

DEVELOPED STATUS AND GENDER EFFECTS ON COMPUTERIZED TOEFL
ACCEPTANCE

Kubra Karakaya Ozyer

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Masters of Art in the Department of Educational Psychology, Measurement and Evaluation in the School of Education.

Chapel Hill
2014

Approved by:

Gregory J. Cizek

William B. Ware

Jeffrey A. Greene

© 2014
Kubra Karakaya Ozyer
ALL RIGHTS RESERVED

ABSTRACT

Kubra Karakaya Ozyer: Developed status and gender effects on computerized ToEFL acceptance
(Under the direction of Gregory J. Cizek)

The primary purpose of this study was to address gender and country developed status differences in computerized ToEFL acceptance. Based on previous studies, this study developed and tested a model, which included four latent variables: (1) perceived playfulness; (2) perceived usefulness; (3) perceived ease of use, and (4) behavioral intention to use. The study sample consisted of 237 international students, ranging in age from 18 to 44 years. Multigroup Structural Equation Modeling and Multiple Indicator Multiple Cause methods were used to test differences between subgroups for computerized ToEFL acceptance. This study did not find evidence of gender differences on computerized ToEFL acceptance. Moreover, participants' native country's developed status did not have an impact on the behavioral intention to use computerized ToEFL exam for international students. The implications for future research were discussed.

ACKNOWLEDGEMENTS

I am grateful to many people for helping me through this difficult and lengthy journey. Special thanks to Dr. Gregory Cizek for letting me continue this research under his guidance and for pushing me to do my best.

I also offer thanks to Dr. William Ware and Dr. Jeffrey Greene for their willingness to remain on my committee, despite the extended time this has taken me to complete.

To my husband, thank you for giving up a wintry home in Turkey near friends and family for the heat of North Carolina and the stress of life with a graduate student. Words cannot express how much the love and support have meant to me.

Finally, I am so happy that I have a supportive family around me. They are always next to me whenever I need. The strengths of this dissertation belong to them as much as to me. The weaknesses belong to me alone.

TABLE OF CONTENT

LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1: INTRODUCTION.....	1
Introduction.....	1
Research Questions.....	4
CHAPTER 2: LITERATURE REVIEW	5
Assessment Technologies.....	5
Social Validity and CBA	7
The ToEFL as a High-stakes Test.....	13
Theory of Acceptance.....	15
Theory of reasoned action.....	16
Limitations and proposed additions to the TRA model.....	17
Technology acceptance model.....	18
Criticism of the TAM framework.....	19
Theory of planned behavior.....	20
Extended technology acceptance model.....	21
Combined technology acceptance model and theory of planned behavior.....	24
Social cognitive theory.....	26
The model of PC utilization.....	26
Innovation diffusion theory.....	28

Unified theory of acceptance and use of technology.	29
Computer-based assessment acceptance model.	31
Individual Differences in Acceptance.....	36
Country development differences and acceptance.	36
Gender differences and acceptance.	39
Conclusions.....	41
CHAPTER 3: METHODOLOGY	43
Research Design	43
Variables	44
Perceived playfulness.....	46
Perceived usefulness	47
Perceived ease of use.....	48
Computer self-efficacy.....	50
Social influence.	51
Facilitating conditions.	53
Goal expectancy	54
Content of examination	55
Behavioral intentions to use..	57
Methodological Approach	60
Research Participants	60
Instrumentation	62

Validity and Reliability of Instrument	63
Ethical and Security Concerns	65
Procedures.....	66
Data Analysis	67
Developed status analysis.....	72
Gender effect analysis.	73
CHAPTER 4: RESULTS	76
Data Preparation	76
Descriptive Statistics.....	76
Confirmatory Factor Analysis	79
CFA model for full data	79
Modified full CFA model.....	81
CFA for new model.....	83
Modified CFA for new model.....	84
Multigroup Structural Equation Modeling Analysis	84
New structural equation model..	85
Test of configural model for developed status	85
Test of MIMIC model for developed status	85
Test of configural model for gender.....	88
Test of weak factorial invariance for gender.....	89
Test of MIMIC model for gender.....	90

Summary.....	92
CHAPTER 5: DISCUSSION.....	94
Research Purpose and Framework.....	94
Research Design	95
Discussion of Findings.....	96
Implications	99
Limitations	100
Future Directions	101
Summary.....	102
APPENDIX: COMPUTERIZED ToEFL ACCEPTANCE QUESTIONNAIRE.....	104
REFERENCES	117

LIST OF TABLES

Table

1. Hypotheses for Gender Differences	58
2. Hypotheses for Developed Status Differences	60
3. Sample Characteristics: Age	78
4. Sample Characteristics: Gender	78
5. Sample Characteristics: Developed Status.....	79
6. Sample Characteristics: Educational Level.....	79
7. Sample Characteristics: Discipline	80
8. Results of the Model Fit Tests for Full Dataset	83
9. Results of the Model Fit Tests for Developed Status.....	87
10. MIMIC Model Estimates for Developed Status	88
11. Results of the Model Fit Tests for Gender	90
12. Results of X^2 Difference Test	91
13. MIMIC Model Estimates for Gender	92
14. Summary of Hypothesis Testing Results	99

LIST OF FIGURES

Figure 1 - Theory of reasoned action model.....	17
Figure 2 - Technology acceptance model	18
Figure 3 - Theory of planned behavior model.	20
Figure 4 - Extended technology acceptance model.	23
Figure 5 - Combined technology acceptance model and theory of planned behavior.....	25
Figure 6 - The model of PC utilization model.	28
Figure 7 - Unified theory of acceptance and use of technology model	31
Figure 8 - Computer-based assessment acceptance model.....	35
Figure 9 - Computerized ToEFL acceptance model.....	45
Figure 10 - Revised computerized ToEFL acceptance model	83
Figure 11 - MIMIC model diagram for developed status difference.....	88
Figure 12 - MIMIC model diagram for gender difference	92

CHAPTER 1: INTRODUCTION

Introduction

One who speaks only one language is one person, but one who speaks two languages is two people (Turkish proverb).

In a rapidly changing communication environment, knowing another language, especially English, is crucial for interacting with the rest of the world. English is accepted by many people as a global language that enables individuals to share information easily on economic, health, and political matters. As a result, in many countries people are taught English as a second language at an early age. For international students pursuing education at American universities, English proficiency is often measured by norm-referenced tests to inform admission and placement procedures. The *Test of English as a Foreign Language* (ToEFL) is a widely accepted instrument used to measure reading, listening, speaking, and writing skills (Anderson, 2009). Because the ToEFL is a widely accepted and used test, it has been taken by over 27 million people in the world.

International students who take either the paper-based or computer-based ToEFL examination come from “developed” nations, as well as countries that are recognized as “developing,” based on criteria established by the World Bank (World Bank, 2014). Developed or developing countries are divided based on their Gross National Income (GNI) per capita per year. GNI is converted to U.S. dollars using the World Bank Atlas method and divided by the midyear population. The World Bank described GNI as “the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net

receipts of primary income (compensation of employees and property income) from abroad” (World Bank, 2014). The World Bank named countries with a GNI more than US\$11,905 as developed and countries with a GNI of US\$ 11,905 or less as developing (World Bank, 2014). Each year, students taking the ToEFL come from both developed and developing nations. Since universities require ToEFL scores for admission and placement decisions, the demand for both testing opportunities and high scores on examinations is great.

Advances in technology have provided opportunities to expand education and measurement methods in order to meet this demand. One application that is frequently used is computer-based assessment (CBA) (Pino-Silva, 2008). Computer-based assessment encompasses a range of activities such as scoring, administration, and analyses of student assessment processes. Researchers (Bugbee, 1996; Drasgow, 1999; Parshall, Spray, Kalohn, & Davey, 2002) have emphasized the advantages of using this technology in the assessment process. Some identified advantages of CBA include increased objectivity and consistency of scoring, as well as extending the range of available assessment methods. Some researchers have argued that computerized assessments are preferred more than paper-based assessments by students (Croft, Danson, Dawson, & Ward, 2001; Sambell, Sambell, & Sexton, 1999). Students have reported that computer-based assessments are more realistic, objective, interesting, fun to use, fast, easy, and less stressful (Croft, Danson, Dawson, & Ward, 2001; Sambell, Sambell, & Sexton, 1999). In addition, there is some empirical evidence that students who use computer-based assessments achieve better results than those taking paper-based tests (Bocij & Greasley, 1999; Noyes, Garland, & Robbins, 2004). There is some debate within the research literature as to what factors influence whether a student selects CBA or traditional paper-based tests as well as what can affect performance outcomes on each.

Schneberger, Amoroso, and Durfee (2008) conducted a study to identify factors that influenced performance on computer-based assessments. They developed a revised version of the technology acceptance model (TAM). The TAM is a conceptual framework that includes perceived usefulness and perceived ease of use as predictors of behavioral intention to take CBAs. Their research study showed that students who have a higher level of expertise with computers perform better on computer-based assessments. Moreover, the level of expertise factor was correlated with perceived ease of use of CBA. In addition to expertise, attitude toward using computer-based assessment was a factor that affected assessment performance. These research findings indicate that acceptance of CBA can affect test performance (Schneberger, Amoroso, & Durfee, 2008).

In contrast to these findings, the Educational Testing Service (ETS) published a study addressing the transition from paper-based assessments (PBA) to CBA for ToEFL examinations (Breland, Lee, Najarian, & Muraki, 2004). The study examined whether there was a difference between computer-based and paper-based ToEFL acceptance across different subgroups. The results did not show a performance difference between two groups. Later, ETS published a subsequent report that examined the factor structure of the internet-based ToEFL test (IBT) across subgroups comprised of people from different nationalities (Sawaki, Strinker, & Oranje, 2009).

To make better informed decisions and to meet the challenge of developing computer-based testing systems that satisfy stakeholder needs, test developers need to better understand the factors that influence the acceptance and use of CBAs. To enhance this understanding, this study posited a holistic framework to examine several constructs suggested in the literature that may lead to the behavioral intention to use a computerized ToEFL exam. For the purpose of this

study, *behavioral intention to use* refers to the examinees' perceptions, beliefs, and intentions to take a test, in this case a future form of the ToEFL in a computer-delivered administration.

The aim of this study was to gain an understanding of factors that affect students' intention to use computer-based assessments in general and specifically the computerized version of the ToEFL. To perform this task, this study used a computer-based assessment acceptance model (CBAAM) (Terzis & Economides, 2011) and adapted it to investigate the effects of gender and students' native country developed status on computerized ToEFL acceptance

Research Questions

Two research questions (RQ) were explored in this study. They were:

RQ1: Are there differences between students from developed and developing countries on their behavioral intention to use, perceived usefulness, perceived ease of use, and perceived playfulness that affect acceptance of computerized ToEFL?

RQ2: Are there differences between female and male international students' behavioral intention to use the computerized ToEFL?

CHAPTER 2: LITERATURE REVIEW

The purpose of this study was to investigate whether international students' behavioral intention to use the computerized ToEFL can be predicted either by the developed status of students' native country or gender. The following literature review first examines the relevant research of assessment technology, acceptance theories, and behavioral intention to use models. Second an explanation of the computer based assessment acceptance model is provided as well as a justification for its use in this study. Third, the studies examining the role of country developed status and gender difference for computerized ToEFL acceptance are discussed

Assessment Technologies

Assessment is a critical factor in student learning and there is considerable pressure on higher education institutions to measure learning outcomes more formally both frequent assessment procedures and high-stakes testing (Farrer, 2002; Laurillard, 2002). According to Shohamy (2001a), a "high-stakes test" is a test the results of which affect important decisions in the life of the test takers. For example, the score of the ToEFL test is used to inform decisions about whether or not the applicants will be accepted into higher education, allowed to continue in future studies, and enrolled in specific programs such as medical degree.

Current technological advances offer exciting opportunities to design assessments that are active and flexible. Moreover, measuring complex student knowledge and providing rich observations for student learning are two crucial advantages of new assessment technologies.

Since the 1970s, there have been advances in information and communication technology with assessment technology. In the early 1970s, clinical psychologists and the U.S. military

pioneered the development of computer-based tests (Russell, Goldberg, & O'Connor, 2003). Psychologists realized that computerized assessments could be used as a control method. Moreover, psychologists could optimize the use of trained personnel by freeing them from the routine and time-consuming functions of test administration and scoring (Russel et al., 2003).

The decade of the 1990s witnessed the rapid expansion of computer-based assessment (CBA). Although CBAs were administered for many years, the availability of faster, more powerful, and cheaper computers made large-scale computer-delivery of tests feasible (Mills, Potenza, Fremer, & Ward, 2005). In 2001, the National Research Council (NRC) published a report, *Knowing What Students Know*. This report highlighted innovative projects that used technology to assess learning. This report foreshadowed how advances in technology and statistical analysis would provide new models for assessment (Russel et al., 2003). However, it was not until 2006 that state, national, and international high-stakes testing programs begin to deliver high-stakes tests via computer-based technology. For instance, in 2006 the Programme for International Student Assessment (PISA) piloted online versions of its items prior to moving into online delivery. In 2011, the National Assessment of Educational Progress (NAEP) piloted technology-based items in math, science, and literacy in the U.S. The results of the computer-based testing initiatives that began in the 1970s and continued through subsequent decades were significant and vast. Presently, ETS administers ToEFL, GRE, SAT, SMAT examinations via computer-based or internet-based systems worldwide. In addition to the U.S., nations such as Singapore and Norway are beginning to consider ways in which computers might be used to enhance student assessment of knowledge. The success and impact of computer based assessment programs has not yet been evaluated. Nevertheless, it is likely that other nations will

also begin to transition their assessment process to a computer-based format in order to improve the objectivity, consistency, and validity of the testing instruments.

For computer-based assessment, there are a range of activities such as scoring, administration, and analyses of student assessment process. Prior research findings indicate several advantages to using computer-based technology versus traditional paper-based testing. Two identified benefits include increased scoring objectivity and extending the range of assessment methods (Bugbee, 1996; Drasgow, 1999; Parshall, Spray, Kalohn, & Davey, 2002). According to Ricketts and Wilks (2002), CBA improved students' test performance if they were provided an appropriate and clear computer interface. In other words, screen arrangement should be simple and understandable for all types of students. The researchers reported that if students did not have to scroll down the page during testing, they were more likely to obtain higher scores (Ricketts & Wilks, 2002). Moreover, Noyes et al. (2004) conducted a research study comparing performances of undergraduate students on paper-based and computer-based assessments. The undergraduate students selected either multiple choice paper-based or computer-based tests. The results indicated that students who preferred computer-based tests achieved better scores than those taking the paper-based version. However, traditional concerns about paper-based testing such as test bias, content validity, familiarity of tests, and psychological effects of tests are also relevant to CBAs. These concerns also include issues such as whether prior access to technology affects test performance for students of different gender or whose native countries are of different developed status.

Social Validity and CBA

Messick (1981) argued that the social dimension was one of the crucial aspects test validity. Social influence factors are determined by how testing or test results may affect a test

taker's life. It has been argued that researchers should consider both the usefulness and appropriateness of test scores for the students (Broadfoot, 2005). McNamara (2001) also argued that a fundamental component of test validity research was to understand the social value of the test. Therefore, studying the impact of a test on the lives of students has become increasingly important (Broadfoot & Black, 2004). According to He and Shi (2008), the social influences of a test emphasize the importance of including students' perceptions of standardized English writing tests. These reactions could be used as evidence for the construct validity of test inferences. Also, Bachman (2000) claimed that analyzing test usage and test bias was of critical importance for language testers and the individuals affected by its use.

Pino-Silva (2006) reported that students' have positive thoughts about CBA and believe it has more advantages than disadvantages. This perception was supported by a research study conducted by Escudier, Newton, Cox, Reynolds, & Odell and published in 2011. This study examined the advantages of computer-based assessments over traditional paper-based assessments. The results indicated that undergraduate students perceive computerized assessment to be both fair and acceptable (Escudier, Newton, Cox, Reynolds, & Odell, 2011).

However some studies have indicated concerns that CBA did not measure some of the knowledge and abilities that they were supposed to measure (He & Shi, 2008; Tsaia & Tsou, 2009). In particular, a test that was designed for test takers with different language, cultural, and educational backgrounds may be inaccurate on what it reportedly measures. Subjects or topics that were deemed appropriate by test designers may not be that understandable for the test takers from different backgrounds than those that developed the assessment.

Some studies on the students' perceptions reported perceived problems in the test content. For example, a study conducted by Tsaia and Tsou (2009) investigated students'

perceptions of a paper-based English Language Proficiency (ELP) test. The findings indicated that most of the students perceived the ELP test as boring and too broad in scope. In addition, the students reported that the material covered in the test was different from what they learned in their foreign language classes. They believed that the test did not yield scores that accurately reflected their English language abilities. This study highlighted important problems of paper-based English language proficiency tests. It clearly showed that students did not perceive standardized paper-based language tests as enjoyable, interesting, or motivating.

There are indicators that some standardized tests do not measure what they were supposed to measure. To address this issue, organizations and administrators should investigate examinee's perceptions. For example, examining the content of computerized ToEFL exam from examinee's perspective would be extremely important if valid inferences are to be made based on the test results.

Another benefit to investigating students' perceptions about tests is to gather valuable data on how the tests affect them. Studies conducted by Triplet and Barksdale (2005) and Klinger and Luce-Kapler (2007) reported on the emotional effects of tests on students. In both studies, students were asked about their perceptions of taking high-stakes tests. The results of both studies indicated that students have negative feelings about the tests in general and specifically the length and difficulty of the test, and the possible consequences of failure. Moreover, students were angry to be assessed by a single test rather than throughout the school year (Triplet & Barksdale, 2005). Negative feelings were also reported by the students in Klinger and Luce-Kapler's (2007) study. Their research study showed that students developed test-anxiety when they take the high-stakes tests and prepare those tests. The students even claimed that the class preparation also frustrated them. This indicates that high-stakes testing has negative

effects for many people, especially students, and that they feel that they have to spend a lot of time preparing for the test. Preparation time, testing length and difficulty, and the perceived consequences of failure are elements of social validity that should be considered when developing or using high-stakes tests.

With CBA well-established, researchers have been interested in (1) whether or not test takers' apprehension about computer use affects their language performance on proficiency examinations; (2) whether negative effects could be documented; and (3) how those effects could be minimized. These questions lost importance as examinees in the U.S. improved their feelings of computer self-efficacy. However, little is known about how computer anxiety affects test takers using CBA with less experience or feelings of self-efficacy with regard to technology. The possibility of variance in test performance would be in part attributable to variance in computer familiarity. To address this concern many test developers (e.g., ETS) provide support for test takers who may not yet feel at ease with the technology. One example is the tutorial for the computer-based ToEFL (Taylor, Kirsch, Jamieson, & Eignor, 1999). The aim of the tutorial is to give all examinees an opportunity to familiarize themselves with editing and system functioning.

