexing Services, and ejournals.ph

---

Copyright © 2017 by the Philippine Educational Measurement and Evaluation Association.

This journal is open-access and users may read, download, copy, distribute, print, search, or link to the full texts, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

The articles in the EMEReview are open access at  
<http://www.pemea.org/emereview>



Publication Division of PEMEA  
Philippine Educational Measurement and Evaluation Association

## ◆ Editorial Board

### Editor:

Dr. Adonis P. David, Philippine Normal University  
editorpemeajournals@gmail.com

### Managing Editor:

Dr. Carlo Magno, Mapua University  
crlmgn@yahoo.com

### Associate Editors:

Dr. Marcos Lopez, Centro Escolar University-Malolos, Philippines  
Dr. Richard Gonzales, Development Strategists International Consulting  
Dr. Marilyn Balagtas, Philippine Normal University, Philippines  
Dr. Teresita T. Rungduin, Philippine Normal University, Philippines  
Mr. Jesus Alfonso Datu, The University of Hong Kong, Hong Kong  
Ms. Belen Chu, Philippine Academy of Sakya Philippines. Arellano University  
Ms. Marife Mamauag, HELP University, Malaysia

### Editorial Advisory Board

Dr. John Hattie, University of Melbourne, Australia  
Dr. Jack Holbrook, University of Tartu, Estonia  
Dr. Anders Jonsson, Malmo University, Sweden  
Dr. Timothy Teo, University of Macau, China  
Dr. Tom Oakland, University of Florida, USA  
Dr. Jimmy dela Torre, Rutgers University, USA  
Dr. Jose Pedrajita, University of the Philippines-Diliman, Philippines  
Dr. Shu-ren Chang, Department of Testing Services, American Dental Association, USA  
Dr. Karma El Hassan, Office of Institutional Research and Testing, Americal University of Beirut, Lebanon  
Dr. Alexa Abrenica, Professional Regulation Commission  
Dr. Marie Ann Vargas, University of Sto. Tomas, Philippines

### Reviewers for this issue:

Dr. Adonis David, Philippine Normal University  
Dr. Carlo Magno, Mapua University  
Dr. Marilyn Balagtas, Philippine Normal University  
Dr. Jennie Jocson, Philippine Normal University  
Dr. Niclie Titatira, University of Rizal Systems  
Dr. Ma. Jenina Nalipay, Philippine Normal University

Implementing Outcome-Based Education (OBE) Framework: Implications for Assessment of Students' Performance

*Jonathan V. Macayan* ..... 1

Assessing Achievement Goals as Mediator on the Effect of Perfectionism on Procrastination

*Carlo Magno, Queen Marie Katrina O. Alcantara*  
*Christian Dave N. Domingo, Maria Queenie Y. Isidro* ..... 11

A Cross-Sectional Analysis of Classroom Assessment Literacy of English Teachers in Secondary and Tertiary Levels

*Tristan L. Piosang* ..... 30

Validation of the CEU-Lopez Critical Thinking Test Using Multidimensional IRT Model

*Johnny T. Amora and Marcos Y. Lopez* ..... 49



---

# Implementing Outcome-Based Education (OBE) Framework: Implications for Assessment of Students' Performance

Jonathan V. Macayan  
*Mapúa University*

## Abstract

This paper initially traces the roots of Outcome-Based Education (OBE) and introduces key concepts at the level of school-wide implementation based on Spady (1994). It then proceeds with defining what outcomes are and discusses how the definition of outcomes demands paradigm shift in assessment and evaluation practices. Finally, the paper tackles important implications of carrying out the framework for the practice and methods of assessment and evaluation of students' performance in schools. These implications are meshed with discussion of the four operating principles of OBE.

*Keywords:* outcome-based assessment, outcome-based evaluation, outcome-based education

## Introduction

In response to the need for standardization of education systems and processes, many higher education institutions in the Philippines shifted attention and efforts toward implementing Outcome-Based Education (henceforth OBE) system on school level. The shift to OBE has been propelled predominantly because it is used as a framework by international and local academic accreditation bodies in school- and program-level surveillance, on which many schools invest their efforts into. The Commission on Higher Education (CHED) even emphasized the need for the implementation of OBE by issuing a memorandum order (CMO No. 46, s. 2012) entitled, "Policy-Standard to enhance quality assurance in Philippine Higher Education through an Outcomes-Based and Typology Based QA". Then, in 2014, it was followed

by a release of the Handbook of Typology, Outcomes-Based Education, and Sustainability Assessment.

Given the current status of OBE in the country, this paper intends to shed light on some critical aspects of the framework with the hope of elucidating important concepts that will ensure proper implementation of OBE. Also, the paper zeroes in on inferring implications of OBE implementation for assessment and evaluation of students' performance.

## What is OBE?

Outcomes-based education as defined by Spady (1994, p. 12) means “clearly focusing and organizing everything in an educational system around what is essential for all students to be able to do successfully at the end of their learning experience.”

The definition explicitly specifies certain markers, which should serve as bases of actions and procedures that schools must undertake to ensure the proper institution-wide implementation of OBE. Tucker (2004) further emphasized this in his description of OBE as a process that should involve the restructuring of curriculum, assessment, and reporting practices in education. The changes that OBE entails put emphasis on students' demonstration of learning outcomes rather than accumulation of course credits. Also, these definitions of OBE emphasize the need to accordingly align all aspects of educational processes and systems to the expected outcomes that all students should be able to proficiently exhibit at the end of the curriculum, and that outcomes should not be viewed synonymously with grades or simply curricular completion, but rather authentic demonstrations of expected competencies as a result of significant learning experiences.

It can be surmised then that the implementation of OBE requires consistency across desired outcomes of education, teaching and learning activities, and assessment methods and practices.

To organize everything in the educational system (curriculum, resources, facilities, curricular and co-curricular activities, etc.) and align them with the desired outcomes of education, it would be necessary first for schools to have a clear understanding of what outcomes are. Thus, the next section addresses the following questions: What are outcomes and how are they derived and stated? The next section of this article will provide thorough discussions on the *outcomes* according to the OBE framework.

## What are Outcomes?

The term *outcome* is lexically defined as “*something that follows as a result or a consequence*”, “*an end-product or a result*”, and “*the way a thing turns out*”. One common denominator among these definitions is that they all concur that outcomes happen as a product or an end-result of processes or any antecedent factors or events. In education, outcomes are viewed as the learning results that students are expected to demonstrate across the curriculum. Hence, outcomes in education may vary in terms of levels or forms. According to Killen (2000), some outcomes are expected to be demonstrated at a course level (subject-related academic outcomes), and some are at the program and institutional levels (cross-discipline outcomes). However, according to Spady (1994, p. 49), the most important form of outcomes with which other forms or levels of outcomes should be aligned are those that reflect real life roles that learners will perform the moment they exit the education system – these are called ‘culminating outcomes.’

Simply, the course/subject-related and program level outcomes should be fundamentally linked to the culminating or exit outcomes of education. This practice ensures that education prepares students to perform future life-roles. Thus, the focus of OBE is more on the results or products of education, rather than on the content and curricular processes (Morcke et al., 2012).

In an OBE set-up, the first thing that should be identified and explicitly stated is the culminating or exit outcomes, *what we want our students to be able to do successfully at the end of their learning journey in school*. Again, these culminating outcomes should be based on life-roles that students will perform in the real world. One of the operating principles of OBE in Spady’s (1991; 1994) framework is the *Design Down* principle, which should be simultaneously applied together with the other operating principles (i. e., clarity of focus, high expectation, and expanded opportunity). The design down principle is like a top-down approach of formulating and stating outcomes. The culminating outcomes should be stated first, followed by some enabling outcomes (program level), then by some discrete outcomes that are measured in terms of specific learning tasks (course level). The backward design of outcomes would somehow guarantee that all the forms and levels of outcomes across the curriculum are systematically and intentionally aligned and connected. Then, the implementation of this design should be forward.

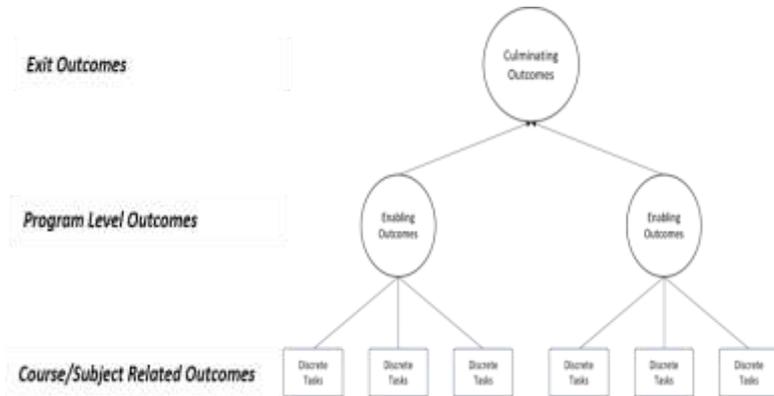


Figure 1. Levels of Outcomes of Education

## Implications to Learning Assessment Practices

As discussed in the preceding section, the implementation of OBE in the institution level would entail restructuring of relevant systems and procedures to constructively facilitate the attainment of the desired outcomes of education. This includes the critical restructuring of assessment methods and procedures employed by educators and institutions in evaluating student performance, which serves as evidence of the attainment of outcomes.

Assessment plays an important role in the educative process. It serves as basis for determining the rate of learning progress of students as well as the source of information of opportunities for further improvement. One of the most comprehensive definitions of assessment is provided by the American Association for Higher Education (Angelo, 1995, p. 7):

An ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance.

In OBE schools, assessment, when implemented appropriately, would have manifold purposes and benefits. Aside from providing educators ideas about the progress of students, it also informs them about the effectiveness of their teaching methodologies and approaches. Moreover, assessment results in an OBE school are used as bases to improve educational services and systems on an institutional level (Bresciani et al., 2012).

Proper implementation of OBE both in the classroom and institutional levels would demand paradigm shift. The following summarizes the shifts of assessment practices moving from the traditional practices to OBE practices:

### **Paradigm Shift 1: Teacher-Centered to Learner-Centered Approach**

Assessment in outcome-based education require a shift in mindset of educators and educational leaders. The shift requires a turnaround of approach from teacher-centered to learner-centered education (Bresciani, 2012; Bresciani et al., 2009; Ramoroka, 2006; Nieburh, 1996).

Table 1

*Assessment: Traditional vs. OBE*

Traditional	OBE
<ul style="list-style-type: none"> <li>• What are our (educators) practices</li> <li>• Teaching (inputs, content)</li> <li>• Teaching and Learning (TLA) as the end</li> <li>• Practice determines the outcomes</li> </ul>	<ul style="list-style-type: none"> <li>• What our students have become and able to demonstrate</li> <li>• Learning (demonstration of skills and competencies, outcomes)</li> <li>• Teaching and Learning (TLA) as the means to an end</li> <li>• Outcomes inform the practice</li> </ul>

### **Paradigm Shift 2: Being Outcomes-Minded**

Needless to say, in outcome-based education framework everything should be based on outcomes. Thus, assessment methods and techniques should be consistent with the stated outcomes of education. According to Bresciani (2006), outcome-based assessment is a systematic and intentional process. This means that the assessments used in this set-up are deliberately designed and administered in pursuit of outcomes attainment. Along with teaching and learning activities, assessments used in OBE classrooms should be constructively aligned with the outcomes that are expected to be

successfully demonstrated at specific stages and curricular levels (Biggs, 2011; Biggs & Tang, 2007).

Spady (1994) specified four operating principles that will guide educators and academic leaders in the implementation of OBE. When applied consistently, systematically, creatively, and simultaneously the efforts of shifting to OBE can be almost guaranteed.

The four operating principles of OBE and their implications for assessment are as follows:

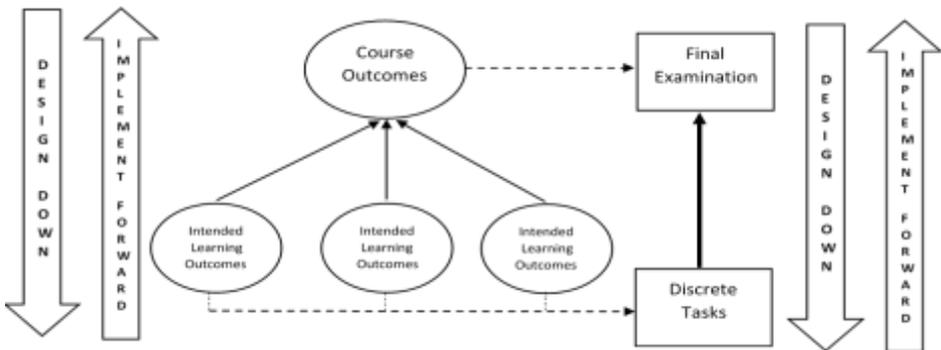
**Clarity of focus.** Educators should be made aware and conscious about the outcomes of education each student must manifest or demonstrate at the course level and that these outcomes at the classroom level are connected to the attainment of higher level outcomes (i. e., program/institutional outcomes and culminating outcomes). Thus, at the initial stage of academic or course planning, the higher outcomes serve as guide for educators in defining and clearly stating the focus of the course/subject. This principle implies that the criteria of attainment of learning outcomes (students' learning performance) that can be elicited through assessments should exhibit a particular standard that applies to all learners. In effect, this standardizes the assessment practices and procedures used by educators in specific subject/course.

**High expectations.** As stated in the clarity of focus principle, learning outcomes at the course level are necessarily connected to higher level outcomes. These connections warrant educators from eliciting high level of performance from students. This level of performance ensures that students successfully meet desired learning outcomes set for a course, and consequently enable them to demonstrate outcomes at higher levels (program or institutional level). Thus, the kind of assessments in OBE learning context should challenge students enough to activate and enable higher order thinking skills (e. g., critical thinking, decision making, problem solving, etc.), and should be more authentic (e. g., performance tests, demonstration exercise, simulation or role play, portfolio, etc.).

**Expanded opportunity.** The first and second principles importantly necessitate that educators deliver students' learning experiences at an advanced level. In the process, many students may find it difficult complying with the standards set for a course. As a philosophical underpinning of OBE, Spady (1994, p. 9) emphasized that "all students can learn and succeed, but not on the

same day, in the same way.” This discourages educators from generalizing manifestations of learned behavior from students, considering that every student is a unique learner. Thus, an expanded opportunity should be granted to students in the process of learning and more importantly in assessing their performance. The expansion of opportunity can be considered multidimensional (i. e., *time, methods and modalities, operational principles, performance standards, curriculum access and structuring*). In the assessment practice and procedures, the *time* dimension implies that educators should give more opportunities for students to demonstrate learning outcomes at the desired level. Thus, provisions of remedial, make-up, removal, practice tests, and other expanded learning opportunities are common in OBE classrooms. Methods and modalities of assessment can also be expanded depending on the types of learners. Students vary in many ways. One important aspect of diversity among learners for example is their thinking style. In studies on thinking styles (e. g., Abdi, 2012; Zhang, 2002), findings revealed that students vary on thinking or cognitive styles. These manifold styles when accommodated appropriately not only on the delivery of lessons but also on the type of assessments would yield more productive and successful results from students in terms of demonstrating the learned outcomes.

**Design down.** This is the most crucial operating principle of OBE. As mentioned in the previous section, OBE implements a top-down approach in designing and stating the outcomes of education (i. e., culminating --- enabling --- discrete outcomes). The same principle can be applied in designing and implementing outcomes’ assessments in classes.



*Figure 2.* Backward Design-Forward Implementation: Course Level Outcomes and Assessments

Traditionally, the design of assessments for classes is done following a bottom-up approach. Educators would initially develop measures for micro learning tasks (e. g., quizzes, exercises, assignments, etc.), then proceed to develop the end-of-term tasks (e. g., major examination, final project, etc.). In OBE context, since the more important outcomes that should be primarily identified and defined are the culminating ones, it follows that the same principle should logically apply. Thus, the first assessment that should be developed and designed for a course is the final assessment; from this, smaller measures (discrete tasks) can be logically designed and progressively implemented. This process employs the top-down approach, which guarantees that all course assessments are constructively linked and aligned to the desired outcomes of the course/subject, and ultimately to the culminating outcomes of education (i. e., program/ institutional, and exit).

## Discussion

Having discussed the fundamental concepts and principles of OBE, as well as the implications of this framework for schools systems and processes, the following can be deduced:

(1) The implementation of OBE in schools requires an academic organization to realign and adjust educational processes and systems in accordance with the desired outcomes of education. In effect, this necessitates major stakeholders of education (e. g., academic leaders, educators, academic external partners, etc.) to work together in determining, defining, and stating outcomes at various curricular levels (i. e., culminating, program, course outcomes).

(2) Proper implementation of OBE requires schools to undergo paradigm shift and consequently adopt some redefinition of the kind of education and educational services that they deliver to students. One of the shifts or changes that schools must adopt is learner-centeredness, not only in principle but importantly in practice. This change in approach cascades necessarily to assessment and evaluation practices. Another critical shift is not only on the awareness but on a serious adherence with the operating principles of OBE. To ensure proper implementation of the framework, these principles should be applied consistently, systematically, creatively, and simultaneously.

(3) Assessment plays a very important role in an OBE set-up. When implemented appropriately, assessment results serve as reliable bases in

determining whether students are on the right track in attaining the outcomes (formative) or have actually attained the desired outcomes at the course or program level (summative).

(4) Outcome-based assessment provides feedback and informs educators as regards the effectiveness of the teaching and learning practices that they employ in classes. This constructively and significantly redounds to the development of more responsive and adaptive teaching techniques that support students in attaining the desired outcomes of education.

Finally, it is clear that the optimal benefits of OBE can be realized if schools will seriously anchor the implementation of the framework on the philosophical underpinnings of outcome-based education. That is, everything in the educational processes and systems should be based on the outcomes; outcomes which extend beyond academics and reflect real-life attributes that the various stakeholders deem pivotal among students who graduate from schools and then integrate to the society as professionals. Teachers and academics must espouse the true-to-form purpose of OBE, which transcends accreditation and goes beyond preparing students for high-stakes assessments. This, in turn, challenges educators and assessment experts to develop and implement authentic assessments that measure real outcomes of education, be they quantitative or qualitative measures. Eventually, outcomes-based assessment should encourage the reshaping of the various levels of outcomes and the rethinking of teaching and learning and assessment tasks to ultimately prepare students not only for academic success, but also importantly for life success.

## References

- Angelo, T. (1995). *Reassessing (and Defining) Assessment*. The AAHE Bulletin, 48(2), 7-9.
- Biggs, J. (2011). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, 1, 15 -22.
- Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Berkshire: Society for Research in Higher Education & Open University Press.
- Bresciani, M. J., Gardner, M. M, & Hickmott, J. (2012). *Demonstrating student success : A practical guide to outcomes-based assessment of learning and development in student affairs*. Sterling, USA: Stylus Publishing.
- Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2009). *Demonstrating student success in student affairs*. Sterling, VA: Stylus Publishing.

- Bresciani, M. J. (2006). *Outcomes-based academic and co-curricular program review: A compilation of institutional good practices*. Sterling, VA: Stylus Publications.
- Killen, R. (2000). Outcome-based education: Principles and possibilities. Unpublished manuscript, University of Newcastle, Faculty of Education.
- Morke, A. M., Dornan, T., & Eika, B., (2013). Outcome (competency) based education: An exploration of its origins, theoretical basis, and empirical evidence. *Adv in Health Sci Educ*, 18, 851–863. DOI.1007/s10459-012-9405-9
- Niebuhr, G. A. (1996). *Lifelong learning through a National Qualifications Framework: A discussion document*. UK: Pretoria.
- Ramoroka, N. J. (2006). *Educators' understanding on the premises underpinning outcomes-based education and its impact on their classroom assessment practices*. UK: University of Pretoria.
- Spady, W. (1994). *Outcome-based education: Critical issues and answers*. Arlington, VA: American Association of School Administrators.
- Spady, W., & Marshall, K. (1991). Beyond Traditional Outcome-Based Education. *Educational Leadership*, 49(2), 67–72.
- Tucker, B. (2004, October 19). *Literature review: Outcomes-focused education in universities*. Retrieved from <http://lsn.curtin.edu.au/outcomes/docs/LitReview.pdf>



## Assessing Achievement Goals as Mediator on the Effect of Perfectionism on Procrastination

**Carlo Magno**

*Mapua University*

**Queen Marie Katrina O. Alcantara**

*Nacoma Services*

**Christian Dave N. Domingo**

*Xavier School*

**Maria Queenie Y. Isidro**

*Ateneo Center for Organizational Research and Development*

### Abstract

The present study advances the role of achievement goals on perfectionism and procrastination. It was hypothesized in the present study that achievement goals mediate the effect of perfectionism on procrastination. The 2 x 2 achievement goals framework by Elliot and McGregor (2001) was used to determine how achievement goals are predicted by perfectionism. The Achievement Goal Questionnaire-Revised (AGQ-R), Multidimensional Perfectionism Scale (MPS), and Procrastination Assessment Scale-Students (PASS) were administered to 349 Filipino college undergraduate students. The mediational analysis was conducted following the procedure by Hayes (2009) using PROCESS. Results showed only performance and avoidance goal orientations act as mediators between the effects of perfectionism on procrastination. Theoretical insights regarding the academic and contextual conceptualizations of achievement goals and perfectionism were discussed.

*Keywords:* Perfectionism, achievement goals, procrastination

### Introduction

Some students delay their work depending on the motivation they have on a task. Reasons for delaying tasks involve the feeling that students do not have the competence to complete it, when the intention is mainly to outperform others, or when students fear of performing poorly compared to

others. Delay of one's work is referred to as procrastination (Solomon & Rothblum, 1984). Certain sources of procrastination are explained in the motivation literature as achievement goals. The goal orientation theory or achievement goal theory states that these [achievement goals] are "competence-relevant" goals that a person aims for in a setting where one sees the opportunity to achieve (Elliot, Maier, & Pekrun, 2009). These goals are influential because they help individuals in constructing possible future achievement outcomes or events, and help them in engaging into more productive and adaptive ways of learning. Recently, a 2 x 2 achievement goal framework was proposed by Elliot and McGregor (2001) where the mastery/performance dimension are combined with the approach/avoidance dimension to come up with four types of goal orientation namely mastery approach goals, mastery avoidance goals, performance approach goals, and performance avoidance goals. The notion of an avoidance goal shown in studies was an intuitive aspect in explaining procrastination (Ferrari, Johnson, & McCown, 1995). It has been consistent in various studies that avoidance goals are positively related with procrastination. This was further clarified in a competitive model by Seo (2009) where mastery avoidance and performance avoidance were used as predictors of procrastination.

The type of achievement goals that students adopt varies depending on the standards individuals set for themselves. These standards are defined as perfectionism (Hewitt & Flett, 1991). This claim is supported by the previous studies that investigated the effect of perfectionism on achievement goals (Gaudreau & Verner-Fillion, 2010). Results from the previous studies showed that Self-Oriented Perfectionism (SOP) is positively related with achievement goals, particularly on the mastery-approach and performance-approach goals (Gaudreau & Verner-Filion, 2010). Socially Prescribed Perfectionism (SPP), on the other hand, showed a positive relationship with mastery avoidance, performance-approach, and performance-avoidance goals (Gaudreau & Verner-Filion, 2010). Most of what the researches have done that examine perfectionism and achievement goal orientations together seem to be focused in performance in sports (e. g., Hotham, Stoeber, & Uphill, 2009), and studies about achievement goals and perfectionism in an academic context needs to be further investigated (Elliot & McGregor, 2001; Gaudreau & Verner-Filion, 2010).

Another important development in the achievement goal theory literature is the studies that supported the relationship between achievement goals and procrastination (e. g., Howell & Buro, 2009; Howell & Watson, 2007). These studies have shown that that the delay of one's work due to failure in their self-regulatory processes is associated with their mastery-

avoidance and performance-avoidance goal orientations. This provides evidence that achievement goals explain variation in procrastination.

Previous studies also showed that one's perfectionist tendencies are said to be associated with students' procrastination (e. g., Onwuegbuzie, 2000; Ferrari & Tice, 2000); however, the said relationship still seems to be equivocal due to the inconsistencies in the results between these two variables (Blankstein, Flett, Hewitt, & Koledin, 1992).

In recent reviews, there seem to have no studies done where the relationship between perfectionism and procrastination is mediated by achievement goals as added effects. Hence, the present study proposed that achievement goals can act as a mediator that explains the equivocal relationship of the dimensions of perfectionism and procrastination. One major strength of the present study is the hypothesized prediction among the variables rather than just correlations. Hence, if perfectionism predicts achievement goals, and achievement goals in turn predict procrastination, it is possible to come up with a more integrative model involving these three variables.

The current study utilized the 2 x 2 achievement goal framework by Elliot and McGregor (2001). The 2 x 2 achievement goal framework posits that achievement goals influence students' performance depending on how they view their learning and how much they engage in the task (Covington, 2000). Achievement goals have four basic factors namely mastery approach, mastery avoidance, performance approach, and performance avoidance. Mastery approach goal is a type of goal orientation wherein an individual engages in tasks with the purpose of seeking to learn and to master the task (Elliot & McGregor, 2001). Mastery avoidance goal is a type of goal orientation wherein an individual avoids tasks due to feelings of incompetence and incapability of accomplishing the task (Elliot & McGregor, 2001). Performance approach goal is a type of goal orientation wherein an individual engages in tasks with the purpose of outperforming others (Elliot & McGregor, 2001). Performance avoidance goal is a type of goal orientation wherein an individual avoids tasks due to fear of performing poorly relative to others (Elliot & McGregor, 2001). The present study is focused on investigating the mediating role of the achievement goals on the effect of perfectionism on procrastination.

Hewitt and Flett's (1991) conceptualized perfectionism with three subscales: Self-oriented perfectionism (SOP), socially prescribed perfectionism (SPP) and other-oriented perfectionism (OOP). Self-oriented perfectionism means that individuals have a strong motivation to be perfect and have high expectation to one's self and to one's performance. On the other hand, socially prescribed perfectionism means that individuals believe that others have set high expectations and high critical evaluation towards them and their

performance. Lastly, other-oriented perfectionism means that individuals set high standards towards their significant others (Hewitt & Flett, 1991). The present study only included the self-oriented perfectionism and socially prescribed perfectionism, as these two factors are focused on the perfectionistic expectations of individuals towards themselves. Other-oriented perfectionism was not included in the present study because perfectionism in the present study is oriented towards others and not to the self (Chang & Rand, 2000).

According to Solomon and Rothblum (1984), procrastination is the act of intentionally delaying a task to the point of distress. In the study done by Howell and Watson (2007), procrastination is the outcome of achievement goals. While, in another study, achievement goals were used as a mediator between perfectionism and academic achievement (Gaudreau & Verner-Filion, 2010), providing evidence that perfectionism is a predictor of achievement goals. Given this sequence, an integrative model was built where achievement goals mediates the effect of the factors of perfectionism on procrastination. This sequence explains students who set high standards for themselves adapt specific learning goals on a task. When standards and goals are set, students do not likely procrastinate or delay their work.

### **Self-Oriented Perfectionism, Achievement Goals and Procrastination**

**Self-oriented perfectionism, mastery approach, and procrastination.** When individuals set high standards for themselves, they seek to engage on their tasks with the purpose of learning and mastering the tasks, and hence they do not delay their work. This means that mastery approach goal mediates the effect of self-oriented perfectionism on procrastination. This assertion was supported by the previous studies done by Howell and Watson (2007) and Gaudreau and Verner-Fillion (2010) where the results of their study indicate a positive relation between self-oriented perfectionism and mastery approach goal, as well as a negative correlation between mastery approach goal and procrastination. Furthermore, since one common element between self-oriented perfectionism and mastery approach goal orientation is a strong intrinsic motivation; such individuals will engage on their tasks and therefore will be less associated with procrastinating behaviors.

**Self-oriented perfectionism, mastery avoidance, and procrastination.** When individuals set high standards for themselves, they do not avoid tasks in which they know that they will fail as they consider the task as a challenge, but they will tend to delay their work. This shows that mastery

avoidance mediates the effect of self-oriented perfectionism on procrastination. This conceptualization was shown to be feasible in the study done by Gaudreau and Verner-Fillion (2010) and Howell and Watson (2007) where they showed that self-oriented perfectionism is negatively related with mastery avoidance goals, and in turn, mastery avoidance goal showed a positive relationship with procrastination.

**Self-oriented perfectionism, performance approach, and procrastination.** When individuals set high standard for themselves, they seek to engage in tasks with the purpose of outperforming others, but they tend to delay their work. This means that performance approach mediates the effect of self-oriented perfectionism on procrastination. It was mentioned in the study done by Gaudreau and Verner-Filion (2010) that individuals who have a performance approach goal orientation are just interested in achievement outcomes, but not on achievement activities. Hence, even though self-oriented perfectionists engage in tasks, if their purpose for engaging is not to learn, but rather to surpass the performance of others, they will have the tendency to postpone their work.

**Self-oriented perfectionism, performance avoidance, and procrastination.** When individuals set high standard for themselves, they tend to avoid tasks that will make them perform poorly relative to others, and hence, they tend to delay their work. This indicates that performance avoidance mediates the effect of self-oriented perfectionism on procrastination. According to Solomon and Rothblum (1984), one common characteristic between perfectionists and procrastinators is their extreme fear of failure. Therefore, even though self-oriented perfectionists have a strong motivation, such said individuals will still have the tendency to procrastinate as they constantly avoid tasks that will make them demonstrate poor performance relative to others.

### **Socially-Prescribed Perfectionism, Achievement Goals, and Procrastination**

**Socially-prescribed perfectionism, mastery approach goals and procrastination.** When individuals believe that others expect high standard on them, they do not engage in tasks because the intrinsic interest is not coming from themselves, but from others. They tend to delay their work. This indicates that mastery approach mediates the effect of socially prescribed perfectionism on procrastination. According to Gaudreau and Verner-Fillion

(2010), mastery approach goals are positively associated with the individual's interests and the improvement of their self-regulation. However, since the drive of socially-prescribed perfectionists is coming from others and not from themselves, they do not engage in tasks, and therefore having the tendency of delaying their work.

**Socially-prescribed perfectionism, mastery avoidance goals and procrastination.** When individuals believe that others have high standard on them, they tend to avoid tasks which they think they will fail. Therefore, they tend to delay their work. This implies that mastery avoidance goals mediate the effect of socially prescribed perfectionism on procrastination. One example provided by Pintrich (2000) regarding mastery avoidance is the experience of students who have perfectionistic standards. It was mentioned that perfectionists do not want to make mistakes to avoid appearing incompetent and incapable, because of their fear of failure, socially-prescribed perfectionists tend to delay their work.

**Socially-prescribed perfectionism, performance approach goals, and procrastination.** When individuals believe that others have high standard on themselves, they engage in tasks to perform better than others, but they tend to delay their work. This means that performance approach goals mediate the effect of socially prescribed perfectionism on procrastination. Previous studies have confirmed that there is a positive relationship between socially prescribed perfectionism and procrastination (e. g., Blankstein, Flett, Hewitt, & Koledin, 1992). In addition, studies done on perfectionism and achievement goals showed that there is a positive relationship between SPP and performance approach goals (e. g., Stoeber, Stoll, Pescheck, & Otto, 2008). Given this, conceptualization of the relationship of these three variables can be studied.

**Socially-prescribed perfectionism, performance avoidance goals, and procrastination.** When individuals believe that others have high standard on themselves, they tend to avoid tasks that will make them perform poorly relative to others, and hence, they tend to delay their work. This denotes that performance avoidance goals mediate the effect of socially prescribed perfectionism on procrastination. Since it was asserted that SPP has a positive relationship with procrastination, and that socially prescribed perfectionists is characterized by intense fear of failure, then such individuals engage in procrastination with the purpose of avoiding negative performances relative to others.

## The Present Study

The reviews points out how different achievement goals are predicted by and perfectionism and as a predictor of procrastination. The nature of achievement goals as a mediator for perfectionism and procrastination is explained in the incorporation of intrinsic motivation and tendency to approach rather than to avoid achievement situations in perfectionism (Flett, Blankstein, & Martin, 1995). This notion suggests a negative direction with variables such as procrastination. If perfectionism would be studied with procrastination, the notion of approach and avoidance needs to be part of the model tested. The incorporation of the tendency to approach or avoid situation needs to be demonstrated as a separate construct with that of perfectionism. The purpose of this study is to test whether the factors of achievement goals mediate the effect of the two factors of perfectionism on procrastination.