In summary, tests, especially high-stakes tests, could cause anxiety for the students and influence students' learning if students become most interested in learning devoted to preparing for the test (Klinger & Luce-Kapler, 2007; Triplet & Barksdale, 2005). In addition, teachers and school administrators were also influenced by the test, they felt the need to devote time to test preparation, and encouraged the examinees to take the test (Klinger & Luce-Kapler, 2007).

Another factor that was reported to affect achievement on tests was the student's familiarity with the subject of test. Familiarity with genres and topics on a writing test had

positive effects on students' confidence in the test. In other words, when students encountered unfamiliar topics and genres, they demonstrated less confidence and pleasure with the test (Moni, Kraayenoord, & Baker, 2002).

Technology based assessment researchers have also examined the equivalence of scores from paper-based and computer-based test version. In most of the research, the focus was how computer familiarity impacted the performance of examinees taking a computer-based test. Research reported that some examinees that were less familiar with computers perform worse when the test format was computer-based. In one study, students who were not familiar with typing on computers got lower grades on the computer-based tests (Russell, 1999). The familiarity with computers affected their computer self-efficacy which was a component of computer-based testing acceptance. Another study focused on the Massachusetts Comprehensive Assessment System's (MCAS) Language Arts Tests. It indicated when the researchers could alter the administration environment for the writing section, examinees' test results dramatically increased (Russell & Plati, 2002). In other words, facilitating conditions and computer system familiarity impacted the test results. The authors recommended that high-stakes language test programs should give both paper-based and computer-based options to students. It is evident that students' achievement on a test may not accurately reflect their true ability because of the many factors that influence performance such as familiarity with the test topics and the testing.

In addition to social and emotional effects, student's attitude toward the test is also an important factor in students' testing experience. Attitude is defined as a learned tendency to respond to an object in a favorable or unfavorable way (Fishbein, 1967; Fishbein & Ajzen, 1975). A study carried out by Brown and Hirschfeld (2008) indicated that students with positive attitude toward tests received higher scores than those with negative attitudes. Brown and

Hirschfeld (2008) suggested that if tests were presented as measures of students' individual learning rather than a mechanism of school or teacher's accountability, then students' performance on a test could increase. Fishbein and Ajzen (1975) stated that behavioral intentions are influenced by attitudes and subjective norms. Results of previous studies of students' behavioral intention to use computer-based tests were inconsistent (Fischer & Kopp, 2006; Fluck, Pullen, & Harper, 2009). Although students were familiar with computer and internet usage (Karsten & Schmidt, 2008; Kennedy, Gray, & Tse, 2008; Kennedy, Judd, Churchward, Gray, & Krause, 2008; Link & Marz, 2006), there were several concerns regarding CBA. Students reported feelings of apprehension over testing security (Cassady & Gridley, 2005), possibility of cheating (King, Guyette, & Piotrowski, 2009; Ozden, Erturk, & Sanli, 2004) and the typing element of the testing process (i.e., responding to a task requires quick input of free text answers by keyboard) (Cassady & Gridley, 2005; Fluck et al., 2009). For the successful implementation of computer-based assessment performance, it is helpful to understand these factors may influence student's acceptance.

In conclusion, the studies of students' perceptions of tests provided empirical-based information on the tests and their impacts on students as test takers. These studies showed that tests could be problematic if their mode of administration prevents them from accurately measuring knowledge or skills. First, testing could negatively influence students' learning. Second, they may test students' attitudes toward the test instead of their knowledge and skills. Finally, students' performances were influenced by their familiarity with the test environment and topics. The literature presented thus far is of students' perception of tests in general. The next section focuses on the ToEFL examination as a high-stakes test.

The ToEFL as a High-stakes Test

ToEFL scores are used widely to make important decisions about second language speakers of English. Therefore, this test can be categorized as a high-stakes test. The original ToEFL was administered in paper-based form and was divided into three sections: reading and listening comprehension, and structure and written expression (Anderson, 2009). From 1963 to 2012, the ToEFL evolved from a paper-based test (ToEFL PBT) to a computer-based test (ToEFL CBT) in 1998, and an internet-based test (ToEFL IBT) in 2006. The computer-based version included only reading, listening, and writing sections, while the new internet-based ToEFL has an additional section for speaking.

The computer-based version of ToEFL has the advantages of more security, complexity, and visuals with real life graphs (Ginther, 2001). In addition, because of faster scoring, it allows for a faster admissions process. This is an improvement for test takers and universities. In 2001, ETS stated that in most countries the ToEFL CBT replaced the paper-based version. However, the paper-based ToEFL was still administered on certain days in many countries where infrastructure issues such as unreliable electrical service and internet connectivity limited the ability to offer the ToEFL CBT test. After 2006, ETS discontinued administration of the computer-based ToEFL and announced that they would discontinue the ToEFL PBT after 2012. Prior to 2013, in some locations ETS offered paper-based and internet-based ToEFL options to examinees. However, in most places, there was only one available option. For example, in Turkey, examinees were required to take the ToEFL IBT version. ETS has reported that 96% of ToEFL test takers worldwide took the ToEFL IBT test in 2011, but there is no information regarding whether ToEFL IBT takers had an opportunity to choose a ToEFL type before the

other formats were discontinued. As of 2013 the ToEFL IBT is the only option available to students.

High-stakes tests have power over what knowledge is considered valuable. Tests can dictate which information is important to learn and which information is valued less. In the long term, only selected information will be learned based on what is assessed. Shohamy (2001a) argued that “the power of tests has reached such high levels that are now common belief that what is being tested is important” (p. 109). She further stated that only selected knowledge become important if tests determine the knowledge which is important and which knowledge is not. For instance, in ToEFL IBT reading sections, the texts are mostly expository, argumentation, and historical narrative (Alderson, 2009). Because these text genres are in the ToEFL test, the teachers and students will only be interested in these types of genres, and other text genres may be considered less important. The tasks in the ToEFL IBT reading test require test takers to answer multiple choice questions and to make inferences and summaries from the reading texts (Alderson, 2009). Because these skills were tested, there was a possibility that students thought that the best way to interact with the reading test was to make inferences and summaries, whereas other ways to interact with reading texts might be less emphasized or even neglected. For example, argumentation or poetic writing skills are two important skills but they can be ignored by standardized testing system. There was a tendency that “simply the fact that something is being tested creates a belief that it had reached status and importance” (Shohamy, 2001a, p.113). Because entities develop and administer high-stakes, these tests may affect the determination of what knowledge is important. Thus, the test developer (ETS) is the control mechanism, and this often makes the test takers change their behavior to align with the agendas of the tests in order to maximize their scores (Shohamy, 2001a).

The ToEFL test is intended to measure the examinees' ability to understand English as it is spoken in North America (Jang & Roussos, 2007; Sawaki et al., 2009). The ToEFL test developer has the power to decide what components of North American English are important for the test takers to know and understand and measure these constructs via computers. Therefore, someone wants to take the ToEFL test, regardless of his or her cultural, demographic and educational background, whether or not he or she is familiar with North American English and computer usage, she or he has to learn a certain type of English and computer-based assessment. It shows how "tests are used to redefine knowledge, change the test takers' behavior according to the set agendas of those in power and impose the values and knowledge of those in authority" (Shohamy, 2004, p. 73).

The obligation to use ToEFL IBT limits the research of students' reactions and perceptions. These reactions may affect their perceptions of themselves and of the test users. Acceptance by test takers, test users, and the public is essential to the continued viability of the ToEFL. Because there was ambiguity about how and why people prefer the ToEFL IBT, I will explore how examinees' perceptions of computerized testing relate to their intention to use a ToEFL IBT. The results of previous studies indicated that examinees' had positive feelings about computerized tests, and they preferred such kind of tests in the future. However, more research is needed to understand acceptance, or, why and how students intended to use the computerized ToEFL.

Theory of Acceptance

User technology acceptance has been examined extensively in prior information system and learning management research. Most studies examined behavioral intention in the analysis of accepting a particular technology. After the 1980s, various models of technology acceptance

were developed and tested. There are nine principle models in the information technology acceptance literature. Each model attempted to describe the determinants of acceptance of the proposed technology. Those models are theory of reasoned action (TRA), technology acceptance model (TAM), theory of planned behavior (TPB), combined TAM and TBP (C-TAM-TPB), extended technology acceptance model (TAM2), social cognitive theory (SCT), the model of PC utilization (MPCU), the innovation diffusion theory (IDT), and unified theory of acceptance and use of technology (UTAUT).

Theory of reasoned action. Acceptance models were developed from several base theories, but Theory of Reasoned Action (TRA) is considered one of the most important models (Fishbein & Ajzen, 1975). According to TRA, *behavioral intention* (BI) predicted the performance of behaviors that were under a person's volitional control (see Figure 1). Intention was modeled as a function of *attitude towards behavior* and *subjective norm*. According to the theory, external variables influenced behavior only indirectly by affecting attitude or subjective norm.

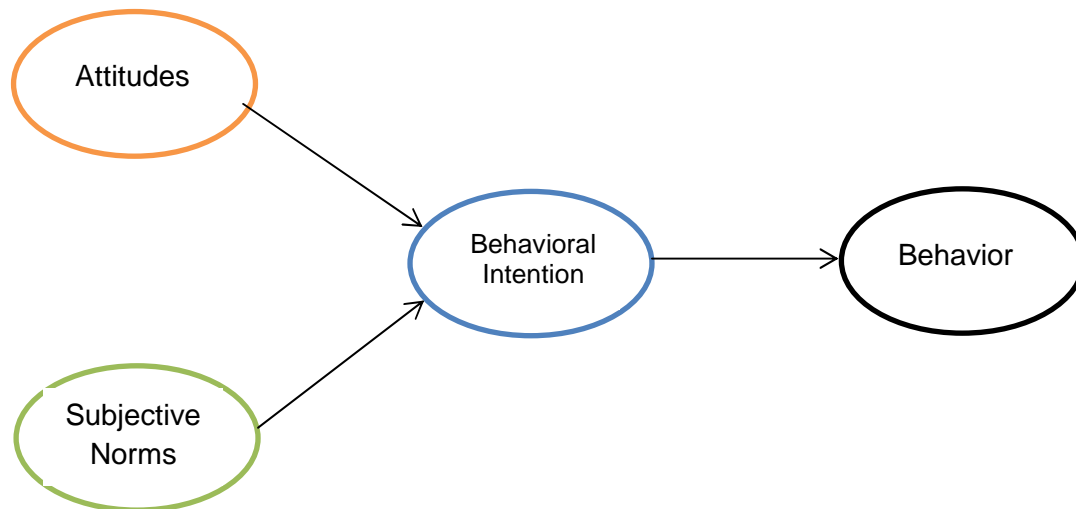


Figure 1. Theory of reasoned action model. Adapted from “Prediction of Goal-directed Behavior: Attitudes, Intentions, and Perceived Behavioral Control,” by I. Ajzen and T. M. Madden, 1986, *Journal of Experimental Social Psychology*, 22(5), p. 454. Copyright 1986 by Elsevier.

Limitations and proposed additions to the TRA model. Some researchers identified limitations of the TRA. Venkatesh, Morris, Davis, and Davis (2003) identified three limitations with the implementation of the TRA. First, the authors noted the lack of a clear differentiation between attitude and subjective norms. This could create confusion when applying these constructs. Second, TRA explores only attitudes and subjective norms. It does not account for other external variables, such as gender, culture, individual differences and unconscious habits. Thus, there might be behaviors that were not explained by TRA. Third, the observations made were based on the self-report of individuals rather than direct observation. The fourth limitation of TRA concerns the perceptions of individuals. TRA does not have the capacity to clearly explain certain types of behaviors. In order to compensate for these limitations, Ajzen developed a new model referred to the theory of planned behavior (TPB) in 1991 and that model is going to be described later in this chapter.

Technology acceptance model. Davis and colleagues developed the technology acceptance model, an adaptation of the theory of reasoned action (TRA) (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989). Davis et al. (1989) found TAM to be a better predictor of intention to use technology. The TAM predicts that user acceptance of technology is determined by three factors: *perceived usefulness* (PU), *perceived ease of use* (PEOU), and *behavioral intention* (BI) (see Figure 2).

Their research showed that behavioral intention was mediated by PU and PEOU (Davis et al., 1989). In addition, behavioral intentions were a strong predictor of actual use (Davis et al. 1989; Taylor & Todd, 1995).

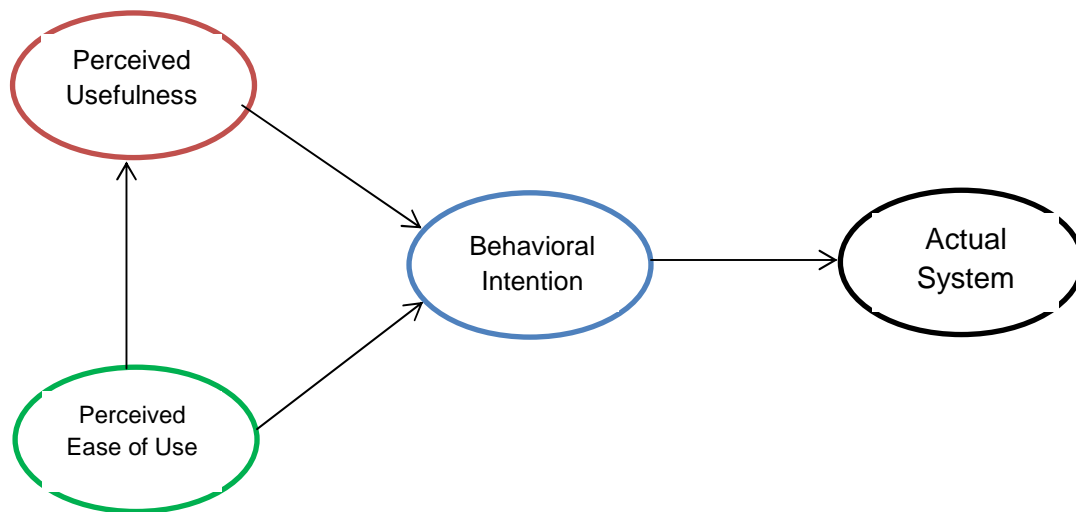


Figure 2. Technology acceptance model. Adapted from “User Acceptance of Computer Technology: A Comparison of Two Theoretical Models,” by F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, 1989, *Management Science*, 35(8), p. 985. Copyright 1989 by the INFORMS.

In the technology literature, the TAM instrument has been used to examine the problems of user behavioral intention (Dishaw & Strong, 1999; Lederer, Maupin, Sena, & Zhuang, 2000; Moon & Kim, 2001). Lately, educational research used TAM to investigate different technology based issues in educational settings such as student acceptance of online courses and gender

differences in preservice teachers, (Drennan, Kennedy, & Pisarski, 2005; Gao, 2005; Kelleher & O'Malley, 2006; Ma, Anderson, & Streith, 2005; Ngai, Poon, & Chan, 2007; Ong & Lai, 2006; Pan, Sivo, Gunter, & Cornell, 2005; Pituch & Lee, 2006; Selim, 2003; Yuen & Ma, 2002).

Davis (1989, 1993) has suggested that further studies need to be performed to extend TAM to determine the types of external variables such as computer self-efficacy and training that may influence the motivating belief factors of *perceived usefulness* and *perceived ease of use*.

Criticism of the TAM framework. In a modern society driven by technological developments, many people are affected by organizations mandating the use of information technology (Frank, 2011; Mitra, Sambamurthy, & Westerman, 2011). A direct result of this enforcement of technology usage has been the investigation by information technology researchers to examine and analyze its effect on TAM. Although many research lent some support the TAM as an acceptable model, numerous studies (Bagozzi, 2007; Burton-Jones & Straub, 2006; Karahanna & Straub, 1999) also identified concerns or inconsistencies.

Venkatesh and Morris (2000) argued that social aspects of acceptance were not addressed in the TAM framework. Another deficit of the TAM framework is that it does not address affective or emotional influences that may encourage system use (Bagozzi, 2007).

The extended TAM (TAM2) and the decomposed theory of planned behavior (C-TAM-TPB) were developed to further expand TAM to incorporate factors that were not addressed by Davis's original framework (Venkatesh & Davis, 2000).

Theory of planned behavior. Ajzen (1991) introduced a theory of planned behavior that incorporates some central concepts of the social and behavioral sciences. It was developed from the TRA framework (Fishbein & Ajzen, 1975). In addition to TRA, TPB contains perceived behavioral control as a variable to predict behavioral intention (see Figure 3). According to Ajzen (1991), attitudes toward the behavior, subjective norms, and perceived behavioral control predict behavioral intentions with a high degree of accuracy. As a result of these interactions, behavioral intentions could account for a considerable proportion of the variance in behavior. In addition, the model has determinants to measure attitudes, subjective norm, and perceived behavioral control.

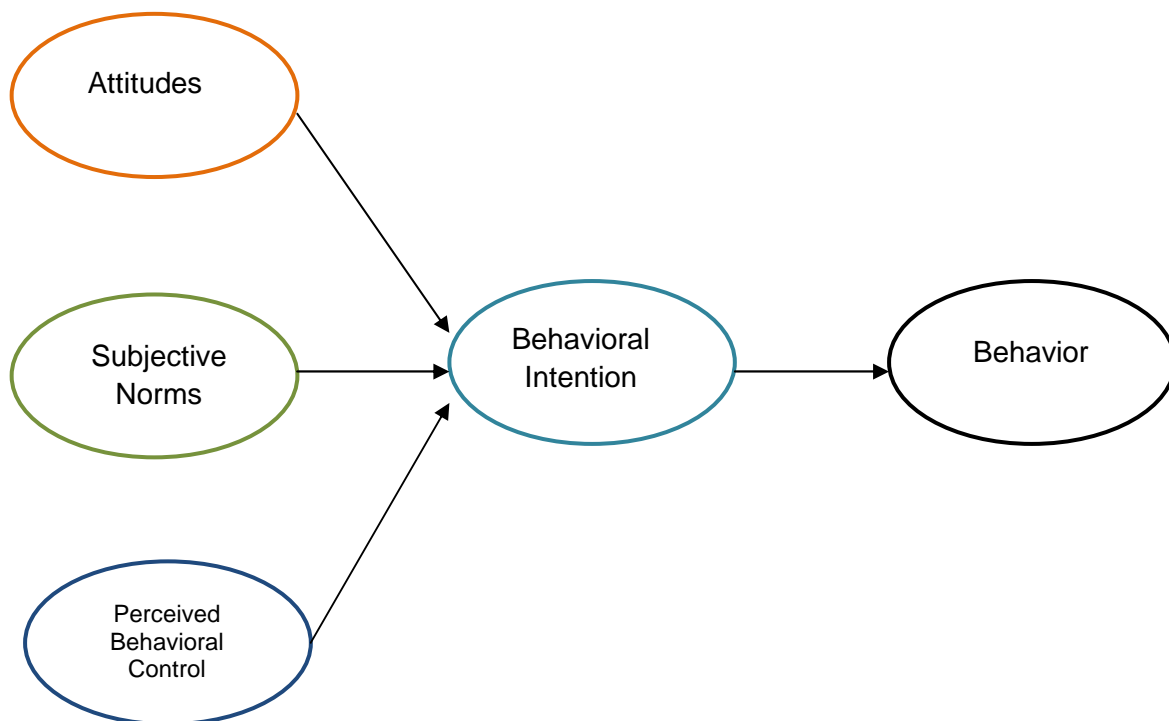


Figure 3. Theory of planned behavior model. Adapted from “The Theory of Planned Behavior,” by I. Ajzen, 1991, *Organizational Behavior and Human Decision Processes*, 50(2), p. 182.

Copyright 1991 by Elsevier.

Extended technology acceptance model. Davis (1989) pointed out that some external variables (e.g., user training, system characteristics) which may affect the technology acceptance model. However, theoretical and empirical studies did not find the supportive results for them. Since the development of TAM, there have been a few scholars devoted to the study of the original model and its improvements. For the core model amendment, behavioral intention to use was removed in the amended TAM model presented by Davis (1993). Davis and Venkatesh (1996) claimed that attitude is only the user emotion, and the preference of information technology cannot completely convey the impact of perceived usefulness and perceived ease of use on behavioral intention to use. Finally, Venkatesh and Davis (2000) developed a new amendment of the TAM model for external factors. The new model was called TAM2. This model adopted a social influence process and a cognitive instrumental process. Perceived usefulness was determined by those process variables (see Figure 4).

In the model of TAM2, subjective norm directly impacts perceived usefulness (PU), and indirectly impacts behavioral intention (BI). Subjective norm (SN) also indirectly impacts PU through image. At the same time, subjective norm has a direct effect on BI (see Figure 4). Therefore, Venkatesh and Davis (2000) showed that in an involuntary use environment, subjective norm had a greater impact on BI. The effect of the subjective norm on BI was also impacted by voluntariness. Voluntariness refers to when a user does not feel forced to accept a technology. When using a system, BI changes in accordance with the level of whether the user is willing to accept it. The subjective norm is also affected by experience. Hartwick and Barki (1994) showed that when users were less familiar with the system, the subjective norms variable had a significant impact. Nevertheless, increasing experience reduced the social influence for user acceptance.

TAM2 extended of TAM by showing that subjective norm exerts a significant direct effect on behavioral intentions above and beyond perceived usefulness and perceived ease of use for mandatory systems.

Because TAM was developed for the workplace context, it is difficult to use this model in a voluntary information technology system. However, Schneberger, Amoroso and Durfee (2008) questioned the factors that influence performance on computer based assessments. Based on the revised TAM model, they found that most of the predicted relationships from the TAM model were valid and that level of expertise was a strong factor that enhanced assessment performance. Level of expertise can be improved by training students for computer based assessment. In conclusion, there are also some factors (e.g., level of support, perceived usefulness, perceived ease of use, level of skill expertise, attitude toward using) that can affect performance on computer based assessments.



Figure 4. Extended technology acceptance model. Adapted from “User Acceptance Enablers in Individual Decision Making About Technology: Toward an Integrated Model,” by V. Venkatesh, C. Speier, and M. G. Morris, 2002, *Decision Sciences*, 33(2), p. 302. Copyright 2014 by John Wiley & Sons.

Combined technology acceptance model and theory of planned behavior. The combined TAM-TPB was proposed by Taylor and Todd (1995) and adapted from the theory of planned behavior (TPB). Taylor and Todd (1995) determined attitudes, subjective norm, and perceived behavioral control as important factors to explain technology use behavior. These researchers identified perceived usefulness, ease of use, and compatibility beliefs to explain attitudes; peer influence and superior's influence to explain subjective norm and self-efficacy and facilitating conditions to explain perceived behavioral control (see Figure 5).

The C-TAM-TPB is a more complicated model that increases the predictor power of behavior. This new model has some advantages over the prior acceptance models. First, administration of the C-TAM-TPB is less difficult and more time efficient. Second, it provides a more complete understanding of behavior and behavioral intention than other models such as TAM and TPB. A third advantage is that there is no need to develop new scales as the scales of TAM and TPB have been administered in hundreds of studies (King & He, 2006; Manning, 2009; Schepers & Wetzels, 2007).



Figure 5. Combined technology acceptance model and theory of planned behavior. Adapted from “Understanding information technology usage: A test of competing models,” by S. Taylor and P. A. Todd, 1995, *Information Systems Research*, 6(2), p. 163. Copyright 1995 by INFORMS.

Social cognitive theory. Social cognitive theory (SCT) has been utilized recently by information technology researchers not specifically to predict acceptance behaviors but rather to provide additional insights into the determinants of acceptance behaviors. Current social cognitive theory is rooted in the research of Albert Bandura and his colleagues (1986). The essence of Bandura's (1986) SCT rests in the notion of reciprocal triangle. Individual behavior is posited to be an outcome of a complex set of interactions between individual characteristics, environmental factors and situational factors. Behaviors, individual differences, and situational contingencies mutually influence one another. Although the theory is rich and complex, particular elements have been utilized to inform information technology research. The effects of the individual characteristic, self-efficacy, on technology acceptance outcomes are one such element.

Compeau and Higgins (1995a) found support for the positive effects of computer self-efficacy on computer usage, affect, and outcome expectations related to performance. In another study, Compeau and Higgins (1995b) examined the role of computer self-efficacy in the context of computer training. They empirically identified the influence of self-efficacy on performance as well as personal outcome expectations.

The model of PC utilization. The model of pc utilization (MPCU) was adapted from Triandis' (1971, 1979) theory of human behavior. This model presents a competing perspective to that proposed by TRA and TPB. In the earlier work by Triandis (1971), attitudes, social norms, habits, and consequences of behavior were considered what determined the behavior. In addition, Triandis (1971) suggested that attitudes involve cognitive, affective, and behavioral determinants. The cognitive component of attitudes includes beliefs and the affective component

of attitudes which has a like or dislike connotation. Triandis (1971) defined behavioral intentions by explaining simply what individuals intend to do.

Later, Triandis (1979) presented a more comprehensive model of human behavior. The major statement of this model is that social factors, affect, and perceived consequences determine behavioral intentions which in turn influence behavior (see Figure 6). In addition, Triandis (1979) claimed that habits were both direct and indirect determinants of behavior. He acknowledged that facilitating conditions are crucial factors for behavior even if the intention is high. Facilitating conditions are objective factors in the environment that people agree make an act easy to accomplish. For instance, PC usage support might be considered as a facilitating condition in information technology (Thompson, Higgins, & Howell, 1991). For example, if someone intends to use a PC but does not have easy access to one, usage is less likely to occur. The model includes other variables, such as culture, the social situation, and genetic biological factors that may influence behavior. Thompson et al. (1991) adapted Triandis' model for information system environment. According to these authors, long-term consequences of PC use, job fit with PC use, complexity of PC use, affect toward PC use, social factors, and facilitating conditions were determinants of PC utilization. Although the model by Thompson et al. (1991) is about personal computer utilization, the nature of the model makes it well-suited to predict individual acceptance and use of a range of information technologies.

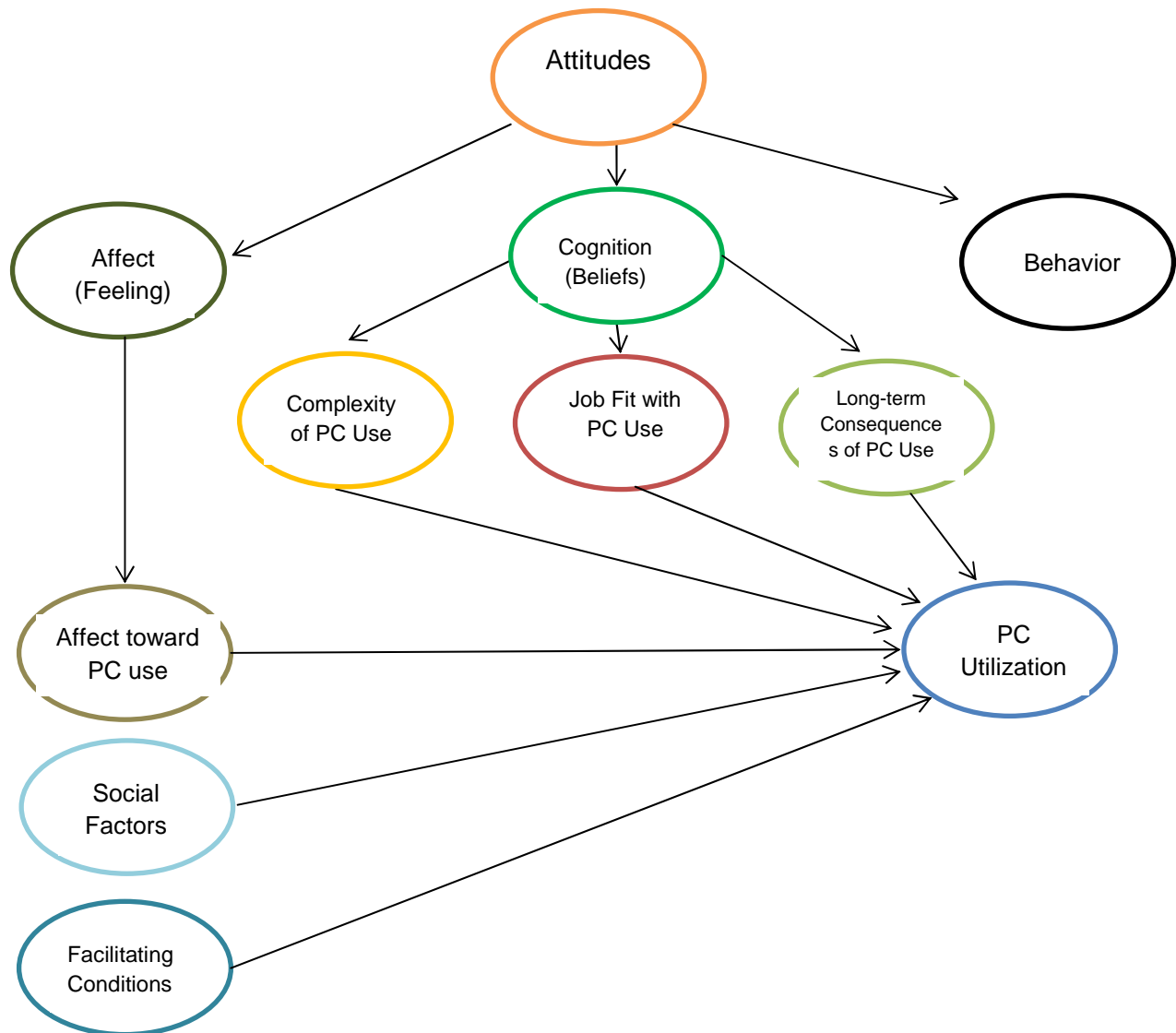


Figure 6. The model of PC utilization model. Adapted from “Personal Computing towards a Conceptual model of Utilization,” by R. L. Thompson and C. A. Higgins, 1991, *MIS Quarterly*, 15(1), p. 131. Copyright 2014 by the MISQ.

Innovation diffusion theory. An innovation is "an idea, practice, or object that is perceived as new by an individual or another unit of adoption" (Rogers, 1995, p. 14). Diffusion is the social process by which an innovation is communicated through certain channels over time among members of a social system (Rogers, 1995). The innovation diffusion theory (IDT) argues that "potential users make decisions to adopt or reject an innovation based on beliefs that they

form about the innovation" (Agarwal, 2000, p. 90). From the perspective of the IDT, adoption is predicted by perceived attributes of innovations, social norms, and individual characteristics (Rogers, 1995)

Rogers (1995) popularized the innovation diffusion theory. He stated that there were five significant innovation characteristics: relative advantage, compatibility, complexity, trialability, and observability. The first characteristic and the best predictor of the innovation adoption was relative advantage. It was defined as the degree to which an innovation is considered as being better than the idea it replaced. Second, compatibility was explained by consistency with the potential individuals' existing values, prior experiences, and needs. Complexity was another characteristic in the model. It related to perceived level of difficulty in understanding innovations and their ease of use. Trialability was defined as the degree to which innovations can be tested on a limited basis. Finally, observability referred to visibility of the innovation's results by other. These characteristics were used to explain user adoption of innovations and the decision-making process.

Diffusion of innovation research has been widely applied in education and information technology disciplines (Rogers, 1995; Karahanna & Straub, 1999; Agarwal, Sambamurthy, & Stair, 2000). Based on descriptions of the characteristics, the IDT research on those areas focused on the investigation of the process through which innovation was diffused through a social system over a time. Therefore, IDT provides an applicable theoretical basis for the computer technology.

Unified theory of acceptance and use of technology. Venkatesh, Morris, Davis, and Davis (2003) proposed a unified theory of acceptance and use of technology (UTAUT). This model contains constructs from the previously described models. Unified theory of acceptance

and use of technology posit that there are four key variables that determine information technology (IT) acceptance and four variables that are moderators of the main relationships. Performance expectancy, effort expectancy, social influence, and facilitating conditions are determinants of behavioral intention. Gender, age, experience, and voluntariness are moderators of acceptance of information technology. According to this model, performance expectancy, effort expectancy, and social influence have positive effects on behavioral intention. Additionally, behavioral intention and facilitating conditions have positive effects on user behavior (see Figure 7).

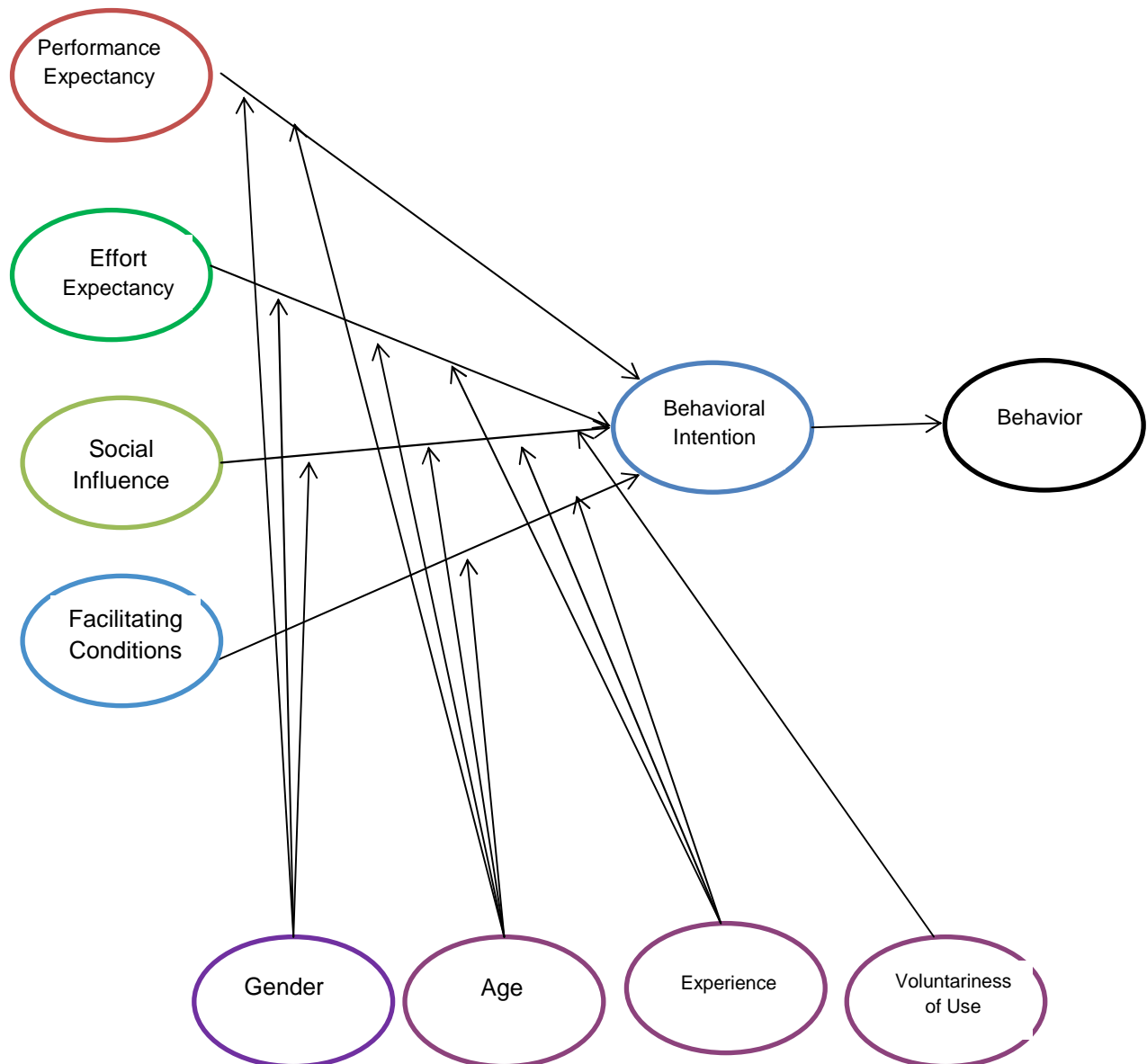


Figure 7. Unified theory of acceptance and use of technology model. Adapted from “User Acceptance of Information Technology: Toward a Unified View,” by V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, 2003, *MIS Quarterly*, 27(3), p. 447. Copyright 2014 by MISQ.

Computer-based assessment acceptance model. Computer-based assessment acceptance is defined as a test taker’s willingness to employ computer for the tests it is designed to support.

Terzis and Economides (2011a) developed a conceptual model called computer-based assessment acceptance model (CBAAM). It supports previous research in the fields of learning management systems acceptance and information technology acceptance. Because most of the previous studies focused on e-learning environments, CBAAM provided a first step toward the analysis of computer-based assessment. The model uses seven variables from previous models and adds two new variables. The CBAAM variables were adapted from the corresponding models of perceived ease of use and perceived usefulness from the technology acceptance model (Davis, 1989); social influence and facilitating conditions from the unified theory of acceptance and use of technology (Venkatesh, Morris, Davis, & Davis, 2003); perceived playfulness from an extended TAM version (Moon & Kim, 2001); and computer self-efficacy from Compeau and Higgins's model (1995b). In order to explain the intention to use a computer-based assessment, the CBAAM model proposed two additional variables: content and goal expectancy (Terzis & Economides, 2011a). Finally, behavioral intention to use was adapted from Davis's model (1989).