## Method

### Participants

The participants in the present study are undergraduate students from private universities in Manila. There are 349 students (statistical power of 1.00) 209 females and 140 males with mean age of 18.29. The current study made use of a purposive sampling to gather respondents to participate in the study. The criteria included in the sample are college students that are currently enrolled in a college/university, have engaged in doing several assignments and projects in school, and have experienced joining extracurricular activities in school. This ensures that they may at least have experienced procrastinating in a school-related task.

### Instruments

**Hewitt and Flett's Multidimensional Perfectionism Scale (HFMPs).** The HFMPs is an instrument with 45 items that has three dimensions (Hewitt & Flett, 1991). This scale tests a person's perfectionism on three accounts, when they have high standards for themselves (SOP), when they have high standards for their significant others (OOP), and when they think others have high standards for them (SPP). The scale is rated by the use of 7-point Likert scale (*7 being strongly agree and 1 being strongly disagree*) where each subscale/dimension has 15 items in the scale. The scale has a test reliability coefficient of .88, .85 and .75 for the subscales SOP, OOP and SPP

respectively (Hewitt, Flett, Turnbull-Donovan & Mikail, 1991). Clara, Cox, and Enns (2002) tested the questionnaire in both university and clinical samples and reported the model has a GFI value of .71 for both sample areas.

**Procrastination Assessment Scale-Students (PASS).** The PASS is an instrument developed by Solomon and Rothblum (1984) to check for the procrastinating tendencies of college students. This scale is divided into two sections. The first section attempts to look for the occurrence of procrastination in college students when (1) writing a term paper, (2) studying for an exam, (3) keeping up with weekly reading assignments, (4) performing administrative tasks, (5) attending meetings, and (6) performing academic tasks in general. The questionnaire can be answered by a 5-point Likert scale ranging from 5 –always procrastinate to 1 – never procrastinate. Another set of 5-point scale measures the degree to which procrastination causes a problem on the given task (5 – always a problem to 1 – not at all a problem). And the last set of 5-point Likert scale is used to measure the degree to which they want to decrease procrastinating on the said academic task (5 – definitely want to decrease to 1 – do not want to decrease).

The second section of the PASS provides a scenario for students to respond where they would probably procrastinate (Solomon & Rothblum, 1984). There are 26 items answered in a 5-point Likert scale (5 – definitely the reason to 1 – not at all the reason). The statements measure how much respondents delay their task the last time they are in the given scenario. The PASS arrived with one score to represent the students' procrastination tendencies. PASS has a coefficient of .80 for the test retest correlation (Yong, 2010) and a coefficient alpha of approximately .77 (Yao, 2009).

**Achievement Goal Questionnaire Revised (AGQ-R).** The AGQ-R is a 12 item questionnaire that measures ones achievement goal orientation (Elliot & Murayama, 2008). It is a recent development from AGQ (Elliot & McGregor, 2001) which was replaced due to concerns about statistical validity and reliability. The questionnaire was meant to measure short but reliable and valid measures of the four types of achievement goals (mastery-approach goal, mastery-avoidance goal, performance-approach goal, and performance avoidance goal). There are three for each of the different types of achievement goals, and is measured using a 7-point Likert scale (7 – very true of me to 1 – not at all true of me). Elliot and Murayama (2008) checked for confirmatory factor analysis which was a value of .99. Their internal reliabilities were as follows: performance-approach goals = .83, mastery-avoidance goals = .87,

mastery-approach goals = .85 and performance avoidance goals = .77 (Eren, 2009).

## Procedures

The researchers asked potential participants if they are willing to participate in the study. Once they have agreed, they were briefed about the study, and were given instructions on how to answer the questionnaires. After which, they were asked to sign the demographic information sheet as a proof of their willingness to participate. The individuals who agreed to participate are asked to complete three sets of measure, the HFMPs, PASS, and AGQ-R and a sheet for the demographic information. The respondents were given 15-25 minutes to answer the questionnaire and were asked to return the papers immediately upon completion. Once all the questionnaires were returned, the respondents were thanked by the researchers and they were informed that the researchers are willing to answer any questions regarding their participation in the study.

## Results

The means and standard deviations for the factors of perfectionism, achievement goal orientations, and procrastination were determined. Pearson  $r$  was used to determine the relationship among the factors of perfectionism, achievement goal orientations, and procrastination. Lastly, the mediation analysis following the steps by Baron and Kenny (1986) was used. Specifically, the mediation tested whether each of the achievement goal orientation mediates effect of the factors of perfectionism on procrastination.

All factors of achievement goal orientations obtained high mean values, with performance avoidance goal ( $M = 5.34$ ,  $SD = 1.26$ ) attaining the highest mean and variability. For the factors of perfectionism, both self-oriented perfectionism ( $M = 4.66$ ,  $SD = .81$ ) and socially prescribed perfectionism ( $M = 4.14$ ,  $SD = .63$ ) also attained high mean values. Procrastination ( $M = 2.88$ ,  $SD = .44$ ) got the lowest mean and variability among all the variables. The whole achievement goal orientation scale was able to obtain an alpha of .87, while whole procrastination scale obtained an alpha of .86 and the whole perfectionism scale had an alpha of .80, which are all good indicators of scale internal consistency.

Among the two factors of perfectionism, only self-oriented perfectionism showed a significant relationship with all types of goal orientations (mastery approach,  $r = .36$ ,  $p < .05$ ; mastery avoidance,  $r = .19$ ,  $p < .05$ ; performance approach,  $r = .38$ ,  $p < .05$ ; performance avoidance,  $r = .24$ ,

$p < .05$ ), while socially prescribed perfectionism showed a significant relationship only with the two performance goal orientations (performance approach  $r = .22$ ,  $p < .05$ ; performance avoidance  $r = .21$ ,  $p < .05$ ). It was also confirmed that self oriented perfectionism and socially prescribed perfectionism has a significant relationship with procrastination ( $r = -.14$ ,  $p < .05$ ;  $r = .16$ ,  $p < .05$  respectively).

Table 1

*Means, Standard Deviations, Cronbach's Alpha, and Correlations of the Achievement Goals, Factors of Perfectionism and Procrastination*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 Mastery Approach	---						
2 Mastery Avoidance	<b>.46*</b>	---					
3 Performance Approach	<b>.53*</b>	<b>.35*</b>	---				
4 Performance Avoidance	<b>.32*</b>	<b>.41*</b>	<b>.65*</b>	---			
5 Self Oriented Perfectionism	<b>.36*</b>	<b>.19*</b>	<b>.38*</b>	<b>.24*</b>	---		
6 Socially Prescribed Perfectionism	.05	.01	<b>.22*</b>	<b>.21*</b>	<b>.21*</b>	---	
7 Procrastination	-.01	-.06	.02	.11*	<b>-.14*</b>	<b>.16*</b>	---
8 <i>M</i>	5.65	4.94	5.30	5.34	4.66	4.14	2.8
9 <i>SD</i>	.99	1.29	1.12	1.26	.81	.63	.44
10 Cronbach's Alpha	.82	.71	.81	.81	.84	.65	.86

\* $p < .05$

PROCESS was used to test whether the four domains of achievement goals mediate the effect of self-oriented perfectionism and socially prescribed perfectionism on procrastination. The mediation model was tested following Hayes (2009) procedure. Direct and indirect effects were obtained using an Ordinary Least Square for parallel mediators.

There was a direct effect between SOP and SPP with procrastination. SOP had a significant path coefficient for all four achievement goals. On the other hand, SPP only had a significant path coefficient for performance approach and performance avoidance. Only performance approach and performance avoidance had significant paths to procrastination.

Significant indirect effects were obtained for SOP on procrastination through performance approach and avoidance. The same pattern of indirect effects for SPP on procrastination was obtained through performance approach and avoidance.

Table 2  
*Coefficients of Direct Effects*

Mediators	SOP	SPP	Procrastination
Mastery Approach	<b>0.21*</b>	0.01	0.02
Mastery Avoidance	<b>0.15*</b>	0.01	0.01
Performance Approach	<b>0.30*</b>	<b>0.16*</b>	<b>0.19*</b>
Performance Avoidance	<b>0.11*</b>	<b>0.18*</b>	<b>0.16*</b>
SOP	--	--	<b>-.12*</b>
SPP	--	--	<b>.15*</b>
<b>Indirect Effects</b>			
Indirect effect of SOP (through Mastery Approach)	.06		
Indirect effect of SOP (through Mastery Avoidance)	.04		
Indirect effect of SOP (through Performance Approach)	<b>.20*</b>		
Indirect effect of SOP (through Performance Avoidance)	<b>.23*</b>		
Indirect effect of SPP (through Mastery Approach)	.08		
Indirect effect of SPP (through Mastery Avoidance)	.08		
Indirect effect of SPP (through Performance Approach)	<b>.22*</b>		
Indirect Effect of SPP (through Performance Avoidance)	<b>.24*</b>		

\* $p < .05$

## Discussion

The purpose of the present study is to test whether the types of achievement goals can mediate the effect of the factors of perfectionism on procrastination. The results of the correlation showed that the factors of perfectionism are related with procrastination. The self-oriented perfectionism factor showed a significant positive relationship with all types of achievement goal orientations, while the socially-prescribed perfectionism factor is only associated with certain types of achievement goal orientations. Lastly, only the two performance goal orientations are related with procrastination. The

findings from the mediation analysis revealed that performance approach goal and performance avoidance goal orientations act as significant mediators between factors of perfectionism and procrastination.

The present study confirmed the previous assertion that self-oriented perfectionism is negatively correlated with procrastination while socially-prescribed perfectionism is positively correlated with procrastination. This finding confirms the study done by Blankstein, Flett, Hewitt, and Koledin (1992) where they stated that one common component present in both socially-prescribed perfectionism and procrastination is extreme fear of failure. All types of achievement goal orientations are positively related with self-oriented perfectionism. Only the two performance goal orientations showed a significant positive relationship with socially prescribed perfectionism.

The findings from the mediation analysis revealed that among the types of achievement goal orientations, only performance approach goal and performance avoidance goal orientation act as mediators when perfectionism factors predict procrastination.

It was found that the performance approach goal orientation mediated the effect of self-oriented perfectionism on procrastination. The results of the present study is supported by the previous study done by Gaudreau and Verner-Filion (2010) where they asserted that individuals who adopt a performance approach goal orientation are just interested in achievement outcomes, but not in achievement activities. Hence, even though self-oriented perfectionists engage in their tasks, if their orientation is not coming from oneself but from external factors, they will still postpone their work.

The performance avoidance goal orientation mediated the effect of self-oriented perfectionism in predicting procrastination. This means that individuals will have that tendency of delaying their work even if they have high standards for themselves, as they will not engage in tasks in order to avoid performing poorly relative to others. The study of Solomon and Rothblum (1984) supports the results of the present study as they indicated that one common characteristic between perfectionists and procrastinators is their extreme fear of failure. Hence, even though self-oriented perfectionists have a strong motivation to achieve, they will still procrastinate because of their fear of appearing incompetent relative to others.

When mastery goal orientations were used to mediate the effect of the factors of perfectionism on procrastination, both mastery approach and mastery avoidance goals did not turn out as mediators. One reason could be that these goals are considered to be more adaptive as compared to performance goal orientations, and this assertion is supported by previous studies as mastery goals are positively correlated with task interest, task

engagement, deep processing learning, and academic achievement (Ames, 1992; Bernardo, 2008; Harackiewicz, Barron, Carter, Lehto, & Elliot, 1997; McGregor & Elliot, 2002).

In relation to the present study, both mastery goals did not act as mediators between self-oriented perfectionism and procrastination. However, self-oriented perfectionism showed a direct link with procrastination, and a significant positive relationship with both mastery goals. Given this, individuals who have high standards for themselves will engage on their tasks with the purpose of learning and mastering the task (mastery approach), or they will avoid behaviors that will make them commit mistakes, and be seen as incompetent or a failure (Brophy, 2005). However, the findings suggest that it does not matter whether such individuals will adopt a mastery approach goal or mastery avoidance goal, for as long as they will set high standards for themselves, their tendency to procrastinate is reduced.

The relationship among socially-prescribed perfectionism, mastery goals, and procrastination are more complex. In the present study, socially-prescribed perfectionism did not show a significant relationship with both mastery goals, and these mastery goals did not show a significant relationship with procrastination either, but a direct effect between socially-prescribed perfectionism and procrastination is evident. These results suggest that individuals who believe that others have high standards for them will tend to procrastinate even if they do not adopt these mastery goal orientations at all. The positive relationship between socially-prescribed perfectionism and procrastination is also consistent with the result of the study done by Blankstein, Flett, Hewitt, and Koledin (1992) who also found that among the subscales of HF's multidimensional perfectionism, socially-prescribed perfectionism is the most associated with procrastination. The significant relationship between these two constructs is expected because procrastination is a maladaptive behavior and among the subscales of HF multidimensional perfectionism scale, SPP is considered to be a form of maladaptive perfectionism (Frost, Heimberg, Holt, Mattia, & Neubauer, 1993). However, the non-significant relationship of socially prescribed perfectionism with mastery avoidance and mastery approach goals is inconsistent with recent literatures. In the study done by Hewitt, Flett, and Endler (1995), they support the positive relationship between socially prescribed perfectionism and mastery avoidance goal, as they asserted that such said perfectionists would be more likely "avoid rather than approach a problem." On the other hand, the findings from the study of Gaudreau and Verner-Fillion (2010) suggested the negative relationship of socially prescribed perfectionism and mastery approach goals, as such said perfectionists would only engage in tasks with the purpose of

meeting the high standards imposed by others, and therefore they are not concerned of learning and mastering these tasks. The non-significant relationship between socially prescribed perfectionism and mastery goals can be explained by the extent on which culture influences our conceptualization of these two constructs.

The results of the present study also indicated that performance approach goal did not turn out as a mediator of socially prescribed perfectionism predicting procrastination. However, socially-prescribed perfectionism showed a direct link with procrastination, and a significant positive relationship with performance approach goal. This finding suggests that individuals who believe that others have high standards for them will still tend delay their tasks, even if they do not adopt a performance approach goal orientation at all. As mentioned earlier, the direct effect of socially-prescribed perfectionism and procrastination can be explained by the study done by Blankstein, Flett, Hewitt, and Koledin (1992) for they asserted that the fear of failure is what makes these socially-prescribed perfectionists to delay their tasks. However, the result also indicated that there is a significant positive relationship between socially prescribed perfectionism and performance approach goal, indicating that when individuals believe that others have high standards for themselves, they will engage in the tasks but with the purpose of outperforming others. This finding is supported by the study done by Gaudreau and Verner-Fillion (2010) where they asserted that socially-prescribed perfectionists will also engage in their tasks in the hope of avoiding failure, but this engagement are only concerned with attaining achievement outcomes (e. g., higher grades, academic recognition) and not on achievement activities (e. g., task engagement, task interest). To make the relationship between socially prescribed perfectionism and performance approach goal more cultural-specific, the study done by Bernardo (2008) mentioned that the achievement goals being adopted by Filipinos have its social dimensions, where, to an extent, “significant others and groups define the goals, standards, means of goals attainment, and acceptance of achievement outcomes” of individuals. The result of this study suggested that the performance approach goals of most Filipino students are defined by their personal-standards oriented (individual’s standard for himself), parent-oriented (wishes, goals, and standards set by parents), and teacher-oriented (wishes, goals, and standards set by teachers) motivations.

The findings of the present study advances theoretical insights in the relationships among perfectionism, achievement goals, and procrastination in three ways: (1) Individuals who have high standards for themselves will not delay their tasks, unless they want to outperform others or they want to avoid

performing poorly relative to others, (2) Individuals who believe that others have high standards for them will have the tendency to delay their work, more so when they do not engage to avoid performing poorly relative to others, and (3) Not all achievement goals would lead both types of perfectionists to delay their work, as collectivist values and factors may play an important role in moderating the influence of achievement goal construct and perfectionism on the self.

The results of the study brings a theoretical insight that individuals who have high standards for themselves will not delay their work, unless they want to outperform others or they want to avoid performing poorly relative to others. Setting high standards on the self indicates task engagement and intrinsic motivation, as individuals will always strive to achieve in the hope of fulfilling these self-imposed standards (e. g., Kilbert, Langhinrichsen-Rohling, & Saito, 2005; Gaudreau & Verner-Filion, 2010). The tendency to not procrastinate lies behind the reason of engaging into the tasks. If individuals have high standards for themselves will engage with the purpose of mastering the task, their tendency to procrastinate is lessened (e. g., Howell & Watson, 2007; Gaudrea & Verner-Filion, 2010). However, individuals will delay their tasks even if they have high standards for themselves when their reason for their engagement is to just outperform others or to avoid demonstrating incompetence as compared to others (e. g. Solomon & Rothblum, 1984; Gaudrea & Verner-Filion, 2010).

Another theoretical insight explains that individuals who believe that others have high standards for them will have the tendency to delay their work, more so when they do not engage on their tasks to avoid performing poorly relative to others. Individuals will tend to exhibit procrastinating tendencies if they are driven to work not by their own standards, but by the standards imposed by others on them (Blankstein, Flett, Hewitt, & Koledin, 1992). Failing to meet the standards imposed by others make these kinds of individuals fearful and avoidant of the tasks that they have to do, as they do not want others to associate their failures with their lack of competence or ability in comparison with others (Ferrari & Tice, 2000).

Another theoretical insight explains that not all achievement goals would lead both types of perfectionists to delay their work, as collectivist values and factors may play an important role in moderating the influence of achievement goal construct and perfectionism on the self. As mentioned in the study done by Bernardo (2008), mastery approach goals and performance goals has its individual and social dimensions. In the individual dimension, one's personal performance standard is what motivates individuals to engage on their tasks. On the other hand, the social dimensions pertain to the parent-oriented

and teacher-oriented motivations that were also considered by individuals as these help them to perform better on their tasks. In relation to the present study, the significant positive relationship between socially prescribed perfectionism and performance approach goal orientation is believed to have its cultural-specific underpinnings. Also, the result that mastery goals did not have a significant relationship with socially prescribed perfectionism is inconsistent with the results from the existing literature. However, the researchers believe that the lack of significant relationship between these two variables can be stemmed from the extent to which collectivistic values and practices such as fulfilling role obligations, “basking as a reflection of glory”, maintaining harmonious relationships, and seeing one’s self as important member of social in-groups play an important role when dealing with conceptualizations of socially prescribed perfectionism and achievement goal motivations in the Philippines, as these cultural (particularly collectivistic) factors can explain reason as to why not all achievement goals mediate the effect of factors of perfectionism on procrastination.

### References

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–271.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.
- Blankstein, K., Hewitt, P., Flett, G., & Koledin, S. (1992). Components of perfectionism and procrastination in college students. *Social Behavior and Personality, 20*(2), 85-94.
- Bernardo, A. B. I. (2008). Individual and social dimensions of Filipino students’ achievement goals. *International Journal of Psychology, 43*(5), 886-891.
- Bernardo, A., Salanga, M., & Aguas, K. (2008). Filipino adolescent students’ conceptions of learning goals. In O. S. Tan, D. M. McInerney, A. D. Liem, & A. G. Tan (Eds.), *What the West can learn from the East: Asian perspectives on the psychology of learning and motivation* (pp. 169-190). Greenwich, CT: Information Age Publishing.
- Brophy, J. (2005). Goal theorists should move on from performance goals. *Educational Psychologist, 40*, 167-176.

- Chang, E. C., & Rand, K. L. (2000). Perfectionism as a predictor of subsequent adjustment: evidence for a specific diathesis-stress mechanism among college students. *Journal of Counseling Psychology, 47*(1), 129-137.
- Church, A. T., & Katigbak, M. S. (1992). The cultural context of academic motives: A comparison of American and Filipino college students. *Journal of Cross-Cultural Psychology, 23*, 40-58.
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review Psychology, 51*, 171-200.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 X 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519.
- Elliot, A. J., Maier, M., & Pekrun, R. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 10*, 115-135.
- Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*, 613-628.
- Eren, A. (2009). Examining the teacher efficacy and achievement goals as predictors of Turkish student teachers' conceptions about teaching and learning. *Australian Journal of Teacher Education, 4*, 69-87.
- Ferrari, J. R., & Tice, D. M. (2000). Procrastination as a self-handicap for men and women: A task-avoidance strategy in a laboratory setting. *Journal of Research in Personality, 34*, 73-83.
- Flett, G., Blankstein, K. R., & Martin, T. R. (1995). Dimensions of perfectionism and procrastination. In J. R. Ferrari, J. L. Johnson, & W. G. McCown (Eds.), *Procrastination and task avoidance: Theory, research, and treatment* (pp. 21-34). NY: Plenum.
- Frost, R. O., Heimberg, R. G., Holt, C. S., Mattia, J. L., & Neubauer, A. L. (1993). A comparison of two measures of perfectionism. *Personality and Individual Differences, 14*, 119-126.
- Frost, R. O., Marten, P. A., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research, 14*, 449-468.
- Gaudrea, P., & Verner-Filion, J. (2010). From perfectionism to academic adjustment: The mediating role. *Personality and Individual Differences, 49*, 181-186.
- Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. L., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology, 73*, 1284-1295.

- Hewitt, P. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, *60*, 456–470.
- Hewitt, P. L., Flett, G. L., & Endler, N. S. (1995). Perfectionism, coping and depression symptomatology in a clinical sample. *Clinical Psychology and Psychotherapy*, *2*, 47-58.
- Hewitt, P. L., Flett, G. L., Turnbull-Donovan, W., & Mikail, S. F. (1991). The multidimensional perfectionism scale: Reliability, validity, and psychometric properties in psychiatric samples. *A Journal of Consulting and Clinical Psychology*, *3*(3), 464-468.
- Hotham, S., Stoeber, J., & Uphill, M. (2009). Predicting race performance in triathlon: The role of perfectionism, achievement goals, and personal goal setting. *Journal of Sport & Exercise Psychology*, *31*, 211-245.
- Howell, A., & Buro, K. (2009). Implicit beliefs, achievement goals, and procrastination: A meditational analysis. *Learning and Individual Differences*, *19*, 151-154.
- Howell, A. J., & Watson, D. C. (2007). Procrastination: Associations with achievement goal orientation and learning strategies. *Personality and Individual Differences*, *43*, 167–178.
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, *30*(2), 3-13.
- Kilbert, J. J., Langhinrishen-Rohling, J., & Saito, M. (2005). Adaptive and maladaptive aspects of self-oriented versus socially prescribed perfectionism. *Journal of College Student Development*, *46*, 141-156.
- McGregor, H. A., & Elliot, A. J. (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Educational Psychology*, *94*, 381–395.
- Onwuegbuzie, A. J. (2000). Academic procrastinators and perfectionistic tendencies among graduate students. *Journal of Social Behavior and Personality*, *15*(5), 103-110.
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, *3*, 544–555.
- Salanga, M.G.C. (2003). *Perfectionism in Filipino university students*. Unpublished masters thesis, De La Salle University, Manila, Philippines.
- Seo, E. H. (2008). Self-efficacy as a mediator in the relationship between self-oriented perfectionism and academic procrastination. *Social Behavior and Personality*, *36*(6), 753-764.
- Seo, E. H. (2009). The relationship of procrastination with a mastery goal versus an avoidance goal. *Social Behavior and Personality*, *37*(7), 911-919.

- Solomon, L. J., & Rothblum, E. D. (1984). Academic procrastination: frequency and cognitive-behavioural correlates. *Journal of Counseling Psychology, 31*, 504–510.
- Stoeber, J., Stoll, O., Pescheck, E., & Otto, K. (2008). Perfectionism and achievement goals in athletes: Relations with approach and avoidance orientations in mastery and performance goals. *Psychology of Sport and Exercise, 9*, 102–121.
- Yao, M. P. (2005). *An exploration of multidimensional perfectionism, academic self-efficacy, procrastination frequency, and Asian American cultural values in Asian American university students*. Unpublished master's thesis, The Ohio State University.
- Yong, F. L. (2010). A study on the assertiveness and academic procrastination of English and communication students at a private university. *American Journal of Scientific Research, 6*, 62-71.



## A Cross-Sectional Analysis of Classroom Assessment Literacy of English Teachers in Secondary and Tertiary Levels

**Tristan L. Piosang**

*National University, Manila*

### **Abstract**

The present study reported on the classroom assessment literacy of secondary English teachers (SET) and tertiary English teachers (TET). It specifically compared the level of assessment literacy of both groups. Classroom assessment literacy (CAL) was measured using Classroom Assessment Literacy Inventory (CALI). The results showed that teachers in basic education had higher means on assessment literacy than teachers in college, though, both have poor CAL ( $M=13.78$ ,  $M=10.87$ ). On the level of CAL, 85% of the SET group had poor CAL and 15% had fair CAL. For TET group, 88% had poor CAL, 5% had fair CAL, and 7% had good CAL. Results of t-test for independent samples showed that there is a significant difference between the level of classroom assessment literacy of SET and TET,  $p=0.04$ .

*Keywords:* Assessment Literacy, Language Teachers, CALI, Classroom Assessment

### **Introduction**

Assessment plays a critical role in the 21<sup>st</sup> century educational landscape as it serves critical purposes more than just mere measurement of student achievement. Ewell (2009) has made a distinction on the two critical purposes assessments serve which he termed as *assessment paradigms* in the 21<sup>st</sup> century, namely: (1) Assessment for Improvement Paradigm and (2) Assessment for Accountability Paradigm. On the first paradigm, Ewell (2009) argues that assessment is used for the improvement of learning and instruction. The second paradigm emphasizes accountability on different levels – from classroom to national – wherein teachers, school heads, and the government are held accountable for student achievement. Accountability is dictated by the results of assessments conducted at each level. Across all levels, it is in the

classroom level that these two paradigms are greatly manifested. This is because at this level, teachers make use of assessment results to make instructional adjustments and interventions to improve student learning. At the end, teachers are held accountable for student achievement.

High quality classroom assessment, specifically on its formative assessment component, greatly accounts for improved student achievement (Black & William, 1998; Butler & Winne, 1995; Kingston & Nash, 2011; White & Frederiksen, 1998; Randel, Apthorp, Beesley, Clark, & Wang, 2016). When students are frequently assessed in the course of the learning process, teachers have the opportunity to adjust instruction and make necessary interventions to address learning deficiencies (Stiggins, 1998). This would result to better scaffolding of learners, hence improved student achievement. The concept that explains this is called formative assessment which is also a critical component of classroom assessment. Formative assessment is argued as a powerful way to improve student learning (Black & William, 1998; Magno & Lizada, 2015). Black and William (1998) has argued that an assessment strategy becomes formative when the results it yields are used to adjust instruction and make to meet students' needs. The P21 document on the 21<sup>st</sup> Century Skills Assessment (2007) has been strong on the need to balance the practice of formative assessment with the traditional use of assessment or summative. Assessments, it argues, must be regarded as both an instructional tool that is used to inform necessary adjustments in teaching to scaffold learners' current state of learning to their desired state of learning (formative), and as an accountability tool that shows quantitatively whether learning has actually occurred (summative) (P21, 2007).

One critical component of a high-quality classroom assessment is the alignment of assessment activities to learning goals (Randel et al., 2016; Ayala et al., 2008; Shepard, 2000; Stiggins & Chappuis, 2008; Valencia, 2008; Wiley, 2008). It is crucial, Randel et al. (2016) notes, that students are informed of and are made to understand these learning goals, how these classroom assessment activities are aligned with the goals, and the criteria by which progress in the achievement of goals is assessed. Studies (Fontana & Fernandez, 1994; Frederiksen & White, 1997) show that students who are aware of the learning goals learn more quickly than students who are not (Randel et al., 2016). Because the quality of assessment taking place at the classroom level greatly accounts for the quality of student achievement, teacher's accountability for assessment results is high (Magno & Piosang, 2016).

Because of the critical role classroom assessment plays in the improvement of student achievement, the literacy of teachers on classroom

assessment and on assessment in general is also critical. Assessment literacy, De Luca, McEwan, and Luhanga (2016) citing Popham (2013), has become a priority in educational systems in North America. Popham (2009) has argued that assessment is a necessary condition for competent teachers. In the United States, states have been conducting professional development programs for teachers that focus on classroom assessment. In South Carolina, for example, a professional development program on formative assessment was designed by its Department of Education which broadly addresses classroom assessment. The program specifically focuses on the increasing of teachers' skills in designing classroom assessments that improve student learning (Randel et al., 2016).

### **Standards for Teacher Competence in Educational Assessment of Students**

An effort to improve classroom assessment literacy of teachers necessitates the presence of a specific set of standards on educational assessment which teachers will have to observe. Having a set of specific assessment standards for teachers would set better practice of proper classroom assessment (Magno, 2013). However, Magno (2013) notes these set of specific assessment standards are only present in a few countries or states like the United States, New South Wales, and New Mexico. In the United States, the National Council on Measurement and Evaluation (NCME), American Federation of Teachers (AFT), and the National Education Association (NEA) jointly developed the Standards for Teacher Competence in Educational Assessment of Students or The Standards in 1990. There are seven specific standards under The Standard (AFT, NCME, & NEA, pp. 1-2):

1. Teachers should be skilled in choosing assessment methods appropriate
2. for instructional decisions.
1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3. The teacher should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.

5. Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

To date, there is no set of standards on teacher competence on student assessment yet in the Philippines (Magno, 2013). This may have an effect on the level of classroom assessment literacy of Filipino teachers as it has been argued that the presence of a set of specific assessment standards for teachers would set better practice of proper classroom assessment (Magno, 2013).

The Standards has served as test blueprint for the development of instruments that measure assessment and classroom assessment literacy such as the Teacher Assessment Literacy Questionnaire (TALQ) by Plake and Impara (1992), Assessment Literacy Inventory (ALI) by Campbell et al. (2002), and the Classroom Assessment Literacy Inventory (CALI) by Mertler (2004).

## Studies on Classroom Assessment Literacy

There have been several studies on both assessment literacy (general) and classroom assessment literacy of teacher (Plake, Impara, & Fager, 1993; Campbell et al., 2002; Mertler, 2003; Mertler & Campbell, 2005; Yamtim & Wongwanich, 2013).

Plake et al. (1993) conducted a national study measuring the assessment literacy of teachers in the United States. Using the TALQ, 553 teachers from a total of 98 districts in 45 states participated in the survey. Results of the study showed that teachers obtained an average score of 23 out of 35 items correct. A mean score of 3.45/5.00 was obtained on *Standard 3- Administering, Scoring, and Interpreting Assessment Results* which was the highest performance among all standards. The lowest performance occurred on *Standard 6- Communicating Assessment Results* with  $M=2.70/5.00$ . Plake et al. (1993) recommended inservice materials for teacher training which include standards of assessment (Magno, 2013).

While the study of Plake et al. (1993) focused on inservice teachers, the study of Campbell et al. (2002) focused on measuring the assessment literacy of undergraduate preservice teachers. This time, an instrument identical to TALQ was used – Assessment Literacy Inventory (ALI). It was administered to 220

preservice teachers taking up course on measurement. The data obtained from this study yielded higher reliability ( $\alpha=.74$ ) than in Plake et al (1993) study ( $\alpha=.74$ ). The average score of the preservice teachers was 21 which was two points lower than the inservice teachers in the study of Plake et al (1993). The participants in this study scored highest on *Standard 1- Choosing Appropriate Assessment Methods*. Both groups subjected in the studies of Plake et al (1993) and Campbell et al (2002) scored lowest on *Standard 6-Communicating Assessment Results*.

Mertler and Campbell (2005), in his study, compared the classroom assessment literacy of inservice and preservice teachers. It utilized the Classroom Assessment Literacy Inventory or CALI designed by Mertler (2004). The preservice teacher group (N=67) obtained an average score of 19 out of 35 items correctly. The highest performance of the preservice groups was found for *Standard 1*(M=3.25), while the lowest was obtained for *Standard 5-Developing Valid Grading Procedure* (M=2.06). For the inservice teachers (N=197), they obtained an average of 22 out 35 items correctly which is 3 points higher against the preservice group. The inservice group's highest performance was found on *Standard 3-Administering, Scoring, and Interpreting the Results of Assessments* (M=3.95). The group, on the other hand, obtained the lowest in *Standard 5-Developing Valid Grading Procedures*.

Lastly, the study of Yamtim and Wongwanich (2013) reported the classroom assessment literacy (CAL) of 19 primary school teachers in Thailand. The study used the Classroom Assessment Literacy Inventory (Mertler, 2003) to measure the classroom assessment literacy of the respondents. Findings of the study demonstrated that the mean scores of the participants (M=17.11) indicate poor assessment literacy with 78.95% of the participants who obtained scores that qualify under low level of CAL, and 21.04% who obtained scores that indicate medium level of CAL.