Figure 8 shows the original CBAAM model. This is a structural equation model with four exogenous variables and five endogenous variables. The four exogenous variables are social influence, facilitating conditions, goal expectancy, and content. The endogenous variables are perceived ease of use, perceived usefulness, perceived playfulness, computer self-efficacy, and behavioral intention to use. In previous studies (Moon, & Kim, 2001; Terzis, Economides, 2011a), perceived playfulness was defined by three dimensions: concentration, curiosity, and enjoyment. These linked and interdependent dimensions are considered crucial factors for implementation of a computer-based assessment (CBA). Moreover, if a test taker feels that CBA is useful then it will increase the examinee's concentration, curiosity and enjoyment.

Terzis and Economides (2011a) found that perceived usefulness had a positive effect on perceived playfulness. This link created an indirect effect of perceived usefulness on the behavioral intention through perceived playfulness. Perceived ease of use is defined as the degree of belief that using the system is free of effort. This will increase perceived playfulness because there is no annoying disturbance in the system. Furthermore, perceived ease of use indirectly affects the behavioral intention to use through perceived usefulness (Venkatesh, & Davis, 1996). Computer self-efficacy is determined as the individual's beliefs about his or her ability to use a computer (Compeau, & Higgins, 1998).

Venkatesh and Davis (1996) demonstrated a link between computer self-efficacy and perceived ease of use. They stated that an individual's judgment of their capability to use computers affected their perception of the level of ease of computer technology. Computer self-efficacy had a direct effect on perceived ease of use and an indirect effect on behavioral intention to use. Terzis and Economides (2011a) defined social influence as the effect of other people's opinion, peer influence and superior influence. They also found a causal link between social influence and perceived usefulness. Facilitating conditions were determined as services that facilitated a user to perform a procedure (Terzis, & Economides, 2011a). These services were offered by the system.

In the previous study (Terzis, & Economides, 2011a) showed that facilitating conditions had a positive effect on perceived ease of use. Terzis and Economides (2011a) also proposed goal expectancy as a variable. Goal expectancy influences an individuals' belief that she or he is adequately prepared to use computer-based assessment. It had two dimensions: examinee's preparation to take a CBA and the desirable level of success for each participant. They found that goal expectancy had positive effect on perceived usefulness and perceived playfulness. Finally,

content of examination was introduced to the CBAAM. Terzis and Economides (2011a) indicated that content of examination was important for the CBA' usefulness and playfulness. The questions in the CBA had to be clear, understandable and relative to the course's content in order to increase student satisfaction. They found that content had a direct and positive effect on perceived usefulness, perceived playfulness and goal expectancy, and an indirect effect on behavioral intention.

In summary, Terzis and Economides (2011a) supported the position that perceived ease of use and perceived playfulness had a direct effect on computer-based assessment use, particularly on the behavioral intention. More specifically, perceived ease of use is significantly attributed to computer self-efficacy, and to facilitating conditions. Perceived usefulness, computer self-efficacy, social influence, facilitating conditions, content of examination, and goal expectancy had only indirect effects on behavioral intention to use. Perceived usefulness was significantly attributed to content of examination, goal expectancy, social influence and perceived ease of use. Perceived playfulness was explained by perceived usefulness, perceived ease of use, content and goal expectancy defined by content of examination. Finally, behavioral intention to use a computer-based assessment was significantly attributed to perceived playfulness and perceived ease of use. The results of the study showed that these eight variables explained approximately 50% of the variance of behavioral intention (Terzis & Economides, 2011a).

The current study used an adaptation of the CBAAM and extended it in an attempt to identify possible developed status of country and gender effects among the variables toward computerized ToEFL acceptance.

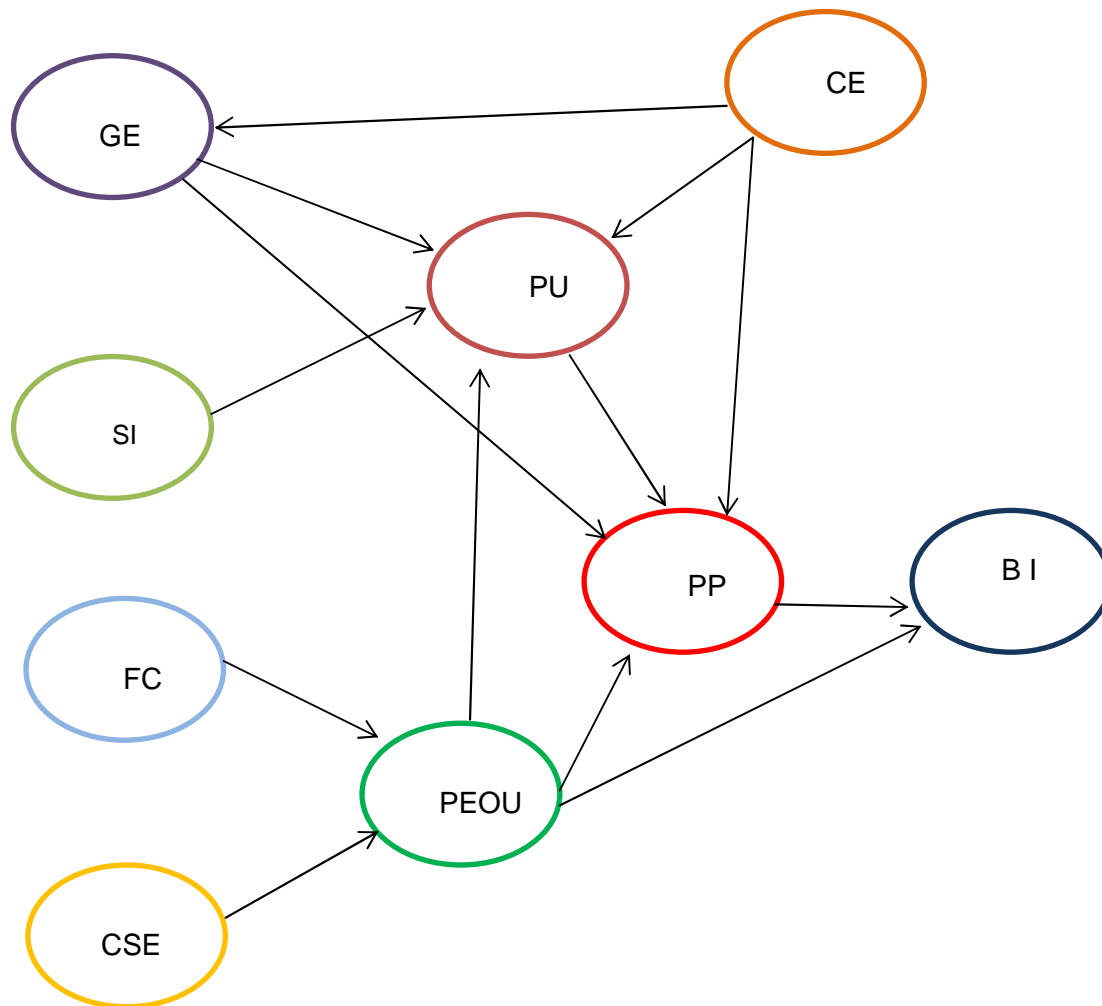


Figure 8. Computer-based assessment acceptance model. Ovals represent latent variables which are perceived playfulness (PP), perceived usefulness (PU), perceived ease of use (PEOU), computer self-efficacy (CSE), social influences (SI), facilitating conditions (F FC), goal expectancy (GE), content of examination (CE), and behavioral intention to use (BI). Arrows indicate factor loadings or regression coefficients. Adapted from “The Acceptance and Use of Computer Based Assessment,” by V. Terzis and A. A. Economides, 2011a, *Journal of Computers & Education*, 56(4), p. 1034. Copyright 2011 by Elsevier.

In the next section, individual differences especially developed status of countries, and gender differences in acceptance were explained.

Individual Differences in Acceptance

The term *individual differences* could be interpreted most generally to predict dissimilarities among people. It includes differences in perceptions, behaviors, individual traits and personality characteristics. In acceptance research, several individual difference variables have been studied. These variables include cognitive style (Benbasat & Taylor, 1978), gender, age, experience, culture, country, and personality (Harrison & Rainer, 1992; Taylor & Todd, 1995; Thompson et al., 1994), and motivation (DeSanctis, 1983). This study extends the literature by focusing on developed status of countries and gender differences.

Country development differences and acceptance. Despite continuing interest and ongoing investment in technology for the facilitation of development efforts, research on individual-level factors that influence users' acceptance of these systems has rarely been conducted. Information systems in developing countries have experienced high rates of failure and show important problems for computer-based testing context (Anandarajan, Igarria, & Anakwe, 2000; Heeks, 2002; Odedra, Lawrie, Bennett, & Goodman, 1993). Because there are concerns about informational technology usage in developing countries, this study investigated differences in the intention to use a computerized exam between developing and developed countries.

Lee, Breland, and Muraki (2005) conducted a study to investigate how ToEFL examinees who came from different language backgrounds performed on the ToEFL's listening, reading, and structure and written expression sections. Later the authors referred to the listening and reading sections as English Language Ability (ELA) and the structure and written expression

section as the writing section. The participants in this study were categorized into two language groups, European and Asian language groups. The results of the study showed that test takers from the European language group obtained higher scores in ELA compare to the Asian language group. The European language group also performed better than the Asian language group on their essays. This may indicate a problem with the test's fairness if European language test takers obtain a higher ToEFL score because they share similar educational knowledge of what was presented on the test that is American knowledge.

Xi (2010) argued the fairness concerns for the ToEFL IBT. One of the relevant fairness issues was whether test tasks were equally relevant to and representative of the subgroups. The ToEFL IBT scores are used for admitting both undergraduate and graduate applicants who do not speak English as their primary language, "A fairness issue is that the tasks did not assess some critical language skills required of undergraduate or graduate students" (Xi, 2010, p. 158).

In summary, high-stakes tests (e.g., ToEFL) created some problems that affect curriculum and student learning. The ToEFL test had issues of fairness because different country groups had different opportunities to succeed, and the test raters' inconsistency in assigning scores made their reliability questionable. I will now explain how test takers' acceptance may affect the fairness of the ToEFL test.

Many theories have been developed to explain individual usage decisions of assessment technologies and researchers have empirically examined these theories for different user groups. Researchers argued that beliefs and perceptions of individuals, which were major determinants of their acceptance behaviors toward technology, reflect the values of different countries (Veiga, Floyd, & Dechant, 2001; Png, Yan, & Wee, 2001; Tan, Watson, & Wei, 1995). Limited research has been conducted from the perspective of country differences in information technology

acceptance and e-learning technology acceptance literature (Veiga et al., 2001; Straub, Keil, & Brenner, 1997; Martinsons & Davison, 2003). Thus far, there is no published research of developed status of country as a factor to explain or predict difference in computer-based assessment acceptance.

Some technology acceptance research has shown that the region or country of the examinees may have an impact on their intention to use technology (Maldonado, Khan, Moon, & Rho, 2011; Zhao & Tan, 2010). Zhao and Tan (2010) stated a motivational perspective to explain behavioral intention to use an e-learning system. Empirical results indicated that Chinese students' and Canadian students' e-learning system acceptance were different from one another. They compared and contrasted that, unlike Canadian students; Chinese students think ease of use had an impact on their intention to use the e-learning systems. On the other hand, Arenas-Gaitán, Ramírez-Correa, and Rondán-Cataluña (2011) did not find significant cultural differences for every relationship of the TAM when they examined cultural differences and technology acceptances for students from Spain and Chile. Even though the results showed that cultural differences existed in both groups, there was no difference in university environments.

Grandon, Alshare, and Kwun (2005) proposed a new research model that was adapted from TAM. Their research study examined factors that influenced students' intentions to take online courses in the U.S. and South Korea. The data were collected from college students. The findings indicated that for American students' convenience, quality, subjective norm, and perceived ease of use were significant predictors of students' intention to use. However, only quality and subjective norms were significant factors impacting the Korean students' intentions.

Stricker and Attali (2010) published a study that assessed examinees' reported acceptance of the internet-based ToEFL. They evaluated differences in the pattern of results for examinees

from China, Colombia, Egypt, and Germany. Except among examinees from Germany, overall attitudes about the ToEFL IBT were moderately positive in other countries. Germany had negative and neutral attitudes toward ToEFL IBT. The divergent attitudes in Germany about the ToEFL raised the question of whether these attitudes were somehow unique to that country and if so, why, or whether they were widespread in other developed countries.

Based on the findings and conclusions of the investigations mentioned above, it could be affirmed that country differences affect the development and use of computer-based assessment technology, including the computerized ToEFL exam.

Gender differences and acceptance. A limited number of studies have examined the relationship between technology-mediated learning environments and individual differences. Some studies concentrated on gender discrepancies about information technology discipline (Arbaugh, 2000; Manochehr (2006); McSporrán & Young, 2001). The findings showed inconsistent and contradictory results. Keasar, Baruch, and Grobgeld-Dahan (2005) examined technology-mediated learning in science education for male and female students separately. They stated that there was no significant gender difference on students' learning for a biology class. However, McSporrán and Young (2001) concluded that technology-mediated learning gave more responsibilities to students. They argued that female students tended to be more effective with time management. Their results indicated that female students learn more effectively in a technology-mediated learning environment than their male counterparts. Analysis of previous research results suggests that, certain individual characteristics may affect learning effectiveness in technology-mediated learning environments.

Another study investigated how the digital divide affects the learning effectiveness of different student groups in technology-mediated learning (Chen, 1986). Chen stated that female

students may be at a disadvantage for technology-mediated learning because female students have lower computer self-efficacy and technology usage than male students. Meyers, Bennett, and Lysaght (2004) investigated adult women in rural areas and their experiences in technology-mediated learning. This study reported on problems of technology usage for women and suggested several strategies for making technology-mediated learning more equitable.

For computer usage and computer self-efficacy, gender may be considered a crucial factor. Males and females use computer technology in different ways. For example, some research studies indicated that male users have more knowledge, usage, and experience in terms of computer technology (He & Freeman, 2010; Li & Kirkup, 2007; Link & Marz, 2006). Those studies revealed that male computer users have higher computer self-efficacy than female users. In addition, males reported a higher degree of information communication technology usage for leisure purposes (Li & Kirkup, 2007; Tomte & Hatlevik, 2011). Another study suggest that if the test is enjoyable and the content is relevant, then both female and male test takers perceived computer-based assessments the same way (Terzis & Economides, 2011a). Female students valued easiness and facilitating conditions of the computer-based assessments more than male students. However, perceived usefulness and social influence are two important determinants of male computer-based assessment acceptance (Terzis & Economides, 2011b). Despite the gender differences described in the literature, the performance difference on computerized tests between female and male students was not reported in some studies (Kies, Williams, & Freund, 2006). Because of these contradictory results, little is known about how gender mediates computer-based assessment acceptance.

Although there are numerous studies relating to gender variations on acceptance of e-learning and information technology systems (Wang, Wu, & Wang, 2009) only one study was

identified to evaluate gender differences on acceptance of computer-based assessments (Terzis & Economides, 2011b). Many studies have examined gender differences regarding e-learning system acceptance (Gefen & Straub, 1997; Ong & Lai, 2006; Wang et al., 2009) and located diverse results. For example, some studies showed no gender gap regarding intentions, (Cheung, Lee, & Chen, 2002; Yuen & Ma, 2002) and alternative studies found that men were more motivated by perceived usefulness on their intention to use the computer for assessment (Gefen & Straub, 1997; Sun & Zhang, 2006). Additionally those studies showed that women were more influenced by perceived ease to use (Ong & Lai, 2006). Terzis and Economides (2011b) found that both genders were equally likely to use the computer-based assessment if it was playful and its content was clear. Moreover, usefulness of CBA was vital for men and their attitudes toward using CBA were influenced by their social environments. However, women preferred CBA if it was straightforward and easy to understand. Based on these results, it was expected that this study will have results similar to previous gender effect studies.

Gender is important and has implications for issues related to diversity and equal opportunity. Understanding gender effects on students' acceptance and addressing the key barriers commonly experienced by the disadvantaged gender was crucial if system developers and instructors were design better computer-based assessment systems.

Conclusions

The evolution of theoretical frameworks and statistical models has identified a number of key variables to explain differences on acceptance and intention to use CBA.

In recent years, a significant number of studies evaluating the impact of contextual factors (such as country and gender difference) on technology usage have occurred. In terms of technology acceptance studies, Venkatesh and Morris (2000) highlighted gender differences

when evaluating new technologies. Technology usage decisions were strongly influenced by perceptions of usefulness for men, while women were strongly influenced by perceptions of ease of use and subjective norm. To extend the current state of knowledge this study explored the developed status of country and gender differences on student acceptance of an internet-based ToEFL.

CHAPTER 3: METHODOLOGY

Computer delivery of assessments remains a new and evolving phenomenon in the educational setting. As student demographics continue to change, test administrators, teachers, and staff continue to develop new techniques to meet the growing demands of their constituents. In order to best meet the needs of today's students for effective assessment delivery, researchers must assess those techniques that students perceive as the most useful, effective, and acceptable.

This chapter describes the methods and procedures that were used in this study. All procedures were approved by the university Institutional Review Board prior to the start of the study. The first part of this chapter explains the rationale for the research design, the variables employed, and a description of the study's participants. Next, the data collection method and procedures are described. Finally, the data analysis procedures are explained to complete the description of the research methods and procedures for this study.

Research Design

The aim of the study was to explore how test takers from different countries and the different genders accept the computerized ToEFL. This study built upon the research of those who examined computer-based testing acceptance in relation to specific potential variables. Because the ToEFL IBT is a type of computer based assessment, the CBAAM model can easily adapted to the ToEFL environment. Therefore, I used the CBAAM as the conceptual model for this study. Participants in this study represented 13 developed countries and 18 developing countries. Using the World Bank definitions of "developed" and "developing" countries, developed countries in this study were South Korea, Taiwan, Canada, France, Italy, Germany,

Hong Kong, Singapore, Japan, Sweden, Finland, Denmark, and Portugal. Countries identified as developing were Turkey, China, Colombia, Paraguay, Thailand, Russia, Tanzania, Vietnam, Mexico, Ukraine, Serbia, Indonesia, Puerto Rico, Malaysia, Romania, India, Brazil, and Chile.