## **Classroom Assessment in the Philippines**

Classroom assessment literacy, defined as the necessary knowledge and skill in compiling data on students' achievement and the effective utilization of classroom assessment tasks and outcomes to better students' achievement (Chappuis, Stiggins, Chappuis, & Arter, 2012), has gained significant attention in educational systems around the world because of the critical role classroom assessment play in the improvement of student learning. Saefurroham and Balinas (2016), in their study on the classroom assessment practices of English teachers in the Philippines and Indonesia, has noted that in the two countries,

premium importance has been accorded to classroom assessment alongside with both countries' change of curriculum paradigm to one that places students as the center of learning. In the Philippines, the passage of Republic Act 10533 or the Enhanced Basic Education Act of 2013, also known as the K to 12 Law, has brought forth the issuance of a comprehensive policy on classroom assessment by the Philippine Department of Education (DepEd). In 2015, the DepEd has issued Order no. 08, s. 2015 titled Policy Guidelines on Classroom Assessment for the K to 12 Basic Education Program. DepEd Order no. 08 is strong on the role of classroom assessment, specifically classroom formative assessment, on the improvement of achievement of Filipino learners. While this order sets a comprehensive policy on the use of classroom assessment, there appears to be a dearth of studies that investigate the classroom assessment literacy of Filipino teachers. This is taken into serious consideration because several studies in other countries (Bol, Stephenson, O'Connell, & Nunnery, 1998; Stiggins & Conklin, 1992; Wiggins, 1989) have consistently demonstrated that teachers are inadequately trained and ill-prepared in developing, administering, and interpreting the results of various types of assessments including classroom assessment (Koh, 2011). This is aside from the observation that teachers were not good judges of the quality of the assessment tasks they develop (Black & William; Bol & Strage, 1996) is indicative of low assessment literacy (Koh, 2011). To pattern an assumption from these studies, if Filipino teachers have low classroom assessment literacy, two questions should be asked: (1) How then the mandate of DepEd Order no. 08 be effectively translated into actual classroom practice? (2) How does their level of CAL affect achievement of Filipino learners? Hence, there is pressing need for studies to be conducted focusing on the level of CAL of Filipino teachers.

### **CAL of English Teachers**

In response to this gap, the present study reports the levels of CAL of English teachers in the Philippines in two levels, secondary and tertiary. With the aim to produce holistically developed and globally competitive Filipino graduates, the achievement of 21<sup>st</sup> century skills have become an integral part of the K to 12 basic education reform program. One of the 21<sup>st</sup> century skills that make graduates globally competitive is communicative competence in English. This puts to critical task English language teachers. As it has been argued that high quality classroom assessment greatly accounts for improved student achievement (Black & Wiliam, 1998; Butler & Winne, 1995; Kingston

& Nash, 2011; White & Frederiksen, 1998; Randel, Apthorp, Beesley, Clark, & Wang, 2016), improving on the CAL of English teachers would therefore account for improved achievement of learners in English. Hence, there is a need to report the levels of CAL of the English teachers in the country as it may serve as foundation or basis for professional development programs specifically designed for the development of teachers' classroom assessment literacy (Koh, 2011). DeLuca and Klinger (2010) has noted that in-service teachers consistently indicate their need of more professional development programs on classroom assessment as teachers often receive inadequate training on classroom assessment (Randel et al., 2016; DeLuca & Klinger, 2010; O'Sullivan & Chalnik, 1991; Schaffer, 1993). The present study reports and compares the levels of CAL of English teachers in secondary and tertiary levels. This is also because previous studies like that of Mertler (2003) have focused on reporting and comparing the levels of CAL of preservice and inservice teachers. The present study, on the other hand, reports and compares the levels of CAL of two groups of inservice teachers teaching in two different levels.

## **Research Questions**

The present study aims to determine the levels of classroom assessment literacy of Filipino English language teachers in secondary and tertiary levels using Mertler's (2004) Classroom Assessment Literacy Inventory or CALI. The specific research questions are as follows:

1. What are the levels of classroom assessment literacy of secondary and tertiary English teachers?
2. Is there significant difference between the levels of classroom assessment literacy of secondary and tertiary English teachers?

## **Method**

### **Research Design**

The present study employed cross-sectional and descriptive research designs. Cross-sectional design is used to measure differences between or from among a variety of people, subjects, or phenomena (Bethlehem, 1999). By employing cross-sectional design, the present study was able to ascertain the differences between the levels of classroom assessment literacy (CAL) of secondary and tertiary English teachers both in overall CAL and in the specific standards of CAL. This study is also descriptive in nature as it determines,

describes, and identifies the levels of CAL of the both groups in overall CAL and in the specific standards of CAL.

## Participants

The participants of the study are 80 English teachers (Male= 16; Female=64) randomly selected from three (3) private and four (4) public secondary schools, and five (5) private and (2) public tertiary education institutions in National Capital Region, Philippines. Of the 80 participants, 40 are teaching in the secondary level and 40 are teaching in the secondary level, 47 are MA/MS degree holder. With respect to the length of service, 14 are 1-5 years in service; 9 are 6-10 years; 6 are 11-15 years; 38 are already 16-20 years; 6 are 21-25 years; and 7 are 26-30 years in service.

## Instrument

The present study utilized the Classroom Assessment Literacy Inventory or CALI developed by Mertler (2004). The CALI consists of two parts. The first part has a total of 35 items anchored on the seven (7) Standards for Teacher Competence in Educational Assessment of Students of AFT NCME, and NEA (1990). Each standard is allotted five questions in the inventory. Some of the items in the inventory are intended to measure general concepts related to testing and assessment. Some measure the use of assessment activities for assigning student grades and communicating the results of assessment. The other items measure the teacher's knowledge of standardized testing. The rest of the items measure teacher's knowledge of classroom assessment. The CALI was pilot tested two times. In the first pilot testing to 152 preservice teachers ( $n=152$ ), the items yielded relatively sufficient psychometric values with overall KR20 ( $r_{KR20}$ ) of .75, mean item difficulty was equal to .64, and the mean item discrimination was equivalent to .32 (Mertler, 2003). In the second phase of pilot testing, the items also yielded relatively sufficient psychometric values with overall KR20 ( $r_{KR20}$ ) of .74, mean item difficulty was equal to .68, and the mean item discrimination of .31 (Mertler, 2003).

In order to obtain a copy of the CALI, the researchers wrote an e-mail correspondence to Dr. Craig Mertler, Associate Professor at Arizona State University, who designed the CALI. He provided the researchers a copy of the CALI, hence authorizing them of its use for the conduct of the present study.

## Data Gathering Procedure

In order to obtain the necessary data, the researchers have first written letters to the school heads for the secondary schools, and to the program chair or deans of colleges for tertiary educational institutions to seek permission for administering the Classroom Assessment Literacy Inventory (CALI) to the English teachers in their respective schools.

The participants were first informed about the study being conducted. Then, informed consent was sought. They were also assured of confidentiality with regard to the reporting of scores. Then, the researcher explained the competencies or standards covered in CALI before the general instructions were read.

## Data Analysis

The data were analyzed using descriptive statistics in terms of frequency distribution, percentage (%), mean, and standard deviation. The levels of classroom assessment literacy were also computed. The scores of CAL could be divided into three levels, namely: (1) poor or needs improvement (lower than 60%), (2) fair (60-79%), and (3) good (80% and higher). To determine the raw number of participants that fall under whichever level, the following tables served as reference.

Table 1

*Interpretation for Levels of Overall Classroom Assessment Literacy*

Number	Level of Classroom Assessment Literacy	Interpretation
28 and higher	High	Good
21 to 27	Medium	Fair
20 and below	Low	Poor

Table 2

*Interpretation for Levels on each Standards of Classroom Assessment Literacy*

Number	Level of Classroom Assessment Literacy	Interpretation
4-5	High	Good
3	Medium	Fair
1-2	Low	Poor

For the overall CAL, the levels were computed by getting the 60%, 79%, and 80% of the total number of items which is 35. The levels of CAL on specific standards were computed by getting 60%, 79%, and 80% of the total number of items which is 5 (per standard).

T-test for independent samples were conducted for the overall classroom assessment literacy (CAL) of the secondary and tertiary English teachers to determine whether there is a significant difference between the levels of CAL of the two groups. The t-test for independent samples was also conducted for the levels of CAL of the two groups on each specific standard.

## Results

The results show that for the overall mean scores, secondary English teachers (SET) obtained higher mean score of 13.75 than tertiary English teachers (TET) with 10.87. Across all standards, SET obtained higher mean scores than TET except for *standard 6: communicating assessment outcomes*. The overall standard deviation of SET was lower than TET at SD= 5.93 and SD=6.57, respectively. Though SET obtained higher mean score than TET, both mean scores indicate poor classroom assessment literacy because the mean scores are lower than 20.

Table 1  
*Descriptive Statistics*

Group Standards	Secondary				Tertiary			
	M	SD	Min.	Max	M	SD	Min	Max
1. Choosing an assessment method	1.87	1.32	0	5	2.02	1.06	0	5
2. Developing assessment methods	2.15	1.21	0	5	1.92	0.87	1	4
3. Administering, assigning, and interpreting learning outcomes	2.15	1.49	0	5	1.46	1.21	0	5
4. Using assessment outcomes in decision making	2.42	1.29	1	5	1.87	0.86	1	4
5. Using assessment to determine levels of learning outcomes	1.92	1.49	0	5	1.12	1.38	0	5
6. Communicating assessment outcomes	1.00	1.28	0	5	0.84	1.50	0	5
7. Knowing unethical practices	2.25	1.53	0	5	1.61	1.06	0	4
Overall	13.78	5.93	8	27	10.87	6.57	5	31

For SET, the overall minimum score is 8 and the overall maximum score is 27. For TET, the overall minimum score was 5 and the overall maximum score was 31. It should also be noted that TET had higher maximum score (Max=31) but SET had higher minimum score (Min=8). Noticeably, SET yielded six minimum scores of 0 on 6 over 7 competencies which was comparably lower than five minimum scores of 0 obtained in the case of TET. While this was the case, SET yielded maximum scores of five (5) in all seven standards and TET yielded maximum scores of five (5) in only four (4) standards.

Table 2

*Levels of Classroom Assessment Literacy of Secondary English Teachers*

Standards	Classroom Assessment Literacy					
	Low		Medium		High	
	Number	Percent	Number	Percent	Number	Percent
1. Choosing an assessment method	28	70%	7	18%	5	12%
2. Developing assessment methods	28	70%	6	15%	6	15%
3. Administering, assigning, and interpreting learning outcomes	24	60%	8	20%	8	20%
4. Using assessment outcomes in decision making	24	60%	6	15%	10	25%
5. Using assessment to determine levels of learning outcomes	26	65%	9	23%	5	12%
6. Communicating assessment outcomes	36	90%	2	5%	2	5%
7. Knowing unethical practices	25	63%	4	10%	11	27%
Overall	34	85%	6	15%	0	0%

After the descriptive statistics were obtained, the levels of classroom assessment literacy (CAL) were computed for the overall CAL and for specific standards. For overall CAL, 34 (85%) of the secondary English teachers (SET)

have poor level of CAL while there were 6 (15%) who have medium level of CAL. The table also shows that none of the SETs were able to qualify as having good classroom assessment literacy (score  $\geq 28$ ).

With respect to the specific standards of CAL, *standard 6: communicating assessment outcomes* obtained the highest number of SETs which yielded low level of literacy with 36 (90%) SETs. There were 10 SETs which yielded high level of literacy on *Standard 4: using assessment outcomes in decision making*. Generally, most of the SETs demonstrated poor CAL across all standards with Standards

Table 3  
*Levels of Classroom Assessment Literacy of Tertiary English Teachers*

Standards		Classroom Assessment Literacy					
		Low		Medium		High	
		Number	Percent	Number	Percent	Number	Percent
1.	Choosing an assessment method	35	88%	-	-	5	12%
2.	Developing assessment methods	35	88%	1	2%	4	10%
3.	Administering, assigning, and interpreting learning outcomes	34	85%	2	5%	4	10%
4.	Using assessment outcomes in decision making	36	90%	-	-	4	10%
5.	Using assessment to determine levels of learning outcomes	35	88%	-	-	5	12%
6.	Communicating assessment outcomes	36	90%	1	2%	3	8%
7.	Knowing unethical practices	33	83%	2	5%	5	12%
Overall		35	88%	2	5%	3	7%

On the level of CAL of tertiary English teachers (TET), there were 35 (88%) TETs who have low level of overall CAL which was a point higher than SET. There were two (5%) TETs who yielded medium level of CAL. The

remaining three (3) TETs obtained high level of CAL which was higher when compared to SET which had none or 0.

When the levels of literacy on the specific standards were analyzed, there were more TETs who yielded poor literacy compared to SET. On *standard 4: using assessment outcomes in decision making* for example, there were 24 SETs who obtained low level of literacy which was comparably lower than 34 TETs who obtained low level of literacy on the same standard. Across all standards, TET had higher numbers of low level literacies (S1=35; S2=35; S3=34; S4=36; S5=35; S6=36; S7=33) than SET (S1=28; S2=28; S3=24; S4=24; S5=26; S6=36; S7=25).

Table 4

*Overall T-test for CAL of Secondary and Tertiary English Teachers*

G1 vs G2	t-value	Df	F-ratio	p variances	p-value
Secondary vs Tertiary	2.06	77	1.22	0.53	0.04*

\* $p < 0.05$

The t-test for independent samples was conducted in order to determine whether there is a significant difference between the CAL levels of secondary English teachers (SET) and tertiary English teachers (TET). The table shows that the variances of the two groups are equal with an F-ratio of 1.22. The p-variances of 0.53 is greater than .05. This signifies that the sample is fit for t-test.

The p-value obtained ( $p=0.04$ ) indicates significant difference between the level of CAL of SET and TER. By looking at the mean scores, SET has higher CAL ( $M=13.78$ ) than TET ( $M=10.78$ ). While the difference is significant at  $p < 0.05$ , both mean scores indicate poor level of CAL (score  $\leq 20$ ).

Table 5  
*T-test for Specific Standards*

Standards	t-value	Df	F-ratio	p variances	p-value
1. Choosing an assessment method	-0.55	77	1.55	0.18	0.57
2. Developing assessment methods	0.95	77	1.93	0.04	0.34
3. Administering, assigning, and interpreting learning outcomes	2.25	77	1.52	0.20	0.02*
4. Using assessment outcomes in decision making	2.22	77	2.26	0.13	0.02*
5. Using assessment to determine levels of learning outcomes	2.46	77	1.17	0.63	0.01*
6. Communicating assessment outcomes	0.49	77	1.36	0.33	0.62
7. Knowing unethical practices	0.00	76	1.00	1.00	1.00

\* $p < 0.05$

The same test was conducted to determine whether there is significant difference between the levels of literacy on each of the seven specific standards under CAL. Of the seven standards, *standard 3: administering, assigning, and interpreting learning outcomes*, *standard 4: using assessment outcomes in decision making*, and *standard 5: using assessment to determine levels of learning outcomes* yielded significant difference with p-values of 0.02, 0.02, and 0.01, respectively. By looking at the mean scores of the two groups, for standard 3, results show that SET has significantly higher literacy on administering, assigning, and interpreting learning outcomes ( $M=2.15$ ) than TET (1.46). Also, SET obtained significantly higher literacy ( $M= 2.42$ ) on the usage of assessment outcomes in decision making (*standard 5*) than TET ( $M=1.87$ ). The same is the case with the

usage of assessment to ascertain levels of learning outcomes where SET obtained higher mean score ( $M=1.92$ ) than TET ( $M=1.12$ ). While there exists a significant difference on the literacy of both groups on these specific standards, it must be noted that all of the mean scores still indicate poor level of literacy ( $\text{score} \leq 2$ ).

As for *standards 1, 2, 6, and 7*, the p-values obtained ( $p=0.75$ ;  $p=0.34$ ;  $p=0.62$ ;  $p=1.00$ ) indicate that there is no significant difference between the levels of literacy of SET and TET on these standards. By looking at the mean scores shown in Table 1, both groups yielded mean scores that indicate poor level of literacy on these standards ( $\text{score} \leq 2$ ).

## Discussion

The present study investigated the levels of classroom assessment literacy (CAL) of inservice English teachers in secondary and tertiary levels. Results showed that English teachers in the secondary has higher CAL ( $M=13.78$ ) than that of the teachers in the tertiary level ( $M=10.87$ ). Though, both mean scores indicate poor CAL of both groups. For the secondary teachers, 85% indicated poor CAL and 6% yielded fair CAL. Meanwhile, 88% of the tertiary teachers indicated poor CAL, 5% yielded fair CAL, and 7% good. Across all standards, the mean scores indicate poor literacy ( $M=1.87$ ;  $M=2.15$ ;  $M=2.15$ ;  $M=2.42$ ;  $M=1.92$ ;  $M=1.00$ ;  $M=2.25$ ). The findings of the present study indicate that the level of CAL of the participants is lower than the levels of CAL of the participants in the studies of Plake et al. (1993), Campbell et al. (2002), Mertler and Campbell (2005), and Yamtim and Wongwanich (2014).