Variables

Based on the review of theories, experiences of experts in this area of research, and the associated literature (Terzis & Economides, 2011a), a hypothesized model consisting of a network of links among the identified nine variables was represented with a path diagram illustrated in Figure 9. To test whether the study model was consistent with the data, a structural equation modeling (SEM) approach was employed, and the hypotheses indicated in the following subsections of this chapter refer to each relationship among the nine variables in the model (Figure 9). The hypotheses describe whether a variable is positively related to another variable, or whether the effect of a variable is mediated by another variable or variables.

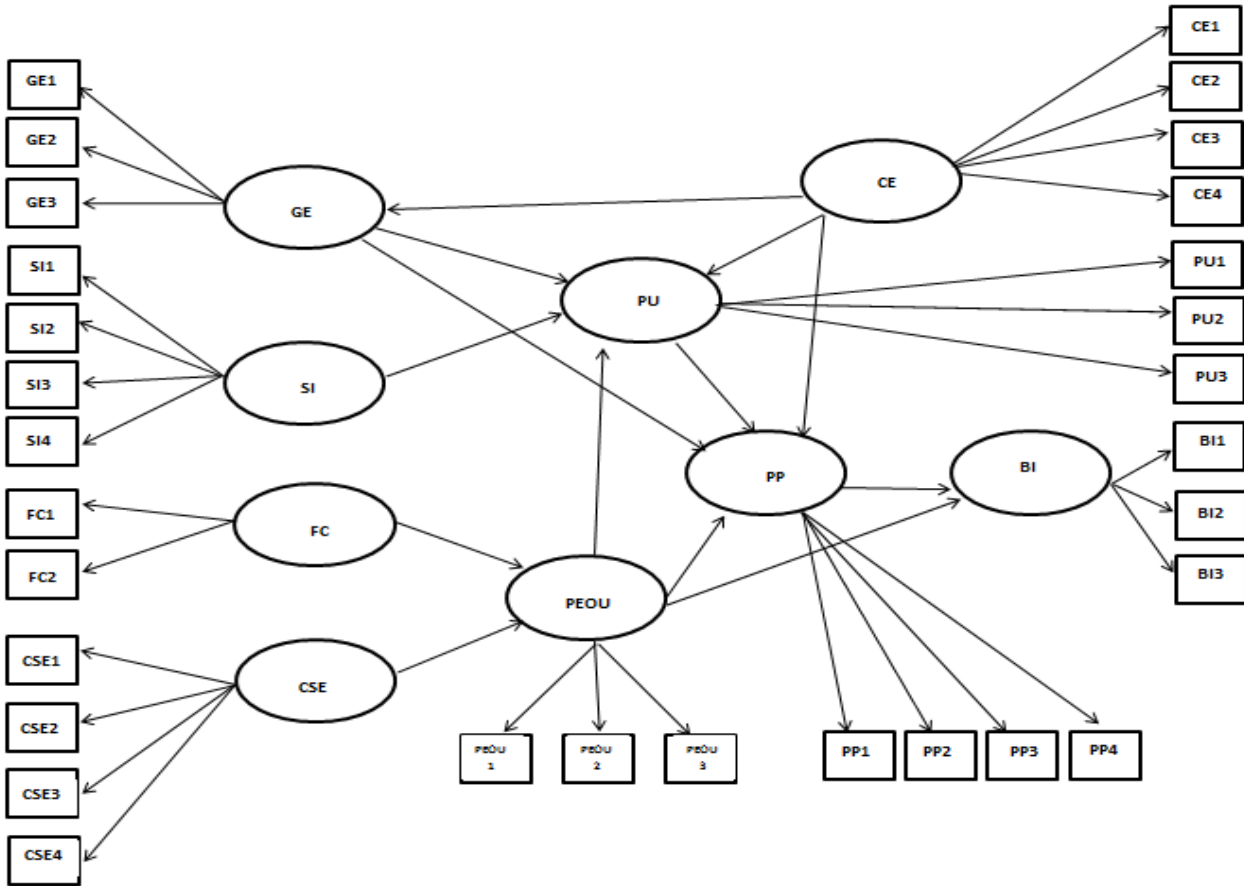


Figure 9. Computerized ToEFL acceptance model. Ovals represent the latent variables which are perceived playfulness (PP), perceived usefulness (PU), perceived ease of use (PEOU), computer self-efficacy (CSE), social influences (SI), facilitating conditions (FC), goal expectancy (GE), content of examination (CE), and behavioral intention to use (BI). Squares indicate the observed variables which are 30 questionnaire items. Arrows indicate the factor loadings or regression coefficients. Adapted from “The Acceptance and Use of Computer Based Assessment,” by V. Terzis and A. A. Economides, 2011a, *Journal of Computers & Education*, 56(4), p. 1034. Copyright 2011 by Elsevier.

Perceived playfulness. Csikszentmihalyi (1975) and Deci and Ryan (1985) proposed that *perceived playfulness* (PP) has an important positive effect on behavioral intention to use the internet. They extended TAM by adding PP and they defined the concept by three dimensions: concentration, curiosity, and enjoyment. These three dimensions are linked and interdependent, but they are not always observed together in practice. The three dimensions of *perceived playfulness* are considered significant factors for the successful implementation of a computerized ToEFL. This type of ToEFL must hold the examinee's concentration, curiosity, and enjoyment at high levels. Because *perceived playfulness* has a positive effect on the behavioral intention of a CBA (Terzis & Economides, 2011a), *perceived playfulness* would have a direct effect on the behavioral intention to use.

Because the computer based ToEFL is more game-oriented than other information technologies, it was expected that perceived playfulness would be higher for men than for women and that it also had greater influence on behavioral intention to use for men than for women.

Previous studies have found that men enjoy playing computer games more than women (Bonanno & Kommers, 2008; Hartmann & Klimmt, 2006; Scott & Rockwell, 1997). Furthermore, men expressed more positive feelings towards multiple choice assessments than females (Birenbaum & Feldman, 1998). As a result, men might use the computerized ToEFL as a game to enhance and test their knowledge. In addition, because perceived playfulness comprises concentration, curiosity and enjoyment, people who are from a developing country would enjoy more than those from a developed country. Computer technology is a new thing for them and this innovation may trigger their curiosity and as a result it affects their perceived playfulness. I hypothesized:

- H1. Perceived playfulness will be higher for men than for women;
- H2. Perceived playfulness will have a direct effect on behavioral intention to use a computerized ToEFL, more strongly for men than for women;
- H3. Perceived playfulness will be higher for persons from developed countries than for those from developing countries; and
- H4. Perceived playfulness will have a direct effect on behavioral intention to use a computerized ToEFL more strongly for people from developed countries than people from developing countries.

Perceived usefulness. *Perceived usefulness* (PU) is defined as a person's belief that using an information technology system increases an individuals' performance (Davis, 1989). This construct is one of the two main TAM determinants. Davis (1989) described PU as the most important determinant of technology acceptance, other than perceived ease of use. Many research studies have shown that there is support for the effect of PU on the behavioral intention to use a learning system (Lee, 2008; Ong & Lai, 2006; Van Raaij & Schepers, 2008). Similarly, examinees believed that a computerized ToEFL system improved their knowledge, comprehension, and performance for English language. If the ToEFL IBT was useful for the test takers, then it might enhance their concentration, curiosity, and probably enjoyment. It was expected that there would be a positive effect of *perceived usefulness* on *perceived playfulness*. This link created an indirect effect of *perceived usefulness* on the behavioral intention to use through the *perceived playfulness*.

Previous studies have shown a moderate effect of gender on PU (Ong & Lai, 2006; Venkatesh & Morris, 2000). Those studies also showed that the direct effect of PU on BI and on PP was stronger for men than for women. However, *perceived usefulness* was more important in

developed countries whereas *perceived ease of use* was more relevant in developing countries (Schepers & Wetzels, 2007). This was consistent with previous findings by Straub, Keil, and Brenner (1997) and McCoy, Everard, and Jones (2005). These studies suggested that TAM did not fit developing countries' attitudes. Mao, Srite, Bennett, and Yaprak (2005) concluded that *perceived usefulness* was less and *perceived ease of use* was more important in developing countries.

I hypothesized:

H5. Perceived usefulness will be higher for men than for women;

H6. Perceived usefulness will have an effect on behavioral intention to use computerized ToEFL more strongly for men than for women;

H7. Perceived usefulness will have a direct effect on perceived playfulness more strongly for men than for women;

H8. Perceived usefulness will be higher for developed countries than for developing countries;

H9. Perceived usefulness will have a direct effect on behavioral intention to use computerized ToEFL, more strongly for developed countries than for developing countries; and

H10. Perceived usefulness will have a direct effect on perceived playfulness, more strongly for developed countries than for developing countries.

Perceived ease of use. Davis (1989, p. 320) defined *perceived ease of use* (PEOU) as "the degree to which a person believes that using a particular system would be free of effort." PEOU has two important roles within TAM. It directly affects behavioral intention to use as well as impacts intention to use over PU. The Technology Acceptance Model postulates a strong

positive relationship between PEOU and PU (Venkatesh, 2000). PEOU enhances PP because perceived ease of use provides a smooth use of the system without annoying disturbances. Thus, it is expected that a positive effect of *perceived ease of use* on *perceived playfulness*.

Furthermore, *perceived ease of use* indirectly affects the *behavioral intention to use*, through its effect on *perceived usefulness* and on *perceived playfulness*.

Perceived ease of use may be more important for women, because it has been routinely reported that men were more familiar than women with computer use (Ong & Lai, 2006; Venkatesh et al., 2003). It was found that the effect of PEOU on BI, PP, and PU was stronger for women. However, developed countries have more opportunities for citizens to access computer technologies. This could affect people's perceptions of the easiness of computer technology.

Consequently, I hypothesized that:

H11. Perceived ease of use will be higher for men than for women;

H12. Perceived ease of use will have a direct effect on behavioral intention to use computerized ToEFL, more strongly for women than for men;

H13. Perceived ease of use will have a direct effect on perceived usefulness to use computerized ToEFL, more strongly for women than for men;

H14. Perceived ease of use will have a direct effect on perceived playfulness to use computerized ToEFL, more strongly for women than for men;

H15. Perceived ease of use will be higher for developed countries than for developing countries;

H16. Perceived ease of use will have a direct effect on behavioral intention to use computerized ToEFL, more strongly for developed countries than for developing countries;

H17. Perceived ease of use will have a direct effect on perceived usefulness to use computerized ToEFL more strongly for developed countries than for developing countries; and,

H18. Perceived ease of use will have a direct effect on perceived playfulness to use computerized ToEFL, more strongly for developed countries than for developing countries.

Computer self-efficacy. In this study, self-efficacy, a key element in Bandura's (1977) social learning theory, refers to belief in one's capability to use the internet. Self-efficacy has been found to influence the decision to use computers (Hill, Smith, & Mann, 1987). As Oliver and Shapiro (1993) presented, previous experience could increase or decrease one's self-efficacy. According to Bandura's (1977) self-efficacy theory, judgments of self-efficacy are based on several kinds of information including performance accomplishments (i.e., using computers successfully), vicarious experiences (i.e., observing others using computers successfully), verbal persuasion, and emotional arousal. From the results of the studies on the direct and mediating effect of technology acceptance models (Agarwal, Sambamurthy, & Stair, 2000; Padilla-Melendez, Garrido-Moreno, & Del Aguila-Obra, 2008; Venkatesh & Davis, 1996), it was theorized that a causal link exists between *computer self-efficacy* (CSE) and *perceived ease of use*. Thus, CSE has an important direct effect on PEOU and an indirect on *behavioral intention to use the system*.

In addition to finding a causal link between CSE and PEOU, researchers also have demonstrated higher levels of CSE for men (Comber, Colley, Hargreaves, & Dorn, 1997; Durndell & Hagg, 2002; Durndell, Hagg, & Laithwaite, 2000; Vekiri & Chronaki, 2008). Moreover, Ong and Lai (2006) suggested that CSE influenced PEOU more strongly for women

than for men. Prior research related to training that was conducted in developed countries showed that user training and computer experience were positively related to usage (Gannon, 1994; Grant, 1989). More specifically, higher computer self-efficacy predicts higher behavioral intention to use computers.

I hypothesized:

H19. Computer self-efficacy will be higher for men than for women;

H20. Computer self-efficacy will influence perceived ease of use more strongly for women than for men;

H21. Computer self-efficacy will be higher for developed countries than developing countries;

H22. Computer self-efficacy will influence perceived ease of use more strongly for developing countries than developed countries;

Social influence. *Social influence* (SI) has been used widely to explain collective behavior (Bagozzi & Lee, 2002). *Social influence* is defined as the degree to which an individual believes that important others (i.e., family, faculty members, colleagues, and students) support his or her system usage. There are three key elements to *social influence*: subjective norms, image and voluntariness (Karahanna & Straub, 1999). *Social influence* is accepted as a direct determinant of behavioral intention to use a technology and is represented as subjective norm, social factors, and image in many different technology acceptance theories. Although they had different names, all of these terms refer to the same construct. For an individual, the opinion of significant others influences the path to user acceptance. The role of *social influence* in technology acceptance decisions was complex and subject to a wide range of contingent influences.

Many students feel insecure regarding the use of computerized ToEFL. They may not have used a similar computerized system. It is evident that students consider the opinions of their colleagues, their friends and their seniors of taken into account their opinions. The primary topic in their discussions is the usefulness and the added value of the system. Thus, it was predicted that *social influence* had a direct effect on *perceived usefulness*. *Social influence* had been used in many proposed models (Agarwal & Karahanna, 2000; Karahanna & Straub, 1999; Taylor & Todd, 1995; Venkatesh & Davis, 2000; Venkatesh et al., 2003; Wang et al., 2009). Additionally, TAM2 supported the causal link between subjective norms with users' perceptions about the system's usefulness and the CBAAM found a causal link of SI to PU (Terzis & Economides, 2011a).

Previous studies have suggested that emotions and social factors more strongly affect women. Therefore, I hypothesized that women's social influence effect was stronger on behavioral intention than men's (Venkatesh & Morris, 2000). The effect on behavioral intention was only indirect through the *perceived usefulness*. Hence, I hypothesized that:

H23. Social influence will be higher for women than for men; and

H24. Social influence will have a direct effect on perceived usefulness more strongly for women than for men.

Technology could be more easily adopted by the students from the developing countries, due to the stronger influence of teachers, friends, and family (Maldonado, Khan, Moon, & Rho, 2009). Therefore, SI was one of the major predictors for BI.

H25. Social influence will be higher for developing countries than for developed countries; and

H26. Social influence will have a direct effect on perceived usefulness more strongly developing than for developed countries.

Facilitating conditions. *Facilitating conditions* (FC) are defined as the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system (Terzis & Economides, 2011a). This definition is a combination of three different constructs: perceived behavioral control (theory of perceived behavior (TPB), decomposed of TPB, and combined-TAM-TPB), facilitating conditions (model of personal computer utilization), and compatibility (innovative diffusion theory). The results of empirical studies suggested that the relationships between each of the constructs and intention were similar. The computer system and the organizational staff comprised the FC. In the computer-based assessment system, tutorials and help tools are designed to help students when they encounter technical difficulties. In the computerized ToEFL experience, the support staff played a significant role. During the ToEFL IBT, the presence of an expert played an important role in order to overcome students' questions about the use of the ToEFL IBT or even questions about the content of the exam.

Because previous studies claimed that women's computer self-efficacy was lower (Comber, Colley, Hargreaves, & Dorn, 1997; Durndell, Hagg, & Laithwaite, 2000; Vekiri & Chronaki, 2008; Whitely, 1997), FC would be a more important determinant for women than for men in order to overcome their computer anxiety. Therefore, the effect of FC on PEOU was stronger for women than for men. However, the positive relationship between organizational support and system usage was also documented in the literature (Fishbein & Ajzen, 1975). These studies indicated that lack of organizational support affected effective utilization of computers (Davis, 1989; Fornell, 1982). It was also found that organizational support was associated with greater system usage (Davis, 1989; Hair, Anderson, Tatham, & Black, 1992).

As a result, I hypothesized that:

H27. The mean of the facilitating conditions variable will be higher for women than for men;

H28. Facilitating conditions will have a direct effect on perceived ease of use, more strongly for women than for men;

H29. The mean of the facilitating conditions variable will be higher for developed countries than for developing countries; and

H30. Facilitating conditions will have a direct effect on perceived ease of use, more strongly for developed countries than for developing countries.

Goal expectancy. *Goal expectancy* (GE) was proposed in the CBAAM (Terzis & Economides, 2011a). *Goal expectancy* is a variable that influences an individual's belief that he or she prepared properly to use the computer-based assessment. *Goal expectancy* has two important dimensions: preparation to the test and desirable level of success. The first dimension is student's preparation to take the CBA, in this case the computerized ToEFL. It is clear that a tutor is not able to measure a student's preparation from either a qualitative or quantitative approach through the questionnaire and the system. Thus, the computerized ToEFL acceptance model actually measured if a student was satisfied with his or her preparation. The students usually tried to predict their performance based on their preparation and the hypothetical difficulty level of the exam. In other words, they evaluated their self-confidence regarding the preparation and the assessment. The second dimension is the desirable level of success for each examinee. Terzis and Economides (2011a) showed that, according to the CBAAM, there were positive effects of GE on PU, and GE on PP. Therefore, it was assumed that there were links between GE and PU, and GE and PP.

Previous studies showed that men were considered to be more competitive and aggressive than women (Eagly, Mladinic, & Otto, 1991; Gneezy & Rustichini, 2004). It was reported that men were also more concerned with winning than women (White & Duda, 1994). The findings of research studies with computer games stated that males were more motivated by challenge than females (Eglesz, Feteke, Kiss, & Izso, 2005; Greenberg, Sherry, Lachlan, Lucas, & Holmstrom, 2010). Because the GE variable was only recently added to a technology acceptance model, there was not much research to support linkages. Without a supporting literature base I was not able to assume differences between developed and developing countries.

I hypothesized:

H31. The mean of goal expectancy will be higher for men than women;

H32. Goal expectancy will have a direct effect on perceived usefulness, more strongly for men than for women;

H33. Goal expectancy will have a direct effect on perceived playfulness, more strongly for men than for women;

H34. The mean of goal expectancy will be the same for developed and developing countries;

H35. Goal expectancy will have a direct effect on perceived usefulness, equally for developed and developing countries; and

H36. Goal expectancy will have a direct effect on perceived playfulness, equally for developed and developing countries.

Content of examination. Based on a study by Doll and Torkzadeh (1988), Wang (2003) proposed the content construct as one of the determinants of e-learner satisfaction. Wang examined whether the content was up-to-date, sufficient, satisfied, and useful. Shee and Wang

(2008) also proposed the content variable as a determinant for students' satisfaction. They mentioned the need for nontechnical experts during the construction, operation, and maintenance of the system. *Content of examination* (CE) was first introduced into the assessment acceptance literature by Terzis and Economides in 2011. They believed that content affected CBA usage and they proposed two dimensions of the content variable. The first dimension was related to the course's content. The course's content was a criterion for the student to evaluate whether the course was difficult or easy, interesting or boring, useful or not useful. CE was related to the CBA's usefulness and playfulness elements (Terzis & Economides, 2011a).

Likewise, ToEFL IBT test takers evaluated the exam with regard to its content. The content of the ToEFL exam affected the usefulness and playfulness of the test. The second dimension was related to the questions during the CBA. The questions had to be clear, understandable and relative to the course's content in order to maximize the student's perceptions of utility and satisfaction. Previous CBAAM research indicated that CE had a positive impact on perceived usefulness, perceived playfulness and goal expectancy (GE) variables (Terzis & Economides, 2011a). I assumed that for computerized ToEFL acceptance, the same relationships held.

Examinees' computer self-efficacy was examined in order to highlight gender differences regarding the effect of CE on PU, PP, and GE. Previous studies have shown that female students were likely to have less positive perceptions towards computer self-efficacy than males (Vekiri & Chronaki, 2008). Therefore, it was expected that men would score higher than women on this variable. As with the GE variable, *content of examination* was a new construct in acceptance theories. There was no research indicating differences in the CE variable between developed and developing countries. Thus, I hypothesized:

H37. Content of examination will be a more important determinant to predict behavioral intention to use computerized ToEFL for men than women;

H38. Content of examination will have a direct effect on perceived usefulness, more strongly for men than for women;

H39. Content of examination will have a direct effect on perceived playfulness, more strongly for men than for women;

H40. Content of examination will have a direct effect on goal expectancy, more strongly for men than for women;

H41. The mean of content of examination will be for developing countries and developed countries;

H42. Content of examination will have a direct effect on perceived usefulness, equally for developed and developing countries;

H43. Content of examination will have a direct effect on perceived playfulness equally developed and developing countries; and

H44. Content of examination will have a direct effect on goal expectancy, equally for developed and developing countries.

Behavioral intentions to use. *Behavioral intention to use* (BI) is defined as the strength of the prospective individuals' intention for usage of information technologies. *Behavioral intention* was theorized to be one of the most important indicators of acceptance (Ajzen & Fishbein, 1980) and was a component of the TAM theoretical model. In the present study, *behavioral intention to use* is the primary dependent variable. Table 1 and Table 2 summarize all of the hypotheses for gender and developed status differences.

Table 1

Hypotheses for Gender Difference

Hypothesis Number	Variable or Relationship	Hypothesized Direction of Effect
H1	Perceived Playfulness	Men > Women
H5	Perceived Usefulness	Men > Women
H11	Perceived Ease of Use	Men > Women
H19	Computer Self-efficacy	Men > Women
H23	Social Influence	Women > Men
H27	Facilitating Conditions	Women > Men
H31	Goal Expectancy	Men > Women
H37	Content of Exam	Men > Women
<i>Relationships</i>		
H2	PP → BI	Men > Women
H6	PU → BI	Men > Women
H7	PU → PP	Men > Women
H12	PEOU → BI	Women > Men
H13	PEOU → PU	Women > Men
H14	PEOU → PP	Women > Men
H20	CSE → PEOU	Women > Men
H24	SI → PU	Women > Men
H28	FC → PEOU	Women > Men
H32	GE → PU	Men > Women
H33	GE → PP	Men > Women
H38	CE → PU	Men > Women
H39	CE → PP	Men > Women
H40	CE → GE	Men > Women

Notes. PP = perceived playfulness. PU = perceived usefulness. PEOU = perceived ease of use. CSE = computer self-efficacy. SI = social influence. FC = facilitating conditions. GE = goal expectancy. CE = content of examination. BI = behavioral intention to use.

Table 2

Hypotheses for Developed Status Difference

Hypothesis Number	Variable or Relationship	Hypothesized Direction of Effect
H3	Perceived Playfulness	Developed > Developing
H8	Perceived Usefulness	Developed > Developing
H15	Perceived Ease of Use	Developed > Developing
H21	Computer Self-efficacy	Developed > Developing
H25	Social Influence	Developing > Developed
H29	Facilitating Conditions	Developed > Developing
H34	Goal Expectancy	Developed = Developing
H41	Content of Exam	Developed = Developing
<i>Relationships</i>		
H4	PP → BI	Developed > Developing
H9	PU → BI	Developed > Developing
H10	PU → PP	Developed > Developing
H16	PEOU → BI	Developed > Developing
H17	PEOU → PU	Developed > Developing
H18	PEOU → PP	Developed > Developing
H22	CSE → PEOU	Developing > Developed
H26	SI → PU	Developing > Developed
H30	FC → PEOU	Developed > Developing
H35	GE → PU	Developed = Developing
H36	GE → PP	Developed = Developing
H42	CE → PU	Developed = Developing
H43	CE → PP	Developed = Developing
H44	CE → GE	Developed = Developing

Notes. PP = perceived playfulness. PU = perceived usefulness. PEOU = perceived ease of use. CSE = computer self-efficacy. SI = social influence. FC = facilitating conditions. GE = goal expectancy. CE = content of examination. BI = behavioral intention to use.

Methodological Approach

A methodological approach was needed to determine a scientific framework to examine computerized ToEFL acceptance. This study used the CBAAM framework which is a common approach for studying technology acceptance. In line with previous empirical research, the CBAAM framework was adapted and applied to the responses of the examinees of a computerized ToEFL exam with respect to *perceived playfulness* (PP), *perceived usefulness* (PU), *perceived ease of use* (PEOU), *computer self-efficacy* (CSE), *social influence* (SI), *facilitating conditions* (FC), *goal expectancy* (GE), *content of examination* (CE), and *behavioral intention to use* (BI). The chosen methodology selected for this study was developed on the premise that CBAAM was a proven model, used in numerous previous technology studies (e.g., Terzis & Economides, 2011b, 2012; Terzis, Moridis, Economides, 2013), as well as innovative scientific framework of prediction, which recorded intentions and attitudes of respondents.

Survey was the data collection method in this quantitative study. A systematic, purposive survey of international students from U.S. universities explored their intentions (BI) regarding PP, PU, PEOU, CSE, SI, FC GE, and CE. The responses from the participants were used as input into the computerized ToEFL acceptance model (CTAM) framework. The survey data were analyzed using the statistical software Mplus 7.1 (Muthen & Muthen, 2012).

Research Participants

The participants in the study were 237 international students attending U.S. universities. These international students represented different countries in Europe, Middle East, East and South East Asia, and South America. International students attending American universities were

chosen as the study sample for two main reasons. First, selecting participants that had certain characteristics in common backgrounds was important (Cohen, Manison, & Morrison, 2011; Krueger & Cassey, 2009). In this research, the common experience of taking a computerized ToEFL test was considered an essential characteristic since it was the main data to be analyzed in order to investigate perceptions about the test. The graduate and undergraduate international students were assumed to have experiences with taking the ToEFL IBT or ToEFL CBT test as one of the requirements that they would have met in order to be admitted to a university. The second reason for selecting international students in the U.S. as the target population of this study was because of assumed familiarity with the ToEFL test and more background knowledge about language proficiency tests. Background knowledge about these tests was expected to provide richer information about their perceptions of the computerized ToEFL test. These two factors made international students in the U.S. a suitable sample for this study.

This study used a convenience sample since the empirical research was based on a self-selection sampling method. The participants were non-native English speaking, international students in the U.S. The survey was administered during fall semester of 2012 and spring semester of 2013 by contacting international student associations from various universities in the U.S. To make it representative, one student from each different region of the U.S. was contacted and asked to distribute the survey in his or her school. Furthermore, the survey link was posted to international student organizations' Facebook and Twitter pages to introduce the study to the people who might be interested. Only students who were willing volunteers participated in the study. There was no compensation for participating rather than Amazon gift card drawing.

Based on the participants' self-reports, 48% were female and 52% were male. In addition, 57 of the participants were from developed countries such as Korea, France, and Canada; and

180 of the participants were from developing countries such as Turkey, China, and Mexico. The response rate was 79% in the current study.

Instrumentation

The Computer Based Assessment Acceptance Questionnaire was developed by Terzis and Economides (2011); I adapted items from that questionnaire based on ToEFL IBT format for this study. The final instrument for this study was a self-report questionnaire (see Appendix: Computerized ToEFL Acceptance Questionnaire) that consisted of two sections. The questionnaire was preceded by a cover letter that explained the nature of the research, the estimated time necessary to complete the survey, the voluntary nature of participation, and a statement regarding informed consent. Section one of the survey contained several demographic questions. Section two contained 30 questions to measure the nine primary variables used in the study. All items were written in English, and participants were not obligated to complete the questionnaires.

In order to examine the nine latent constructs of the model, items were modified based on previous studies and ToEFL content. Modification of items was necessary for the content validity of the study. First, items were altered by changing the words "computer-based assessment" to "ToEFL IBT exam." For example, one item was changed to: "Using the ToEFL IBT will improve my work of learning English." Moreover, some items were changed because they did not directly reflect the purpose of study. For example, the item "Test preparation was sufficient for the internet-based ToEFL" was replaced by "Course preparation was sufficient for the computer-based assessment." The same seven-point Likert-type scale was used as was used by Terzis and Economides (2011a). The scale points ranged from one to seven: *strongly disagree* (1), *disagree*, *somewhat disagree*, *neutral*, *somewhat agree*, *agree*, *strongly agree* (7). Finally,

because of the importance of the background information of the participants, some demographic items were created to gather information on nationality, age, gender, education level, department, and so on. Those items were presented at the beginning of the Computerized ToEFL Acceptance Questionnaire.

At a minimum two observed variables were used to represent each of the latent variables to increase the reliability of the latent constructs. Again, the latent constructs for this study were perceived playfulness, perceived usefulness, perceived ease of use, computer self-efficacy, social influence, facilitating conditions, goal expectancy, content of examination and behavioral intention to use. Similar to the CBAAM questionnaire (Terzis & Economides, 2011a), the four items for *perceived playfulness* (PP) were developed based on two studies one by Moon and Kim (2001) and the other by Wang et al. (2009). Moreover, three items for *perceived usefulness* (PU) and three items for *perceived ease of use* (PEOU) were adopted from Davis (1989). For *computer self-efficacy* (CSE), four items were adapted from Compeau and Higgins (1995). Four items from the unified theory of acceptance and use of technology (UTUAT) were adopted for *social influence* (SI) construct (Venkatesh et al., 2003). For *facilitating conditions* (FC), I preferred two items from Thomson, Higgins, and Howell's study (1991). The *content of examination* and *goal expectancy* (CE and GE) constructs were developed by Terzis and Economides (2011a) and these new constructs were measured using four and three items, respectively. Finally, *behavioral intention to use* (BIU) was measured by three items adopted from Davis (1989) (see Appendix, for the complete scale).

Validity and Reliability of Instrument

In order to evaluate the content validity of the survey, a field study was conducted. Boudreau, Gefen, and Straub (2001) and Straub (1989) suggested using field study to establish

the face and content validity of survey instruments. Based on the recommendations of Yun and Ulrich (2002), the field test was conducted with four international students in the U.S., all of whom took a computerized ToEFL exam. Four participants from different countries were selected to determine if different languages and cultural backgrounds had effects on the understanding of the survey. The four participants for the field test were from China, Turkey, Saudi Arabia, and Korea. The purpose of the field test was to determine the ease of delivering and accessing the survey and to establish if the proposed respondents would have difficulty understanding the survey items as well as the format of the questionnaire.

The field test was administered via e-mail. The objectives and directions of the field test requested the respondents to answer the some questions. For example: “Were you able to access the survey without difficulty?”, “Is the content of the questionnaire appropriate for the audience?”, “Are the items in the survey clear?”, “Do the instructions make sense?”, “Are any of the survey items intrusive, invasive, potentially embarrassing, or of a sensitive nature?” and “Do you have any other comments?”

Following the field test, face-to-face or online interviews with each of the students were conducted. The follow-up communication revealed some concerns regarding the survey instrument. Some participants considered certain survey questions redundant, suggested clarification of the instructions and rewording of certain questions. Based on the feedback, three word and sequence changes were made to the demographic questions. For example, the question 10, “Have you ever taken a ToEFL IBT?” was moved and appeared before than the question about other types of ToEFL. Moreover, a logic to the answer of question 10 was added because depends on the response examinees can continue or not to the survey. Identified technical

problems were also addressed and fixed based on and the respondents' experiences with the survey.

Reliability refers to the consistency or dependability of responses. In order to assess the reliability of the questionnaire, a pilot study was conducted with 10 students randomly chosen from those who agreed to participate in the study. Each participant completed two administrations of the questionnaire on two separate occasions, four weeks apart. The test-retest reliability for acceptance measures as follows: perceived playfulness, $r = .921$; perceived usefulness, $r = .943$; perceived ease of use, $r = .890$; computer self-efficacy, $r = .824$; social influence, $r = .878$; facilitating conditions, $r = .866$; goal expectancy, $r = .854$, content of examination, $r = .865$; and behavioral intention to use, $r = .901$.

Ethical and Security Concerns

Ethical and security concerns were also considered during the study. Even though this study did not involve greater than minimal risk to participants, the online questionnaire was distributed to the graduate and undergraduate international students in the U.S. after IRB review and approval. Permission from students was obtained before they participated in the study. To respect the rights of participants, all participants could withdraw from the study at any time without consequence. To protect the anonymity of participants and keep participants' identity confidential, names of participants were not collected. Although basic demographic information was obtained, the respondents were not asked questions that could divulge their identity. The participants gave their e-mail addresses in a separate online survey from those who wished to participate in the Amazon gift card drawing. Throughout the data collection phase, all data were viewed as confidential and were not shared with other participants or individuals outside of the

study. After data collection, all digital data was password protected and only accessible to the researcher.

Procedures

The survey was administered and the answers recorded via a professional survey site, Qualtrics. Qualtrics was appropriate for this web-based survey because it was convenient for respondents and it had automated management and data compilation. Qualtrics mailer was designed to send multiple emails that include a survey link and a specific message to the recipients at the same time. This feature was used to help distribute the survey to multiple recipients quickly through direct e-mail. An embedded website pop-up that linked the survey to social network webpage was posted. The e-mails and web site links, described the purpose of the survey, asked participants to complete the questionnaire on a separate web site, and assured them that their responses would be confidential and would not affect their ToEFL scores. To increase the response rate I offered participants an opportunity to win a \$10 Amazon.com gift card.

International students received information about the survey by email or through a social media webpage. All participants had to read the informed letter and choose if they wanted to continue with the study or not when they clicked the survey link. The participants then completed the background questionnaire. The background questions always appeared first. The 30 items in the CTAQ were randomized for each subject. After completing the questionnaire, the participants received an end of survey note that provided them with a link to join the Amazon gift card drawing. Finally, they were thanked for their participation. The survey took approximately 10 to 15 minutes to complete for each participant.

Data Analysis

After conducting the survey, the data were analyzed. The purpose of the data analysis was to understand and interpret the responses in order to answer the research questions. There were several steps employed to analyze the data.

Because the data analyzed were the scores obtained from the computerized ToEFL acceptance questionnaire, each response from the online questionnaire was saved into the same .csv file on a server. The saved responses were stored into a Microsoft EXCEL 2007 workbook to be prepared for analysis. Descriptive statistics were used to describe data using the Microsoft EXCEL 2007 program. The hypothesized study model was tested using structural equation modeling (SEM) via Mplus 7.1 (Muthen & Muthen, 2011).

SEM can estimate multiple interrelated dependent relationships that are either limited or unavailable in other multivariate analysis techniques (Hair, Anderson, Tatham, & William, 1998). The purpose of an SEM approach is to ensure the consistency of the model with the data and to estimate effects among the constructs. SEM also has the capability to analyze concepts that are unobservable, such as the nine latent variables contained in the research model. In this study, an advantage of SEM is the ability to overcome the limitation of measurement errors. This feature made SEM a powerful alternative to path analysis with regard to assessing and controlling measurement errors (Kline, 2005). By controlling the measurement error, unbiased estimates of relationships are possible with SEM. Other advantages of SEM include better estimates of path coefficients, ability to estimate both direct and indirect effects, and testable models. Due to these benefits, a SEM model was used in the current study. Estimated path coefficients were used to explore which variables had significant effects, and data-model fit indices were examined to test if the SEM model fit the data well nor not.

Standard SEM analysis steps were employed based on principles recommended by Tate (1996). First, model specification was done based on theory, experience, and the literature. The hypothesized model was specified to consist of a network of direct causal links among the variables. Next, model identification was conducted to determine if there was sufficient information (i.e., an adequate number of observed variances and covariances) to allow for the estimation of all of the model parameters. The t-rule was used to evaluate model identification. The t-rule simply states that there must be more observed sample means, variances, and covariances than there are parameters to be estimated (Bollen, 1989a). After passing the t-rule, SEM was used to evaluate the full study model, including an assessment of the fit of the model to the data. If the model was unacceptable, one or more revisions of the model based on theoretical credibility would have been considered. If a theoretically credible model with acceptable fit was obtained, the associated estimated effects: direct, indirect, and total causal could be described.

This study used maximum likelihood estimation with robust standard errors (i.e., MLR) to test the overall model fit, and several indices of model fit were generated. MLR is both widely-used and efficient when the multivariate normality assumption is not met (Hair et al., 1998). The model fit was measured using Chi-square, the Standardized Root Mean Square Residual (SRMR), the Root Mean Square Error of Approximation (RMSEA), the Tucker Lewis Index (TLI), and the Comparative Fit Index (CFI). The descriptions of these fit indices are provided below.

The Chi-square/degrees of freedom index is the ratio of Chi-square statistics to sample size. A value of less than 3.0 is typically considered to be minimally acceptable (Chin & Todd, 1995; Kline, 1998). Because of the limited use of Chi-square/degrees of freedom index, the *p*-

value of the Chi-square was considered an acceptable model-fit index. The Root Mean Squared Error of Approximation (RMSEA) is another index to test model fit; it was developed by Steiger (1990). Because this measure is based on the analysis of residuals, smaller index values are better. Specifically, values below .05 are interpreted as good fit; while values up to .08 represent an acceptable fit of the model to the data (Tate, 1996). The standardized root mean squared residual (SRMR) index is also a model-fit index; it is a standardized summary of the average covariance residuals. Covariance residuals are the difference between the observed and model-implied covariance matrix. SRMR has a lower boundary of zero and an upper boundary of one. Values below .05 indicate a good model fit to the data (Kelloway, 1998). Finally, the Comparative Fit Index (CFI) was also used in the current study. The CFI index was proposed by Bentler (1990) and it evaluates the level of improvement of the proposed model as compared to an independence model, with a correction for model complexity. The CFI also ranges from 0 to 1, with acceptable fit values that exceeded .95 (Hu & Bentler, 1999).