The poor classroom assessment literacy (CAL) of English language teachers in secondary and tertiary levels is telling of the inadequacy of training and ill-preparedness of teachers in the rudiments of assessment, specifically classroom assessment. This is consistent with the findings of several studies that teachers receive insufficient training and ill-preparation in inadequately trained and ill-prepared in developing, administering, and interpreting the results of various types of assessments including classroom assessment (Bol, Stephenson, O'Connell, & Nunnery, 1998; Stiggins & Conklin, 1992; Wiggins, 1989). This is a crucial finding as when English teachers have poor CAL, hence are hardly able to make use of classroom assessments for the improvement of instruction and, consequently, of student achievement. Classroom assessment becomes inutile to the improvement of student achievement. Specifically, it may fail to account for the development of

necessary English language competencies indicated in the K to 12 curriculum. This may ultimately affect the development of the learners' communicative competence in English which is the core intended outcome spelled out in the Philippine K to 12 English Curriculum.

While the Philippine Department of Education has issued a comprehensive policy on the use of classroom assessment for the improvement of student learning and achievement as spelled out in DepEd Order no. 08, s. 2015 or the Policy Guidelines on Classroom Assessment for the K to 12 Basic Education Program, the findings of the present study indicate that teachers, especially in the secondary or basic education level, are not classroom assessment literate enough to translate the mandates and tasks of the said policy into practice. This is to say that no matter how comprehensive and elaborate the policy is on classroom assessment, it would serve inutile if teachers, similar to the findings of this study, lack the necessary knowledge and skills to successfully and appropriately make use of classroom assessment to improve student achievement.

### **Conclusion and Recommendation**

Because the results of the study revealed an alarmingly low classroom assessment literacy (CAL) of teachers of English in both secondary and tertiary levels, the necessity for professional development programs that focus on the development of the classroom assessment literacy of not only teachers in English but teachers across disciplines in basic and tertiary levels arises as something vital and pressing. The instrumentalities of the government that are in-charge of education such as the Department of Education and Commission on Higher Education, and the education sector as a whole should take a look at recalibrating and upgrading the policy on teacher training and professional development on classroom assessment. This is to enable teachers, most especially in the basic education level, to successfully translate the comprehensive policy on classroom assessment of the DepEd into practice which would account for the improvement of student learning and achievement.

Lastly, future researchers are encouraged to conduct the same study employing a bigger and a more representative sample size in order to determine on a wider scale the classroom assessment literacy of Filipino teachers in English across regions or provinces in the country. Such may serve as a sounder empirical basis for the development of a comprehensive

educational policy on professional development and teacher training programs focused on classroom assessment.

## References

- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: Author.
- Ayala, C. C., Shavelson, R. J., Ruiz-Primo, M. A., Brandon, P. R., Yin, Y., Furtak, E. M., & Tomita, M. K. (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Education, 21*, 315–334.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-71.
- Bethlehem, J. (1999). Cross-sectional Research. In H. Ader, & G. J. Mellenbergh (Eds.), *Research Methodology in the Social, Behavioural and Life Sciences* (pp. 110-143). London, England: SAGE.
- Bol, L., Stephenson, P. L., O'Connell, A. A., & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research, 91*, 323–330.
- Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education, 80*, 145– 163.
- Butler, D. L., & Winne, P.H. (1995). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology, 78*, 210-216.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Chappuis, J., Stiggins, R. Chappuis, S., & Arter, J. (2012). Classroom assessment for student learning: Doing it right-using it well. Assessment Training Institute.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*, 419–438.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and call for research. *Educational Measurement: Issues and Practices, 30*(4), 28-37.

- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276.
- Magno, C. (2013). Standards of teacher competence on student assessment in the Philippines. *The Assessment Handbook*, 10, 42-53.
- Magno, C., & Lizada, G. S. (2015). Features of classroom formative assessment. *Educational Measurement and Evaluation Review*, 6, 23-31.
- Magno, C., & Piosang, T. L. (2016). Assessment schemes in the senior high school in the Philippine basic education. *Educational Measurement and Evaluation Review*, 7, 66-87.
- Mertler, C. A. (2003, October). *Preservice versus inservice teacher' assessment literacy: Does classroom experience make a difference?* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Mertler, C. A., & Campbell, C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, QC, Canada.
- O'Sullivan, R. G., & Chalkin, M. K. (1991). Measurement-related course work requirements for teacher certification and recertification. *Educational Measurement: Issues and Practice*, 10, 17-23.
- Plake, B. S. (1993) Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12, 39.
- Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7<sup>th</sup> ed.). Boston, MA: Pearson.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory and Into Practice*, 48(1), 4-11.
- Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research*, 109(5), 491-502.
- Saefurrohman, R., & Balinas, E.S. (2016). English teachers' classroom assessment practices. *International Journal of Evaluation and Research in Education*, 5(1), 82-92.
- Schaffer, W. D. (1993). Assessment literacy for teachers. *Theory in Practice*, 32, 118-126.

- Shepard, L. A. (2000). The role of classroom assessment in teaching and learning (CSE Technical Report). Los Angeles, CA: National Center for Research and Evaluation, Standards, and Student Testing (CRESST).
- Siegel, M., & Wissehr, C. (2011). *Preparing for the plunge: preservice teachers' assessment literacy*. *Journal of Science Teacher Education*, 22(4), 371-391.
- Stiggins, R. J. (1998). *Classroom assessment for student success*. Student Assessment Series. Annapolis Junction, MD: NEA Professional Library
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.
- Stiggins, R. J., & Chappuis, J. (2008). Enhancing student learning. *District Administration*, 44, 42-44.
- Yamtim, V., & Wongwanich, S. (2014). A study on classroom assessment literacy of primary school teachers. *Procedia: Social and Behavioral Sciences*, 116, 2998-3004.
- Wiggins, G. (1989, May). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Wiley, C. R. H. (2008). Traditional teacher tests. In T. L. Good (Ed.), *21st century education: A reference handbook* (Vol. 1, pp. 431-442). Los Angeles, CA: SAGE
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3-118.



## Validation of the CEU-Lopez Critical Thinking Test Using Multidimensional IRT Model

**Johnny T. Amora**

*De La Salle-College of Saint Benilde, Philippines*

**Marcos Y. Lopez**

*Centro Escolar University-Malolos, Philippines*

### Abstract

The CEU-Lopez Critical Thinking Test was developed as a measure of the critical thinking dimensions (Lopez, 2012b). The present study adds to the emerging literature on critical thinking by assessing the psychometric properties of The CEU-Lopez Critical Thinking Test using the three IRT models, namely: unidimensional, consecutive, and multidimensional models. Emerged in the present study is a shorter version of The CEU-Lopez Critical Thinking test, which is fitted statistically to the multidimensional IRT model. The results from the analyses of the five dimensions of the short version of The CEU-Lopez Critical Thinking Test provided evidence on the reliability, validity, and measurement precision of the Critical Thinking Test.

*Keywords:* Critical thinking, Unidimensional IRT model, Consecutive IRT model, Multidimensional IRT model.

### Introduction

One among the twenty first century skills that are crucially essential to be cultivated by our learners is critical thinking which is not just necessary for the mastery of academic contents but more so for the execution of daily life-changing activities and work-related matters (Fisher, 2011; Halpern, 2014; Huber & Kuncel, 2015; Shaheen, 2016; Wagner, 2014). If critical thinking is a necessity in every aspect of human existence then educators need to do something on how this could be taught and tested effectively which supports

what Ennis (2003, 2011) stated that much work needs to be done not only on how critical thinking can be taught but also in the field of assessment, an area that has not been given due attention in critical thinking research. This purports that there is a dearth of existing critical thinking tests and this is true in Asian context which can be borne out by the list of critical thinking tests prepared by Ennis and Chattin (2015) which shows that majority of the available tests specifically written in English language are developed by Western scholars from North America, United Kingdom, Canada and only one from Asia specifically designed for tertiary learners in the Philippines. Hence, the need to develop more sophisticated critical thinking tests designed for Filipinos and other Asian nationals has to be addressed.

In response to the call for the need to develop a critical thinking test, Lopez (2012a, 2012b) developed and validated a critical thinking test designed for Filipino learners in tertiary level. It is called The CEU-Lopez Critical Thinking Test (2012a) which consists of 87 items and is considered a multi-aspect general knowledge critical thinking test (Lopez & Asilo, 2014). Ennis, Millman, and Tomko (2005) explained that the said type of test uses content that is based on daily life experiences which are assumed familiar to the target examinees. The said test developed by Lopez adopted the conception of Ennis (1996) as regards aspects of critical thinking for purposes of item construction. Considering the relevance of critical thinking for individuals to become more functional citizens of the world, the need to assess it periodically is undeniable. If there is no valid and reliable test of critical thinking available, there will be problems on how the schools and other learning organizations can monitor the progress of their efforts on enhancing critical thinking of learners.

The CEU-Lopez Critical Thinking Test, which is a multiple-choice type of test that consists of 87 items, deals with five aspects of critical thinking, such as: deduction, credibility judgment, assumption identification, induction, and meaning and fallacies with 19, 17, 16, 16, and 19 items, respectively. Though each aspect was tested separately, in practice, they are interdependent (Ennis, 1996, 2011; Facione & Gittens, 2013; Fisher, 2011; Norris & Ennis, 1989). The test focuses only on aspects of critical thinking that can be tested objectively. The disposition aspects of critical thinking, which are attitudinal in nature, cannot be tested objectively (Ennis, 2003; Facione & Gittens, 2013). Hence, they were not directly covered as part of test items.

In addition, the test has been normed locally using the students in the three campuses of Centro Escolar University (CEU), namely: Manila, Makati, and Malolos as samples. The results of such study yielded score interpretations that are categorized into six, namely, unreflective thinker, challenged thinker,

beginning thinker, practicing thinker, advanced thinker, and master thinker which are based on the concept of Paul and Elder (2001) as regards stages of critical thinking development of individuals. Three types of norms were generated such as university norms, curricular year level norms, and program type norms (science vs. non science). Though locally done, the said norms for score interpretation purposes can be adopted by other academic institutions that have similar academic setup with that of CEU. Other institutions that have entirely different academic structure with that of CEU can come up with their own norms for research and other purposes (Lopez, Mendoza, Lucero, Opina, 2014).

Concerning the internal consistency measure of reliability of the said test, KR 20 was used. The obtained reliability is .68 which is within the acceptable range for a critical thinking test considering that a critical thinking is a multidimensional construct. Reliability estimates that can be considered adequate tend to range from about .65 to .75 and may increase with the examinees' level of sophistication and that very high reliability on tests that deal with critical thinking should not be automatically regarded as better than more moderate ones (Norris & Ennis, 1989). Since The CEU-Lopez Critical Thinking Test is a multidimensional type of test, it can be expected that its internal-consistency index is lower than that of other unidimensional standardized tests (Lopez, 2012b).

In reference to its uses, The CEU-Lopez Critical Thinking Test can be utilized to determine the strengths and weaknesses of the students in certain aspects of critical thinking. The results of diagnosis can be used for the development of interventions or remedial measures regarding the enhancement of learners' critical thinking. Since the said test has been normed locally, it can be used to determine and categorize the level of critical thinking of the learners and can come up with a comparative study regarding the level of critical thinking of the learners who belong to different curricular year levels and different curricular areas like science or non-science-related courses. It can also be used as pretest and posttest in certain experimental study specifically in determining the effectiveness of infusing critical thinking into curriculum or to look into the comparison of effectiveness of the four approaches in the teaching of critical thinking, such as, general, infusion, immersion, and mixed. In addition, academic institutions may utilize the test as part of selective and retention process for courses in which board examinations are required as well as industries that value critical thinking as one of the most necessary and desirable attributes employees must possess in certain companies or some other organizations.

Furthermore, critical thinking experts from Association of Informal Logic and Critical Thinking (AILACT) were consulted regarding the content of the test and accuracy and clarity of items. Out of twenty-four recognized critical thinking consultants listed in AILACT website, three experts expressed willingness to review each item of the test. They were David Hitchcock (McMaster University, Hamilton, Canada), Dona Warren (University of Wisconsin-Stevens Point, Wisconsin, USA), and Susana Nuccetelli (St. Cloud University, Minnesota). Their insightful comments were looked into and considered thoughtfully as part of a number of series of revisions of test (Lopez & Asilo, 2014).

Regarding its construct validity, Norris' (1992) procedure in the use of verbal reports of thinking to determine the construct validity of each item was adapted. There were 13 verbal reports of thinking gathered for each test item. A total of 1,131 verbal reports of thinking were gathered with the additional 260 verbal reports of thinking for some items that called for further validation. The said verbal reports were tape-recorded and transcribed verbatim and carefully analyzed. Two types of scores were generated from this activity, such as, performance and thinking scores. The former is based on the examinees' selected answer from the three choices provided in each test item. The latter is based on the justification of the examinees as to why a certain option was their answer. The two scores were correlated to determine whether certain item needs to be retained, revised, or replaced in addition to the insights given by examinees based on their verbalized thoughts regarding test item (Lopez, 2013).

It is interesting to note that The CEU-Lopez Critical Thinking Test has been used by other Filipino researchers. Some of these studies were conducted by Agraan, Amado, Lumunsad, Manalo, and Vidad (2016), Galvez (2015), Gracia, Jose, Espiritu, Geronimo, Estrella, and Gulinao, (2013), Grafilo (2013), Moreno, Braza, De Villa, and Refugido (2016), Parico (2015), Tayao (2014), Reyes (2017), Tayao, Daez, Inigo, Cruz, Nievera, Francisco, and Avinante (2016), and Viray (2014) in which the said test was utilized as their measure in determining the critical thinking level of their students. Unfortunately, all of these researches did not report the reliability of the test for their gathered data. Though the reliability of the original 87-item of The CEU-Lopez Critical Thinking Test was analyzed (which yielded an overall KR-20 coefficient of .68), the reliability coefficient for each dimension was not reported. The present paper assesses the reliability of The CEU-Lopez Critical Thinking Test at the dimension level based on the IRT context.

The aforementioned researchers (i. e., those who utilized The CEU-Lopez Critical Thinking Test) were interviewed as regards their feedback in administering the test. They were one in saying that their examinees found the test too long and tiresome and might have the tendency not to take seriously the latter part of the test due to fatigue. Hence, it was deemed necessary to reduce the number of test items to make the test more manageable on the part of the examinees to finish answering the test without experiencing so much mental or physical exhaustion. The present paper attempts to develop a valid and reliable shorter version of The CEU-Lopez Critical Thinking Test based on the criteria of IRT.

It can be noted that the measurement properties of The CEU-Lopez Critical Thinking Test was established using the Classical Test Theory (CTT) approach. The use of CTT in test development has some limitations, which are mainly related to the validity and reliability of the results (Embretson & Reise, 2000; Singh, 2004). Some major limitations of CTT, which are described as circular dependency, are as follows: (1) Person's ability is test-dependent (i. e., for a fixed test, person's ability is high if the test is easy; and person's ability is low if the test is difficult); (2) item/test difficulty is group-dependent (i. e., item/test is easy if test takers have higher ability; and item/test is difficult if test takers have lower ability); and (3) Test-oriented (i. e., score is given at the test level, but there is no basis in determining how well a person perform on a particular item). According to Fan (1998), the circular dependency poses some theoretical difficulties in CTT's application in some measurement situations such as test equating and computerized adaptive testing. On the other hand, many researchers (e. g., Amora & Bernardo, 1999) stressed that it is difficult to judge whether a person with certain ability level will have a problem on a particular item because, in CTT, the items and persons are calibrated in different scales. All the weaknesses of The CEU-Lopez Critical Thinking Test that arise due to the limitations of the CTT can be fixed using the IRT and/or Rasch modeling approaches.

The present paper aims to analyze The CEU-Lopez Critical Thinking Test using the three IRT models, namely: multidimensional, unidimensional, and consecutive IRT models. The main goal is to find the best IRT model that fits The CEU-Lopez Critical Thinking Test. Then, based on the chosen best IRT model, a valid and reliable shorter version of The CEU-Lopez Critical Thinking Test is produced. The three IRT models are discussed in the next section.

## Method

### Participants

Served as respondents of the study were 1400 college students from the three campuses (CEU-Malolos, CEU-Makati, and CEU-Manila) of Centro Escolar University in the Philippines who enrolled in courses such as Medical and Health-related courses (e. g., Nursing, Medical Technology, Dentistry), Natural Sciences (e. g., Biology), Social Sciences (e. g., Psychology, Social Work), Education, Business-related courses (e. g., Accountancy, Business Administration), and Computer-related courses (e. g., Information Technology/Computer Science). Distribution of the samples taken was based on degree program, year level, gender, age, and SES (social economic status).

### The Instrument

The instrument under study is the 87-item multiple-choice type Critical Thinking Test (Lopez, 2012a and 2012b) which consists of five interdependent dimensions of critical thinking such as deduction, credibility judgment, assumption identification, induction, and meaning and fallacies. Deduction dimension, which consists of 19 items, refers to several principles of critical thinking such as fallacy of affirming the consequent, fallacy of division, fallacy of bandwagon, modus ponens, contraposition, and post hoc fallacy. In the test, the examinees are asked to decide on the given argument or item based on the following options: If the underlined statement follows necessarily from the other statements given, mark letter A. If the underlined statement contradicts the other statements given, mark letter B. If the underlined statement neither follows necessarily nor contradicts the other statements given, mark letter C. A sample item is “People say that Sex Education is not effective in teaching elementary pupils to be reflective and reasonable individuals, so, it should be eliminated from the list of subjects being studied by this young people. Although they may be right in saying it, Sex Education needs to be taught to elementary pupils.”

The credibility judgment dimension consists of 17 items. The criteria used in the item construction for judging credibility of sources and observation statements are expertise, lack of conflict of interest, agreement with other sources, reputation, careful habits, use of established procedures, ability to give reasons, minimal inferring involved, provision of records, and corroboration. In the test, each item has two characters who present two conflicting observation statements and examinees are requested to judge which of the two

contradicting statements is more credible by choosing the following options: If you think the first statement is more believable, mark letter A. If you think the second statement is more believable, mark letter B. If neither statement is more believable than the other, mark letter C. A sample item is “Ace said, “Oliver and I visited the Population Census Office in municipal hall two weeks ago and the chairman of the Population Census informed us that there is a total population of 2,558 living in the barrio Maasin. Searching for a sheet of paper, from his pocket, Oliver said, “After our conversation with the chairman of the Population Census, I immediately jotted down the shared information that there is a total population of 2,988 living in the barrio of Maasin.”

The assumption dimension consists of 16 items. The types of assumption tested in this dimension are presupposition, needed assumption, and used assumption. Each item has one character who makes a proposition that is taken for granted in a situation and that supports a conclusion. Sample item: “Farmers are hardworking individuals, so people from this place must be hardworking. Which is most probably taken for granted?” The response options for such item are as follows: A. Farmers in this place really work hard. B. Hardworking people are usually farmers. C. People from this place are farmers.

The fifth dimension, which is called meaning and fallacies, consists of 19 items. In this dimension, the items deal with difference on the use of necessary and sufficient condition language, judging provided definitions, negation and double negation, such logical words as only, if and only if, some, all, and some fallacies such as non-sequitur, post hoc, straw person, and circularity. There are three options to choose from after every stem of the item. Sample item: “To say that a person is healthy, is to say that a person can move gracefully, work tirelessly, and think rationally. These are all characteristics of a healthy individual. Of the following, which is the best statement that comes closest to the definition of a healthy person? A. To move gracefully, work tirelessly, and think rationally are necessary but not sufficient characteristics of a healthy person. B. To move gracefully, work tirelessly, and think rationally are not necessary but sufficient characteristics of a healthy person. C. To move gracefully, work tirelessly, and think rationally are each a necessary component, with all three being jointly sufficient characteristics of a healthy person.”

The CEU-Lopez Critical Thinking Test was developed and validated by adapting the eight-phase test development model designed by Norris (1992). Included in the eight phases are: (1) Test conceptualization, (2) Development of a test plan, (3) Development of the test items, (4) Face and

content validity of the test, (5) Revision of the test items, (6) Pre- try-out of the test, (7) Actual try-out of the test, and (8) construct validation of the test using verbal reports of thinking. The 87-item of The CEU-Lopez Critical Thinking Test yielded an overall KR-20 coefficient of .68. No reliability coefficients were reported for each dimension.

As discussed in the previous section, there are a number of studies that utilized The CEU-Lopez Critical Thinking Test. Unfortunately, all of these researches did not report the reliability of the test for their gathered data.

## Data Analysis

The gathered data were fitted to the three different IRT models, namely: unidimensional, consecutive unidimensional, and multidimensional models. The goal is to determine the model that fits best The CEU-Lopez Critical Thinking Test. The unidimensional model is the standard Rasch/IRT model where unidimensionality is the fundamental assumption; that is, all the items of the instrument should measure one single trait (Lord, 1980). In this paper, the unidimensionality model is tested because one goal of The CEU-Lopez Critical Thinking Test is to measure the overall critical thinking of the students.

The consecutive approach (Davey and Hirsch, 1991), on the other hand, is simply a unidimensional model repeated a number of times using subsets of the full range of items on a given instrument (Briggs and Wilson, 2003). The consecutive unidimensional can be an appropriate modeling approach to analyze The CEU-Lopez Critical Thinking Test because the goal also of the test is to measure each of the five dimensions of critical thinking.

Multidimensional IRT model, or simply multidimensional model, simultaneously calibrates several dimensions and capture the complexity of tests that measure several traits (Adams et al., 1997; Kelderman, 1996; Rost & Carstensen, 2002; Yao & Schwarz, 2006; Baghaei, 2012). Multidimensional IRT model, also known as multidimensional random coefficients multinomial logit (MRCML) model, is an extension of the Rasch family of item response models (Briggs & Wilson, 2003). There are two types of multidimensional model, namely: between-item multidimensional and within-item multidimensional models (Adams et al., 1997). In the between-item multidimensional model, each item in the instrument belongs to only one dimension, while in the within-item multidimensional model, each item is designed to measure simultaneously more than one dimension. The present paper utilized the between-item multidimensional model. For simplicity, multidimensional model

is used throughout this paper, instead of between-item multidimensional model. Multidimensional model is considered as another option to analyze The CEU-Lopez Critical Thinking Test because the test consists of five interrelated dimensions and the purpose of the test is to measure the five dimensions and then report them as separate scores of critical thinking performance and/or as a single critical thinking performance.

The parameters of the multidimensional model were estimated through the use of the ConQuest version 4.5.2 using the Monte Carlo method with 1000 nodes and .005 convergence. Monte Carlo method was used because The CEU-Lopez Critical Thinking Test consists of five interrelated dimensions. According to Adams and Wu (2010), the Monte Carlo method is generally the preferred approach for problems of more than three dimensions, so that the goodness-of-fit statistics are comparable across the three IRT models. The same estimation method (method=monte carlo, nodes=1000, convergence=.005) was utilized in the estimation of the parameters of the unidimensional and consecutive unidimensional models.

## Results and Discussion

The gathered data of the 87-item of The CEU Lopez Critical Thinking Test were subjected to analysis using the multidimensional model and subsequently using both the unidimensional and consecutive dimensional models. The multidimensional model was performed in two steps using the Conquest software: (1) All the 87 items of The CEU Lopez Critical Thinking Test were subjected to analysis and then the statistically fitted items were identified from the results. (2) All the statistically fitted items identified in step 1 were subjected again to analysis. Step was repeated by including only the fitted items in the analysis. Stop the analysis if all items in step 2 are already fitted statistically. Statistically fitted items are those with mean square (MNSQ) unweighted or weighted fit statistics that are inside the 95% confidence interval or items with MNSQ unweighted or weighted fit statistics with absolute T-values of less than 2.0. Results of the analysis revealed that 56 of the 87 items of The CEU-Lopez Critical Thinking Test statistically fitted the multidimensional model. In this study, the resulting 56-item instrument is called the short version of The CEU-Lopez Critical Thinking Test. Subsequently, the short version was subjected to further analysis using both the unidimensional and consecutive unidimensional models. It is worth noting that all the 56 items of the short version statistically fitted with both models. The distribution of the 56 items of the short version across the five dimensions

is shown in Table 2. Proceed to Table 5 for the item measures (or difficulty statistics) and item fit statistics of the short version.

Table 2

*Distribution of the fit and misfit items of The CEU-Lopez Critical Thinking Test across dimensions*

Dimension	Original version (# of items)	Short version (# of fitted items)	# of Misfitted items
1. Deduction	19	15	4
2. Credibility	17	8	9
3. Assumption	16	5	11
4. Induction	16	13	3
5. Meaning	19	15	4
<b>Total</b>	<b>87</b>	<b>56</b>	<b>31</b>

### Model fit Statistics

Model fit statistics such as Akaike Information Criterion (AIC) and Deviance ( $G^2$ ) of the multidimensional, unidimensional, and consecutive unidimensional models were computed. Generally, the smaller the AIC and  $G^2$ , the better the model fit. As shown in Table 3, the multidimensional model (AIC=87,633.99) has the smallest AIC among the three models, indicating that the multidimensional model fits better than the consecutive dimensional and unidimensional models. Because multidimensional model is hierarchically related to the unidimensional model, the model fit can be compared relative to the change in the deviance ( $G^2$ ) value, where the difference in deviance between the two models is approximately distributed as a chi-square (Briggs & Wilson, 2003) and the degrees of freedom is the difference in the number of parameters of the models. As indicated in Table 3, the difference in deviance is statistically significant ( $\chi^2=348.50$ ,  $df = 14$ ,  $p<0.00$ ), indicating that the multidimensional IRT model significantly fits the data better as compared to both unidimensional and consecutive unidimensional models. The multidimensional model is expected to be a better model since it accounts for the interrelated dimensions of The CEU-Lopez Critical Thinking Test. As shown in Table 4, the five dimensions under multidimensional model are significantly and positively correlated (ranging between .06 and .45), except for deduction and credibility ( $r = .01$ ,  $p>.05$ ). Having a positive correlation coefficient indicates that higher scores on one dimension tend to be paired

with higher scores on the other dimensions. Conversely, the lower the scores on one dimension so does the other dimensions. It is worth noting that of the five dimensions critical thinking, deduction and credibility emerged without significant correlation. This finding contradicts with the results of the previous studies (e. g., Ennis, Millman, & Tomko, 2005; Ennis, 1987, 1996) which connote that deduction and credibility judgments are correlated. The aforementioned authors argued that observation and credibility judgments call for the application of principles of deductive process. This suggests that when an individual is asked to give his judgment on the credibility of a person, deductive thinking may be applied. This may explain the correlation between deduction and credibility judgments. Moreover, there were no studies in the literature reporting that deduction and credibility are not correlated.

Table 3

*Fit statistics of the three models*

Models	AIC	Deviance (G <sup>2</sup> )	No. of Parameters
Unidimensional	87,954.41	87,840.49	57
Consecutive Unidimensional	87,754.96	n/a	61
Multidimensional	87,633.99	87,491.99	71

*Notes.* The deviance difference between multidimensional and unidimensional models is chi-square distributed with  $71-57 = 14$  degrees of freedom:  $\chi^2=348.50$ ,  $df = 14$ ,  $p<.000$ . Data are deviance and AIC of the short version of The CEU-Lopez Critical Thinking Test.

Table 4

*Correlations between the dimensions of The CEU-Lopez Critical Thinking Test short version*

	Deduction	Credibility	Assumption	Induction	Meaning
1. Deduction	1.00				
2. Credibility	.01	1.00			
3. Assumption	.06*	.35**	1.00		
4. Induction	.31**	.14*	.29**	1.00	
5. Meaning	.06*	.33**	.35**	.45**	1.00

*Note.* \*\* significant at .01 level; \* significant at .05 level.

One of the advantages of the present study in comparison with the early development of the scale by Lopez (2012a), which utilized the classical

test theory, is the model fit statistics. These model fit statistics play a very important role in the present study because IRT models a probabilistic model. Having a better model fit statistics implies that the model estimates (i.e., the results generated by the model, e.g., items measures) can be used in making inferences about the items of the scale and the scale as a whole with high degree of confidence, an important property of IRT model that cannot be provided by the classical test theory.

### Item Fit Statistics

Table 5 presents the item difficulty for each of the test items of the short version along with the corresponding fit statistics such as MNSQ fit statistics, the 95% confidence interval for the expected value of the MNSQ, and the T-statistics. According to Wu, Adams, Wilson, and Haldane (2007), MNSQ fit statistics are residual-based indices that are similar in conception and purpose to the weighted and unweighted fit statistics that were developed by Wright and Stone (1979) for Rasch's simple logistic model and Wright and Masters (1982) for partial credit model. An MNSQ fit statistic of 1.0 indicates that the item conforms perfectly with the multidimensional model. If a test item has a MNSQ statistic that lies outside the corresponding confidence interval, then the test item does not conform with the multidimensional model. Moreover, if the MNSQ statistic lies outside the confidence interval then the absolute value of the corresponding T-statistic exceeds 2.0 (Wu, Adams, Wilson, & Haldane, 2007). As shown in Table 5, the MNSQ fit statistics of the 56 items lie within the 95 confidence interval and the absolute values of the T-statistics are less than 2.0, indicating that all the 56 items of the short version of The CEU-Lopez Critical Thinking Test fit statistically with the multidimensional model.

The item measures for the test items are presented in the second column of Table 5. An asterisk next to a parameter estimate indicates that it was constrained. One test item for each dimension needs to be constrained so that the mean of the item measures on each dimension is zero (Adams and Wu, 2010). Theoretically, the values of the item measures range between negative infinity and positive infinity. Practically, the values of the item measure range between -3.0 and +3.0.

The item-person map (Figure 1.0) depicts the plot of the persons (represented by Xs) and test items (represented by item numbers) with the five dimensions of the shorter version of The CEU-Lopez Critical Thinking Test. In IRT, both items and persons are calibrated on a common continuum based

on the amount of traits possessed by each other (Bond & Fox, 2007). In the map of the present study, the values in the continuum approximately range from -2.0 to +2.0. The items are hierarchically arranged in terms of items measures from very easy (bottom) to very difficult (top), while the least able respondents are placed at the bottom and the most able at the top. It can be noted that the bunch of test items (around more than 57%) have difficulty measures that fall within the moderate levels. Only few items are very easy (5%), easy (10%), difficult (23%), and very difficult (4%). The ability levels of the majority of the respondents match with the below average test items.

The quality of the items of The CEU-Lopez Critical Thinking Scale was assessed using the item fit statistics. Items that did not qualify for the item fit statistics criteria were removed from the IRT model. Such quality criteria in selecting the good items for the shorter version of the scale is one of the advantages of the IRT modeling in comparison with the classical test theory utilized in the early development of the scale. The 56-item shorter version of The CEU-Lopez Critical Thinking Scale is an improvement of the 87-item original version which was developed based on the classical test theory. Unlike the longer version, the measures in the shorter version are item-free and person-free.

In terms of item difficulty, the 87-item original version have difficulty levels that range from easy to very difficult with a bunch of items (about 77%) falling within difficulty to very difficult levels, the 56-item short version have more dispersed difficulty levels ranging between very easy to very difficult with only few (27%) difficult and very difficult items. This means that a lot of difficult and very difficult items in the long version were pulled down to the moderately difficult items in the short version. This finding is not surprising since the IRT and CTT are two different methodologies.