I conducted a multivariate normality test to test for skewness and kurtosis in the data. The multivariate Shapiro-Wilks test reported that $W = 0.8087$, $p < .001$. These results showed that the data were not multivariate normally distributed.

Multigroup SEM (MGSEM) analyses (Tate, 1996) were planned to determine the moderating effects of gender and developed status in CTAM. A multigroup SEM is “an SEM extension that permits the comparison of models over multiple populations or groups” (Tate, 1996, p. 219). The main focus of the multigroup analysis was to identify the differences in path coefficients between groups (Kline, 1998). To achieve that purpose, invariance tests across groups were tested using MGSEM (Joreskog & Sorbom, 1993).

Multigroup SEM invariance testing consists of two kinds of invariance: measurement invariance and structural invariance. Generally, the measurement model is tested first. Once that measurement invariance across groups has been established, substantive cross-group comparisons could be conducted to test the structural invariance.

Measurement invariance is tested to ensure that the observed scale items and theoretical constructs of the study function similarly for all groups. In other words, measurement invariance shows that observed variables measure latent variables that are similar for all subgroups. If measurement invariance does not hold, analyses of the corresponding measures would not produce meaningful results and the results could not be interpreted validly. Measurement invariance is tested at different levels such as configural, weak factorial, strong factorial, and strict factorial (Bollen, 1989; Byrne, 2012; Byrne, Shavelson, & Muthén, 1989; Cheung & Rensvold, 2002; Little, 1997; Meredith, 1993; Vandenberg & Lance, 2000). A brief explanation of each is provided below.

The first level of invariance, configural invariance, is a model with no constraint imposed on any parameter across groups. Configural invariance is a necessary condition for testing invariance of measurement parameters that pertain to the measurement characteristics of the observed items. In other words, if configural invariance is not demonstrated, it indicates that the observed items of the scale measure different constructs for different groups. Specifically, if the model shows that there are the same number of factors and the same patterns of free and fixed factor loadings across groups, configural invariance is supported. In order to establish weak invariance model, configural invariance model should be demonstrated first. Weak measurement invariance is defined as invariance of factor loadings. Weak factorial invariance demonstrates if the factors are scaled in the same units. Whether the Chi-square difference test result shows that

two nested models (i.e., configural and weak factorial) was not statistically different, then the equivalence of covariance and mean structures could be tested across groups. If statistically significant differences between configural and weak factorial models were present then constraints on factor loadings that caused the lack of fit would have been removed, not simultaneously but one at a time, until partial invariance of factor loadings was established. The next invariance testing is strong factorial invariance. It is met if the factor loading matrices and intercept factors are equal. More specifically, if the strong factorial invariance test shows significant results, then the factors are measured on the same units in each group and have the same origin. Lastly, invariance of covariance structures of measurement errors, with or without involving the mean structures, could also have been examined, which is called the strict factorial invariance model (Kline, 1998).

As discussed, measurement invariance is a prerequisite condition to test structural invariance. After measurement invariance is demonstrated, researchers could test the invariance of structural parameters, such as structural path coefficients, covariances, and factor means across groups. The steps of testing structural invariance are: testing invariance of factor variance, testing invariance of factor covariance, and testing factor mean invariance. First, the variance of a factor is a measure of dispersion of the factor scores (Vandenberg & Lance, 2000). Second, when factor variance invariance holds, factor covariance invariance is equivalent to factor correlation invariance across groups. Third, an important aspect of multigroup SEM is the ability to test factor mean differences between groups. To estimate the differences in factor means, one of the groups was treated as the reference group (Byrne, 2001). This was tested by fixing the factor mean to zero in the reference group and freeing the factor mean in other groups and comparing these groups.

At each measurement and structural invariance testing step, the model Chi-square statistics between restricted and unrestricted models were used to conduct the Likelihood Ratio test for model comparison. If Chi-square statistics of the model did not change significantly after imposing the restrictions, then the corresponding hypothesis of parameter invariance was retained. In addition to Chi-square difference test, change in CFI has become increasingly used to evaluate invariance in MGSEM. A difference in CFI values of less than or equal to .01 between nested models is considered a criterion of invariance. A CFI larger than .01 indicates meaningful change in the model fit for testing invariance (Cheung & Rensvold, 2002).

In summary, an MLR approach using Mplus 7.1 (Muthen & Muthen, 2012) for multigroup structural equation analysis was used to analyze both the measurement and structural model. Multigroup structural equation modeling (MGSEM) was used to test a model where behavioral intention to use was predicted by students' characteristics. Those characteristics included gender and developed status of countries. This analysis was constructed in two parts. First was the developed status effect analysis followed by the gender effect analysis.

Developed status analysis. Research question one focused on potential developed status differences on international students' behavioral intention to use and other crucial variables regarding computerized ToEFL acceptance. To address the first research question, multigroup SEM analysis for students from developing and developed countries was used. The number of unique variances, covariance and means, and then degrees of freedom for overall tests of model fit were calculated for model identification. In addition, sufficient within-group sample sizes to obtain stable estimates of the within-group moments were needed. However, there were not enough participants from developed countries to provide sufficient within-group sample size. As a result, multigroup SEM could not be conducted for developed status difference.

To investigate factor mean differences, Multiple Indicators and Multiple Causes (MIMIC) model was conducted. MIMIC model is another multigroup analysis in SEM. Because measurement invariance could not be tested due to the small sample size, MIMIC model was the only option. A MIMIC model has been defined as confirmatory factor analysis with covariates (Kline, 2005).

Gender effect analysis. The second research question concerned gender differences on international students' intention to use and other variables in the computerized ToEFL acceptance model. To answer the research question, the same types of analyses were performed as were conducted for research question 1, which asked: "Are there differences between students from developed and developing countries on their behavioral intention to use, perceived usefulness, perceived ease of use, and perceived playfulness that affect acceptance of computerized ToEFL? " In a multigroup SEM, models were fit simultaneously to the data from two or more groups. The number of unique variances, covariance, and means were calculated. Then degrees of freedom for overall tests of model fit were calculated to investigate the model identification. For model identification, the SEM fit within female and male groups was considered. Identification of each group was a necessary but not sufficient condition.

The Chi-square model fit index was a way to evaluate data-model fit and it was more sensitive to misfit in mean structure than covariance structure (Hu & Bentler, 1999). Given simultaneous estimation, testing of the parameter differences over groups was possible with invariance tests. These tests were done via the imposition of equality constraints on the model. More specifically, to test the invariance of two groups, parameter estimates should be held equal over female and male participants. Assuming no other differences, the models with and without

equality constraints were nested and Chi-square differences test could be applied to evaluate group differences (Bollen, 1989a).

Sufficient within-group sample size indicated that multigroup SEM (MGSEM) could be conducted. Thus, to evaluate the gender developed status effect about the latent variables, measurement invariance needed to be confirmed. In a MGSEM, the researchers should determine if the measurement model was equal over groups. This equivalence then indicates that there was no test bias and incomparability of scores over female and male groups (Bollen, 1989a). Testing of measurement invariance of the CTAM was conducted step by step. I started with the least strict form, configural invariance, and moved to weak factorial invariance. Configural invariance means that the two groups, female and male, have the same basic conceptualization of computerized ToEFL acceptance, without restricting any of the nonfixed parameters. A poor fit would signify that it made little sense to move to the more restrictive hypotheses. If the model fit indices were acceptance, moving to weak factorial invariance was appropriate. Weak factorial invariance was provided if the female and male group factor loading matrices were equivalent. These factor loadings should be equal in pattern and in the values of each factor loading. If weak factorial invariance held, then the factors (e.g., perceived usefulness) were measured in the same metric in developed and developing countries groups.

The most restricted invariance was strict factorial invariance which indicates that the measurement models for the groups were equivalent. Hence, if strict factorial invariance was met, factor loading matrices, intercept vectors, and residual variances of female and male groups were equal. If this structural equation model matched the data under this highly restrictive hypothesis, the results were consistent with the assumption that the same model operated in both groups.

After evaluating measurement invariance, the structural invariance for the gender effect should be examined. Evaluation of structural invariance was conceptually the same as the testing of moderation effects.

In order to find the pattern of differences on the parameters of the model, an overall test on a parameter among the groups was conducted using a Chi-square difference test. The results of the test indicated whether the two nested models were statistically significantly different or not. A statistically non-significant p -value of Chi-square difference test was desired to continue invariance testing.

CHAPTER 4: RESULTS

In this chapter, I first present a description of structural equation modeling (SEM) analyses and the computerized ToEFL acceptance model (CTAM). After that, a summary of the participant demographics is provided. Next, for comparison across groups (i.e., developed and developing countries), I conducted a multiple indicator and multiple causes (MIMIC) analysis. For gender difference analysis, I examined: (1) whether the pattern of fixed and nonfixed parameters were similar in the two gender groups (i.e., configural invariance), (2) whether the rating scales were treated similarly in different gender (i.e., metric invariance), and (3) whether the female and male groups have the same item intercepts (i.e., strong invariance). I then examined differences in the SEMs across these groups.

Data Preparation

The data from the 237 returned questionnaire were imported into an Excel worksheet and a random 10% (25) were checked for accuracy of data entry; no errors were found. No missing values for any of the model variables were found as the questionnaire system forced respondents to answer of all questions before it could be submitted.

Descriptive Statistics

The analysis of the sample data included the descriptive statistics of the respondents. The information consisted of gender, years of school experience, level of education, home country, and age of the participants. The age of students ranged from 18 to 44 (see Table 3).

Table 3

Sample Characteristics: Age

	Participant	<i>M</i>	Minimum	Maximum
Age	237	27.48	18	44

The gender breakdown, with more males than females and reported age of the study population (see Table 3), is not similar to those of previous information technology adoption studies of nonphysician health professionals (Terzis & Economides, 2011b). There were 52.3% ($n = 124$) men and 47.7% ($n = 113$) women (see Table 4). Based on the World Bank's definition, South Korea, Taiwan, Canada, France, Italy, Germany, Hong Kong, Singapore, Japan, Sweden, Finland, Denmark, and Portugal were considered developed countries, and Turkey, China, Colombia, Paraguay, Thailand, Russia, Tanzania, Vietnam, Mexico, Ukraine, Serbia, Indonesia, Puerto Rico, Malaysia, Romania, India, Brazil, and Chile were categorized as developing countries in this study. The data show that there were 57 people from developed countries and 180 students from developing countries who participated in the study (see Table 5). Participation was spread over numerous schools including engineering, education, economics and psychology, and different educational level such as undergraduate and doctorate degree (see Table 6 and Table 7).

Table 4

Sample Characteristics: Gender

Gender	Frequency	Percent
Male	124	52.3
Female	113	47.7

Table 5

Sample Characteristics: Developed Status

Developed Status	Frequency	Percent
Developed	57	24.1
Developing	180	75.9

Table 6

Sample Characteristics: Educational Level

Educational Level	Frequency	Percent
Bachelor's degree	37	15.6
Masters' degree	91	38.4
Doctoral degree	99	41.8
Professional degree	6	2.5
Post-doc	4	1.7

Table 7

Sample Characteristics: Discipline

Discipline	Frequency	Percent
Science	87	36.8
Social Science	60	25.3
Education	29	12.2
Engineering	24	10.1
Medicine	16	6.7
Business	16	6.7
Law	4	1.8
Dentistry	1	0.4

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) enables one to test whether the data support the proposed factor model (Kline, 2005). Researchers are able to fit the data to see the problems in the proposed model structure. Moreover, CFA provides evidence that can be used to modify the model to improve data-model fit. As a result, a CFA was conducted before conducting an MGSEM analysis.

CFA model for full data. Before conducting multigroup analysis of measurement and structural invariance, I examined model-data fit and parameter estimates for the entire sample ($n = 237$). For the full model, the fit indices revealed unacceptable data-model fit. The Chi-square test of model fit was 848.351 with 369 degrees of freedom and p -value was $< .001$. In addition to the Chi-square statistic, the Root Mean Square Error of Approximation (RMSEA) estimation

indicated poor data-model fit. The RMSEA value was .074 with a 90% confidence interval of .067 to .081. According to Hu and Bentler (1999), a cut-off value close to .06 for RMSEA is evidence of acceptable data-model fit. Also, a stringent upper limit of .07 (Steiger, 2007) seems to be the general consensus amongst experts in this area. The Comparative Fit Index (CFI) was .871 and Tucker Lewis index (TLI) was .847; the CFI and TLI values of .95 or higher are generally considered evidence of acceptable data-model fit. Because the values that were found were below .95, the CFI and TLI fit indices did not indicate a good model fit (Hu & Bentler, 1999) (see Table 8). Finally, an acceptable model fit should have a Standardized Root Mean Square Residual (SRMR) smaller than .08 (Hu & Bentler, 1999). For these data, the value of SRMR was .077, indicating acceptable data-model fit (see Table 8). Taking Hu and Bentler's (1999) guidelines into consideration, overall these fit indices (i.e., Chi-square, RMSEA, TLI, CFI, and SRMR) fell short of the recommended cut-off values. This led me to consider Lagrangian multiplier tests (i.e., modification indices). Modification indices show the degree of incremental improvement to data-model fit if additional parameters are included in the model. Because the initial model indicated an unacceptable model fit to the data, it is common practice to consider possible model revisions based on the data in the study. Such model revisions usually consist of the addition of one or more paths to the initial model.

Although Mplus 7.1 (Muthen & Muthen, 2012) offered several options for model modification, only five of the suggested model revisions were considered: 1) adding a covariance between the measurement errors of Question 12 (i.e., "People who are important to me think that I should use internet-based ToEFL") and Question 11 (i.e., "People who influence my behavior think that I should use the internet-based ToEFL"); 2) adding a covariance between the measurement errors of Question 19 (i.e., "Internet based ToEFL's questions were relative to the

course's that I took in the university") and Question 20 (i.e., "Internet based ToEFL's questions were useful for my higher education"); 3) adding a path from the PU latent variable to Question 20; 4) adding a path from the PP latent variable to Question 20; and 5) adding a path from the GE latent variable to Question 20.

Modified full CFA model. Based on the modification indices, a modified the full CFA model that included the path between PU variable and Question 20 was analyzed. However, in the revised model, the maximum number of iterations was exceeded; hence, the model fit indices could not be calculated. The same results were observed from the other path additions. Also other modified models were also examined via CFA and get unacceptable fit indices. As a result of these unacceptable results, various combinations of modifications were tries until the best fitting model that would converge was obtained. Modified full CFA model included the covariances of Question 12 and Question 11, Question 15 and Question 13, Question 20 and Question 19, Question 22 and Question 18, Question 7 and Question 4, Question 14 and Question 13, Question 19 and Question 15, and Question 24 and Question 20. A new modified CFA model for the full sample was analyzed. The fit indices improved but still did not yield acceptable data-model fit. For this modified full CFA model, the Chi-square value was 758.656 with 361 degrees of freedom, which was statistically significant. Moreover, the RMSEA was 0.068; the CFI and TLI values were 0.89 and 0.871, respectively; and the SRMR showed a poor fit with a 0.100 value (see Table 8). Because the results did not support the idea that this model is appropriate for the data, I decided to reconsider the model.

The data-model fit indices for the modified CFA model showed that a model with nine latent variables is likely too complex to be estimated using this limited sample size. Hence, a reduced form of the proposed model was selected. The new model included only four latent

variables: perceived playfulness, perceived usefulness, perceived ease of use and behavioral intention to use. The revised model is shown in Figure 10.

Table 8

Results of the Model Fit Tests for Full Dataset

Model	Chi-square	df	p-value	RMSEA	CFI	TLI	SRMR
CFA for full data	848.351	369	< .001	.074	.871	.847	.077
Modified full CFA model	758.656	361	< .001	.068	.890	.871	1.00
CFA for new model	105.061	59	.0002	.057	.973	.964	.047
Modified CFA for new model	84.317	57	.0108	.045	.984	.978	.033
New SEM	84.167	58	.0140	.044	.985	.979	.033

Note. *df* = degrees of freedom. RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker Lewis index. SRMR = standardized root mean square residual.

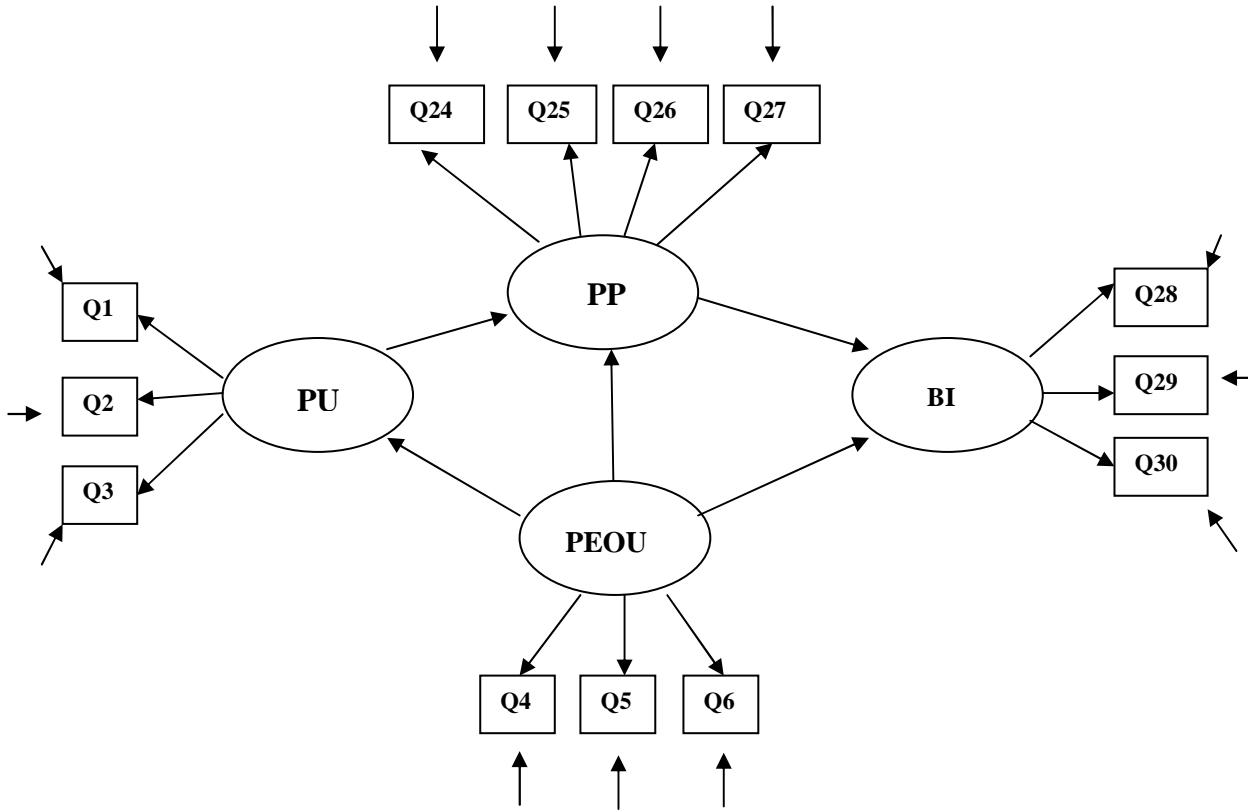


Figure 10. Revised computerized ToEFL acceptance model. Adapted from “The Acceptance and Use of Computer Based Assessment,” by V. Terzis and A. A. Economides, 2011a. *Journal of Computers & Education*, 56(4), p. 1034. Copyright 2011 by the Elsevier.