Table 5

*Item measure and item fit statistics of The CEU-Lopez Critical Thinking Test short version*

Item#	Item Measure	MNSQ	95%CI	T-Value	Item#	Item Measure	MNSQ	95% CI	T-Value
<b>Q01</b>	-2.29	1.00	(.93,1.07)	0.10	<b>Q53</b>	-0.81	1.04	(.96,1.04)	1.80
<b>Q03</b>	0.05	1.00	(.95,1.05)	-0.10	<b>Q54</b>	0.74	1.00	(.87,1.13)	0.00
<b>Q04</b>	0.97	1.01	(.90,1.10)	0.20	<b>Q56</b>	-1.54	1.02	(.98,1.02)	1.80
<b>Q05</b>	1.11	1.00	(.89,1.11)	0.00	<b>Q57</b>	-0.16	0.99	(.93,1.07)	-0.20
<b>Q06</b>	-1.78	1.03	(.95,1.05)	1.20	<b>Q58</b>	0.83	1.00	(.87,1.13)	0.00
<b>Q08</b>	-0.11	1.00	(.96,1.04)	-0.20	<b>Q59</b>	-0.89	1.00	(.97,1.03)	-0.20
<b>Q09</b>	-1.96	1.00	(.94,1.06)	0.10	<b>Q60</b>	-0.30	1.03	(.94,1.06)	0.90
<b>Q10</b>	0.73	1.00	(.91,1.09)	0.10	<b>Q61</b>	0.97	1.00	(.85,1.15)	0.00
<b>Q11</b>	1.95	1.02	(.81,1.19)	0.20	<b>Q62</b>	0.72	1.00	(.87,1.13)	0.00
<b>Q12</b>	0.21	0.99	(.94,1.06)	-0.40	<b>Q65</b>	-0.53	0.98	(.95,1.05)	-1.00
<b>Q13</b>	0.32	1.02	(.94,1.06)	0.70	<b>Q66</b>	0.12	1.00	(.92,1.08)	0.00
<b>Q16</b>	-1.51	1.04	(.96,1.04)	2.30	<b>Q67</b>	0.36	1.00	(.90,1.10)	-0.10
<b>Q17</b>	1.83	1.00	(.82,1.18)	0.00	<b>Q68</b>	0.483*	1.01	(.89,1.11)	0.10
<b>Q18</b>	0.51	0.97	(.93,1.07)	-0.80	<b>Q70</b>	-0.05	0.98	(.96,1.04)	-0.90
<b>Q19</b>	-0.009*	0.98	(.95,1.05)	-0.70	<b>Q71</b>	-0.56	0.99	(.98,1.02)	-1.10
<b>Q20</b>	-0.26	1.00	(.97,1.03)	0.00	<b>Q73</b>	0.61	1.01	(.92,1.08)	0.40
<b>Q21</b>	-0.07	1.01	(.96,1.04)	0.40	<b>Q75</b>	1.10	0.99	(.89,1.11)	-0.10
<b>Q22</b>	0.27	1.02	(.95,1.05)	0.80	<b>Q77</b>	0.13	0.99	(.95,1.05)	-0.30
<b>Q23</b>	0.10	0.97	(.96,1.04)	-1.40	<b>Q78</b>	-0.50	0.99	(.98,1.02)	-0.50
<b>Q24</b>	-0.11	0.98	(.96,1.04)	-1.30	<b>Q79</b>	0.17	1.03	(.95,1.05)	1.00
<b>Q29</b>	-0.32	1.02	(.97,1.03)	1.10	<b>Q80</b>	0.21	1.01	(.95,1.05)	0.20
<b>Q30</b>	0.38	1.00	(.95,1.05)	0.10	<b>Q81</b>	0.18	1.02	(.95,1.05)	0.80
<b>Q35</b>	0.006*	1.00	(.96,1.04)	0.20	<b>Q82</b>	-0.97	1.01	(.98,1.02)	0.80
<b>Q37</b>	-0.15	1.01	(.96,1.04)	0.50	<b>Q83</b>	0.02	1.01	(.95,1.05)	0.20
<b>Q39</b>	0.00	1.00	(.96,1.04)	0.10	<b>Q84</b>	0.42	1.01	(.93,1.07)	0.20
<b>Q40</b>	0.02	1.02	(.96,1.04)	0.90	<b>Q85</b>	-0.11	1.00	(.96,1.04)	-0.10
<b>Q44</b>	0.02	0.99	(.96,1.04)	-0.50	<b>Q86</b>	-0.19	1.00	(.96,1.04)	-0.20
<b>Q46</b>	0.113*	0.99	(.95,1.05)	-0.40	<b>Q87</b>	-0.467*	1.01	(.98,1.02)	0.80

Note: 95%CI = 95% Confidence Interval; MNSQ = Mean square fit statistic; the 95%CI, MNSQ, and T-value are based on the weighted fit statistics.

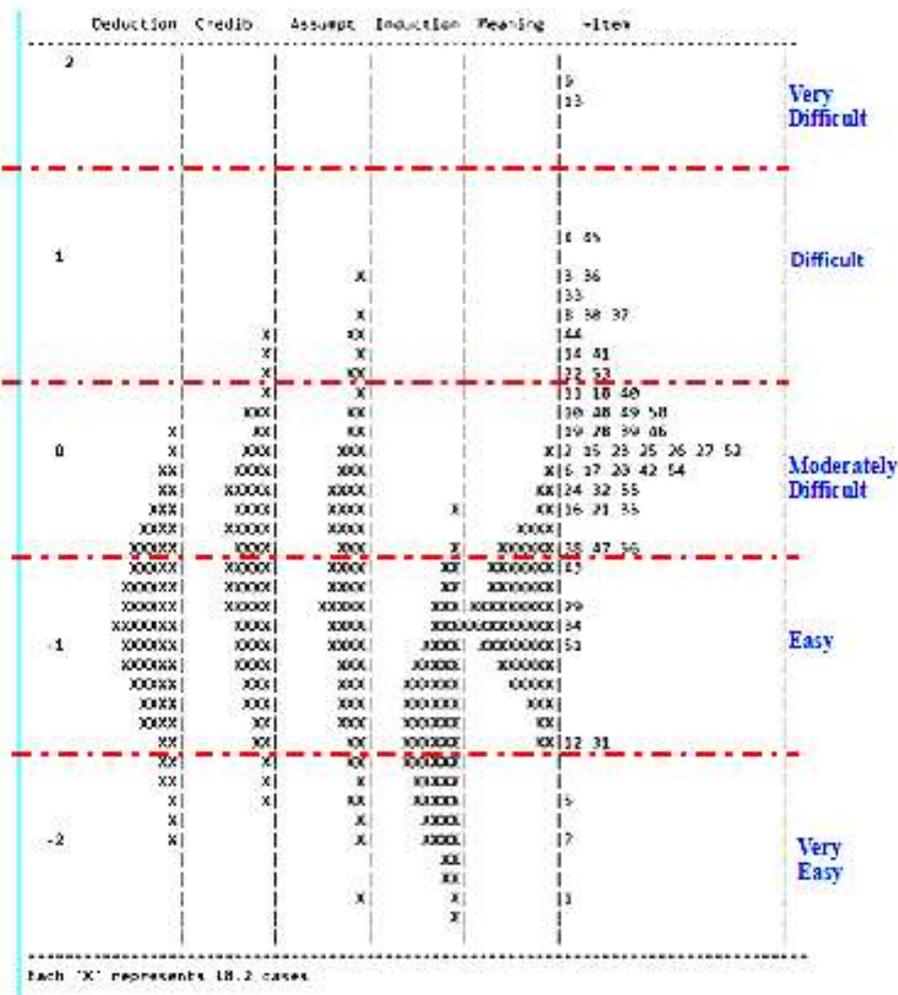


Figure 10. Person Item Map

## Reliability of the The CEU-Lopez Critical Thinking Test

As discussed in the previous section, the model that fits best the CEU-Lopez Critical Thinking Test is the multidimensional IRT model. As shown in Table 6, the model that yields the large reliability coefficients for each of the five dimensions is the multidimensional IRT model. The reliability coefficients vary from the lowest of .60 for credibility dimension to the highest of .69 for the induction dimension. Such reliability coefficients are not far from the .68

overall reliability coefficient of the original version of The CEU-Lopez Critical Thinking Test.

The findings of the present study indicate that each of the five dimensions of The CEU-Lopez Critical Thinking Test is reliable. The scores yielded by The CEU-Lopez Critical Thinking Test are definitely not reliable if both conservative and unidimensional IRT models are utilized. The multidimensional IRT model is expected to yield the more reliable coefficients as compared to the conservative and unidimensional models because the former is the appropriate model for the five interdependent dimensions of the scale. The appropriateness of the multidimensional model for the scale under investigation is supported by the model fit statistics as discussed in the previous section. On the other hand, result of the present study in the aspect of reliability coefficient per dimension is an additional information about the psychometric properties of the scale since its early development reported only the reliability coefficient of the overall scale.

Table 6

*Reliability of The CEU-Lopez Critical Thinking Test*

Dimension	#of Items	Multidimensional IRT	Consecutive IRT
Deduction	15	.66	.34
Credibility	8	.60	.38
Assumption	5	.61	.37
Induction	13	.69	.32
Meaning	15	.64	.25
Total	56		

Unidimensional = .46; KR-20 (original version) = .68

## Conclusion

The present study provides additional evidence on the reliability and validity of The CEU-Lopez Critical Thinking Test as a measure of the five critical thinking dimensions within the context of item response theory. The CEU-Lopez critical thinking test was successfully reduced from its original version with 87 items to the shorter version with 56 items. The five-

interrelated dimensions of the shorter version of The CEU-Lopez critical thinking test can be best modelled using the IRT multidimensional model rather than the IRT unidimensional and consecutive models. Results of the analysis using the multidimensional IRT model show that the shorter version of the test has sound psychometric properties with item measures ranging between easy and difficult, where more than 55% of the items have average difficulty level. Further study about the shorter version may be conducted at the item level (e. g., item bias). On the basis of the findings of the present study, researchers are encouraged to use the shorter version of The CEU-Lopez Critical Thinking Test as a measure of the critical thinking dimensions.

### References

- Adams, R. and Wu, M. (2010). *Multidimensional models. Chapter 10 of the notes and tutorial*. ACER ConQuest Version 4 [computer software]. Camberwell: Australian Council for Educational Research.
- Adams, R. J., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23.
- Agraan, S. M., Amado, A. K., Lumunsad, K. R., Manalo, G. A., & Vidad, M. L. (2016). *An assessment of critical thinking skills among radiologic technology students of De La Salle Health Sciences Institute* (Unpublished undergraduate thesis). De La Salle University-Dasmarinas, Cavite, Philippines.
- Amora, J. T., & Bernardo, A. S. (2009). Testing and reducing L2 vocabulary learning strategies inventory using Rasch model. *Philippine ESL Journal*, 3, 38-73
- Baghaei, P. (2012). The application of multidimensional Rasch models in large scale assessment and validation: An empirical example. *Electronic Journal of Research in Educational Psychology*, 10, 233-252.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd ed.). NJ: Lawrence Erlbaum.
- Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, 4(1), 87-100.
- Davey, T., & Hirsch, T. M. (1991). Concurrent and consecutive estimates of examinee ability profiles. Paper presented at the annual meeting of the Psychometric Society, New Brunswick, NJ

- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists*. Mahwah, NJ: Erlbaum.
- Ennis, R. H. (2011). Critical thinking. Reflection and perspective Part II. *INQUIRY: Critical Thinking Across the Disciplines*, 26(2), 5-19. doi:10.5840/inquiryctnews201126214
- Ennis, R. H. (2003). Critical thinking assessment. In D. Fasko Jr (Ed.), *Critical thinking and reasoning* (pp. 293-313). NJ: Hampton Press, Inc.
- Ennis, R. H. (1996). *Critical thinking*. Upper Saddle River, NJ: Prentice Hall.
- Ennis, R. H., & Chattin, G. S. (2015, May 3). An annotated list of English-language critical thinking tests [Updated list of available critical thinking tests written in English]. Retrieved from [www.criticalthinking.net/TestListDraft050315.docx](http://www.criticalthinking.net/TestListDraft050315.docx)
- Facione, P., & Gittens, A. G. (2013). *Think Critically*. USA: Pearson Education, Inc.
- Fisher, A. (2011). *Critical thinking: An introduction*. UK: Cambridge University Press.
- Galvez, V. B. (2015, April). *The relationship between the level of critical thinking skills of B.S. Psychology students of Centro Escolar University-Malolos and their academic performance in professional Psychology subjects*. Paper presented at the annual research presentation of faculty of Centro Escolar University, Malolos, Philippines.
- Garcia, M. A., Jose, S. J., Espiritu, K., Geronimo, A., Estrella, C., & Gulinao, A. (2013). *The relationship of critical thinking skills and Mathematics performance of freshmen and sophomore Bachelor in Secondary Education major in Mathematics students of polytechnic University of the Philippines-Sta. Maria, Bulacan campus*. (Unpublished undergraduate thesis). Polytechnic University of the Philippines-Sta. Maria, Bulacan, Philippines.
- Grafilo, E.B. (2013). *Critical thinking and reading competence of senior high school students*. (Unpublished master's thesis). Bulacan State University, Malolos, Philippines.
- Halpern, D. F. (2014). *Thought & knowledge: An introduction to critical thinking*. NY: Psychology Press.
- Huber, C. R., & Kuncel, N. R. (2015). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, 20(10), 1-38. doi: 10.3102/0034654315605917
- Kelderman, H. (1996). Loglinear multidimensional item response theory models for polytomously scored items. In R. K. Hambleton, & W. J. van der Linden (Eds.), *Handbook of Modern Item Response Theory* (pp. 287-304). New York: Springer Verlag.

- Lopez, M. Y. (2013). Determining the construct validity of a critical thinking test. *Educational Measurement and Evaluation Review*, 4, 87-99.
- Lopez, M. Y. (2012a). *The CEU-Lopez Critical Thinking Test*. Research and Evaluation Office, Centro Escolar University, Manila.
- Lopez, M. Y. (2012b). *The CEU-Lopez Critical Thinking Test Manual*. Research and Evaluation Office, Centro Escolar University, Manila.
- Lopez, M. Y., & Asilo, M. V. (2014). Development and validation of The CEU-Lopez Critical Thinking Test. *Inquiry: Critical Thinking Across the Disciplines*, 29(1), 32-55. doi: 10.5840/inquiryct20142914
- Lopez, M. Y., Mendoza, E. R., Lucero, R. D., & Opina, A. S. (2014). Establishing the local norms of The CEU-Lopez critical thinking test. *E-Journal of European Academic Research*, 2(3), 3895-3928.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Moreno, R., Braza, M., De Villa, R. D., & Refugido, K. M. (2016). *Intervening factors among grade 10 students' level of critical thinking skills in Mathematics and Science at Manuel S. Enverga University Foundation-Candelaria, Inc.* (Unpublished undergraduate thesis). Manuel S. Enverga University Foundation-Candelaria, Inc., Quezon Province, Philippines.
- Norris, S. P. (1992). A demonstration of the use of verbal reports of thinking in multiple-choice critical thinking test design. *Alberta Journal of Educational Research*, 38, 155-176.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- Paul, R., & Elder, L. (2001). *Critical thinking: Tools for taking charge of your learning and your life*. Upper Saddle River, NJ: Prentice Hall.
- Reyes, E. C. (2017). *Critical thinking and academic performance of junior high school students: Basis for the proposed instructional model in Chemistry* (Unpublished doctoral dissertation). Centro Escolar University, Mendiola, Manila.
- Rost, J., & Carstensen, C.H. (2002). Multidimensional Rasch Measurement via Item Component Models and Faceted Designs. *Applied Psychological Measurement*, 26(1), 42-56
- Shaheen, N. (2016). International students' critical thinking-related problem areas: UK university teachers' perspectives. *Journal of Research in International Education*, 15(1), 18-31. doi: 10.1177/1475240916635895
- Singh, J., 2004. Tackling measurement problems with Item Response Theory: Principles, characteristics, and assessment, with an illustrative example. *Journal of Business Research* 57 (2), 184-208.

- Tayao, J. C. (2014). An assessment on the level of critical thinking skills of BSIT students at Centro Escolar University-Malolos in relation to their Computer Programming and Mathematics academic performance. *E-Journal of European Academic Research*, 2(3), 4394-4411.
- Tayao, J. C., Daez, F. D. P., Inigo, A. N., Cruz, F. D., Nievera, M. L. T., Francisco, N. V., & Avinante, M. D. P. (2016, March). *The relationship of critical thinking skills and decision-making skills of selected Higher Education Institutions' senior Management Accounting, Finance, and Accounting Technology students in the province of Bulacan, Philippines*. Paper presented at International Conference on Business and Social Sciences (ICBASS2016) at Kyoto Research Park, Kyoto, Japan.
- Wagner, T. (2014). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need-and what we can do about it*. New York: Basic Books, A Member of Perseus Books Group.
- Wright, B. D., & Masters, G. N. 1982. *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press
- Wright, B. D., & Stone, M. 1979. *Best Test Design: Rasch Measurement*. Chicago: MESA Press.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S.A. (2007). ACER ConQuest Version 2: Generalised item response modelling software [computer program]. Camberwell: Australian Council for Educational Research.
- Yao, L., & Schwarz, R.D. (2006). A Multidimensional Partial Credit Model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30, 469-492.