The relationships between those latent variables remained the same as was proposed in the previous model. In other words, perceived usefulness had a direct effect on perceived playfulness, perceived ease of use had a direct relationship with perceived playfulness and usefulness, and perceived playfulness and perceived ease of use had a direct effect on behavioral intention to use. Similar to the proposed model, a confirmatory factor analysis was performed before conducting the SEM analysis.

CFA for new model. The new model with four latent variables was analyzed by CFA. The results indicated a moderate fit. The Chi-square value was 105.061 with 59 degrees of

freedom, which is statistically significant. However, the RMSEA estimation was 0.057, which was lower than the stringent upper limit of 0.07 (Steiger, 2007). The CFI and TLI values were also acceptable with values of 0.973 and 0.964, respectively. Finally, the SRMR value was 0.047, which indicated acceptable fit (see Table 8). To make some improvement in the model fit, the modification indices were taken into consideration. The modification indices indicated two covariances to add to the model (i.e., covariances between Q6 and Q4, and between Q29 and Q6).

Modified CFA for new model. The output of the modified new model showed an acceptable fit. Even though the Chi-square test was statistically significant, all other indices supported the model fit (see Table 8). As a result, it was determined that the modified new model was an acceptable model for the data that was collected for this study. Before testing for measurement invariance between genders and countries, an SEM analysis for the new model, with the full sample, was performed.

Multigroup Structural Equation Modeling Analysis

The same fit statistics (i.e., Chi-square, RMSEA, CFI, TLI and SRMR) were used to evaluate the data-model fit of the multigroup SEM models. Unlike many single-group SEMs, multigroup SEMs often include a restricted mean structure. It is important to be aware that the Chi-square and many other model fit statistics are more sensitive to misfit in mean structure than covariance structure. Given simultaneous estimation, it is possible to formally test whether parameters are the same or different across groups. These tests are done via the imposition of equality constraints on the model. It is important to impose equality constraints on raw, not standardized estimates (Kline, 2005).

New structural equation model. As described, the computerized ToEFL acceptance model (CTAM) was modified and only four latent variables were included in the new model. Based on these changes, SEM analysis with the full data was conducted. Despite a statistically significant Chi-square value the fit indices showed an acceptable fit (see Table 8). However, the main analyses of interest involved comparisons of structural paths across gender and developed/developing country status.

Test of configural model for developed status. For developed status of countries, the configural model was tested. There were 57 participants from developed countries and 180 subjects from developing countries in the study. Unfortunately, 57 participants were not enough to calculate a multigroup SEM for the current model. The Mplus 7.1 (Muthen & Muthen, 2012) output provided an error that indicated there were more parameters than the sample size in developed country group (see Table 9). Given these results, a group difference test using SEM was performed which allows analysis with small sample size.

Test of MIMIC model for developed status. For the final CTAM model, a MIMIC analysis was conducted. The goodness of fit indices showed acceptable data-model fit, except for the Chi-square test (see Table 9).

Table 9

Results of the Model Fit Tests for Developed Status

Model	Chi-square	df	p-value	RMSEA	CFI	TLI	SRMR
Configural invariance for country	--	--	--	--	--	--	--
MIMIC for country	111.961	66	.0004	.054	.974	.964	.037

Note. *df* = degrees of freedom. RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker Lewis index. SRMR = standardized root mean square residual.

When investigating the effect of developed status on the four latent factors, developed status was a significant predictor of the PP, PU, PEOU, and BI. To answer the question about differences in factor means, I interpreted the unstandardized parameter estimates between developed status (dummy coded variable) and PU of .58, between developed status and PP of .596, between developed status and PEOU of .349, and between developed status and BI of .840 (see Table 10). In the data file, developed countries were coded as 0 and developing countries were coded as 1. The significant direct effect of developed/developing country status on the latent variables (e.g., PP, PU, PEOU, and BI) indicated that, on average, relative to participants from developed countries, participants from developing countries were significantly higher on perceived usefulness, perceived playfulness, perceived ease of use, and behavioral intention to use (see Figure 11).

Table 10

MIMIC Model Estimates for Developed Status

	Estimate	S.E.	Est. /S.E.	<i>p</i> -value
PU on Country	0.580	0.258	2.251	.024
PEOU on Country	0.349	0.139	2.511	.012
PP on Country	0.596	0.220	2.708	.007
BI on Country	0.840	0.249	3.376	.001

Note. S.E. = standard errors of parameter estimate. Est. / S.E. = estimate divided by standard error. PU = perceived usefulness. PEOU = perceived ease of use. PP = perceived playfulness. BI = behavioral intention to use.

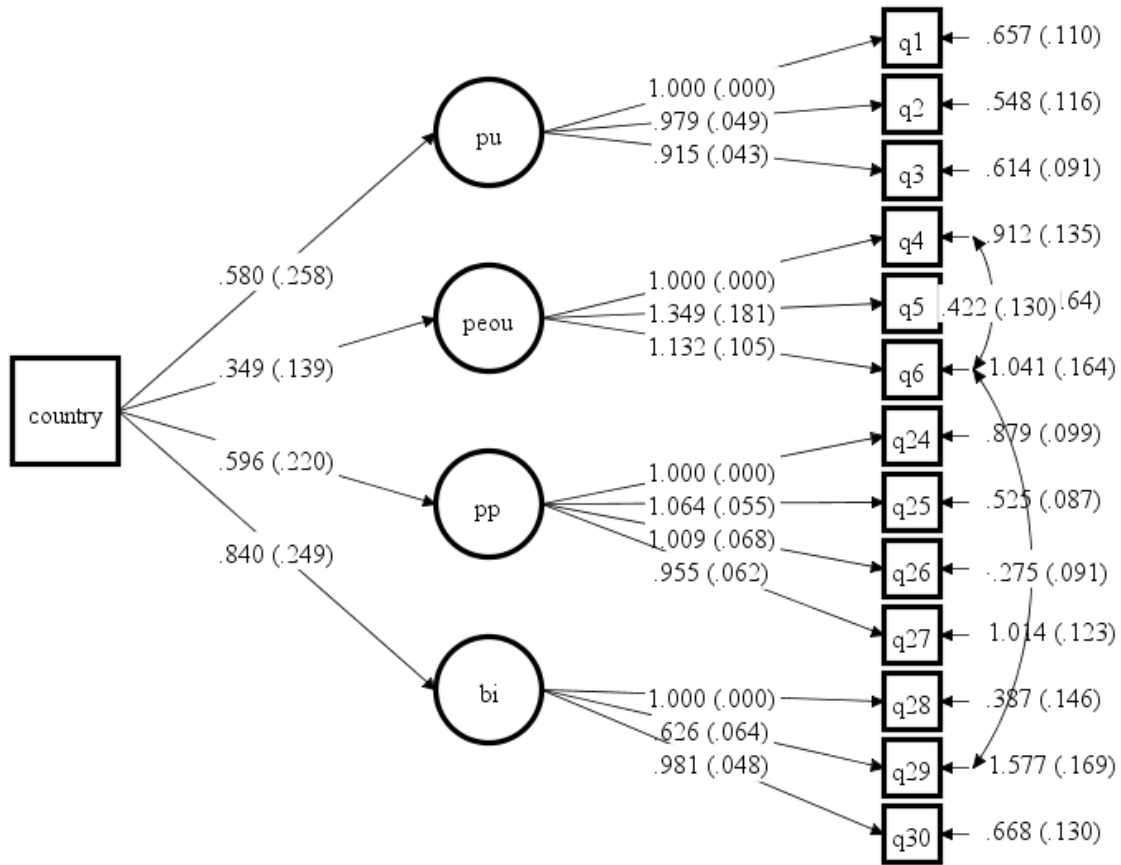


Figure 11. MIMIC model diagram for developed status difference.

Test of configural model for gender. When evaluating a latent variable model with multiple groups, it is important to determine the level of measurement invariance that is present. One way to do this is to start with the least restrictive model and move to more restrictive models as allowed by the data (i.e., from configural invariance to weak, and then strong invariance). In the current study, the configural model showed adequate fit. The CFI and TLI values were .959 and .944, respectively and SRMR was .053 which indicates good fit. However, the Chi-square test of model fit showed a 188.514 value, 114 degrees of freedom and was statistically significant. In addition, RMSEA had a .074 value (see Table 11). As a result, the configural invariance model fit the data reasonably well.

Table 11

Results of the Model Fit Tests for Gender

Model	Chi-square	df	p-value	RMSEA	CFI	TLI	SRMR
Configural invariance for gender	188.514	114	< .001	.074	.959	.944	.053
Weak factorial invariance for gender	206.090	123	< .001	.076	.954	.942	.073
MIMIC for gender	98.465	66	< .001	.046	.982	.975	.034

Note. *df* = degrees of freedom. RMSEA = root mean square error of approximation. CFI = comparative fit index. TLI = Tucker Lewis index. SRMR = standardized root mean square residual.

However, based solely on configural invariance, conclusions regarding group difference would be unique because the factor loadings and intercepts are free to vary across groups. To make inferences about group differences in the factor means, measurement invariance must be demonstrated.

Test of weak factorial invariance for gender. The next most restrictive assumption that could be made was that factor loadings were equivalent for the two groups (i.e., weak factorial invariance). The CFI, TLI, and RMSEA indicated the weak factorial invariance assumption was satisfied for this dataset (see Table 11). Furthermore, the weak factorial invariance model was nested within the configural invariance model. Hence, a Chi-square difference test was conducted to evaluate if the weak measurement invariance assumption should be retained. However, with MLR estimation the Chi-square difference test cannot be conducted without a special calculation (Kline, 2005).

The result of the Chi-square difference test, including the adjustment for MLR, was statistically significant, indicating that imposing equality constraints on the factor loadings did

cause a significant decrement in data-model fit. Thus, the weak invariance assumption was not retained and it was impossible to interpret differences in factor means or structural paths (see Table 12). As a result, further invariance tests (i.e., strong measurement invariance) could not be conducted. Based on the configural invariance model, the results showed that the two gender groups have the same number of latent variables and they were associated with the same manifest variables. However, any other comparisons were not possible with the results. Hence, the Multiple Indicators and Multiple Causes (MIMIC) model was used to complete the analysis.

Table 12

Result of X^2 difference test.

Model comparison	<i>df</i>	X^2 value	<i>p</i> -value	Decision
Test of X^2 difference test between configural invariance and weak factorial invariance	9	17.79	.0376	reject

Note. *df* = degrees of freedom. X^2 value = Chi-square difference test index.

Test of MIMIC model for gender. Multiple Indicators and Multiple Causes (MIMIC) models are a special case of structural equation models. The simplest model for evaluating group differences in SEM is the MIMIC model. In the MIMIC model, the influences of formative indicators on unobservable latent variables are assessed through their impact on the reflective indicators (Lester, 2008). The MIMIC model was constructed for gender difference and the results indicated an acceptable fit (see Table 11).

Interpreting the results of MIMIC model could be problematic because of the lack of measurement invariance, but short of finding measurement invariance, MIMIC model results are the next best option. Table 13 showed the MIMIC model estimations with unstandardized path coefficients for gender groups. These results indicate that gender did not have a statistically significant relationship with any of the latent variables in my modified model. Those

nonsignificant results indicated that men, on average, had equal scores on perceived usefulness, perceived playfulness, perceived ease of use, and behavioral intention to use with women (see Figure 12).

Table 13

MIMIC Model Estimates for Gender

	Estimate	S.E.	Est. /S.E.	<i>p</i> -value
PU on Gender	0.007	0.211	0.033	.973
PEOU on Gender	-0.191	0.139	-1.381	.167
PP on Gender	-0.057	0.178	-0.318	.751
BI on Gender	0.139	0.228	0.611	.541

Note. S.E. = standard errors of parameter estimate. Est. / S.E. = estimate divided by standard error.

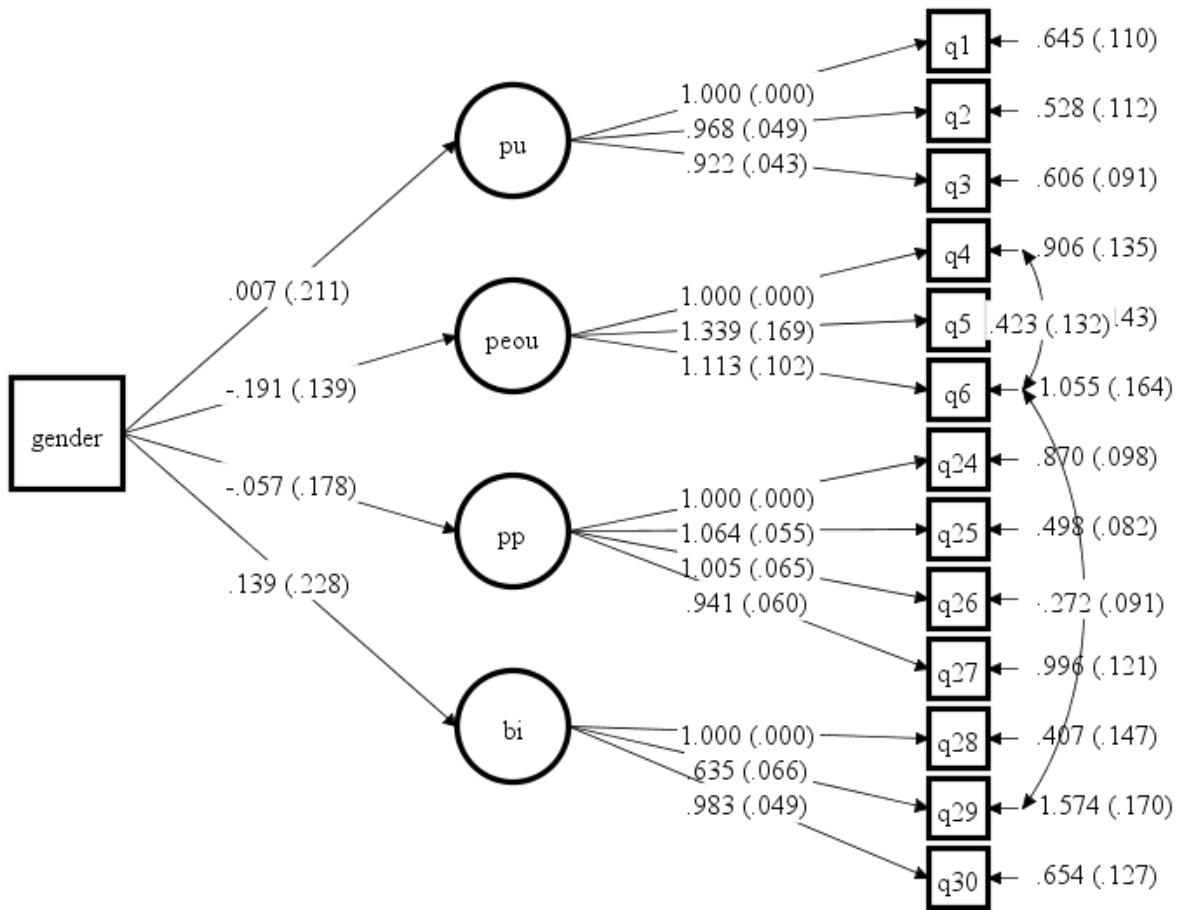


Figure 12. MIMIC model diagram for gender difference.

Summary

In summary, the analyses reported in this section addressed the two primary research questions: 1) Are there differences in behavioral intention to use and on other crucial variables (i.e., perceived playfulness, perceived usefulness, and perceived ease of use) regarding ToEFL IBT acceptance toward students from developed and developing countries? and 2) Are there differences between female international students' and male international students' behavioral intention to use ToEFL IBT acceptance?

The multigroup analysis method recommended by Chin (1998b) was utilized to examine the hypothesis of the moderating effects of gender in the research model. First, the multivariate normality test was conducted to assess the normality of the data. Shapiro-Wilks test revealed that the data was not normally distributed. Hence, maximum likelihood estimation with robust standard errors was used to analyze the multigroup SEM and the MIMIC models. Results of the preliminary confirmatory factor analysis showed that the initial computerized ToEFL exam acceptance model (CTAM) with nine latent variables was too complicated for the collected data, and a revised version of the CTAM that included only four latent variables was proposed. Under this model, weak factorial invariance was not supported by the data. These unexpected results indicated a need to conduct a MIMIC model for gender differences and developed/developing country status. Even though the MIMIC model fitted the data well, the results indicated there was no gender difference on factor means. Finally, for developed status difference, the MIMIC model indicated that participants from developing countries had higher latent factor means, on average, than participants from developed countries.

CHAPTER 5: DISCUSSION

This chapter begins with a review of the study summarizing the purpose, theoretical framework, and research problem. A summary of the research design, procedures, instrumentation, research questions, and associated hypotheses is provided and the results and discussion of the findings are presented. The final section describes study limitations, recommendations for additional research, and an overall summary.

Research Purpose and Framework

Over the past decades there has been a huge increase in the use of large scale computer-based assessment (e.g., ToEFL) (Terzis, & Economides, 2011a); however, little has been published to date on students' views of such computer-based assessments, specifically the ToEFL. Because most of the published work has been on the prevalence of computer anxiety among students, the intention behind the use of computers for the ToEFL examination has been open to question.

In general, the aim of this study was to gain an understanding of international students' intention to use the computerized ToEFL. To perform this task, I used a computer based assessment acceptance model (CBAAM), which was developed by Terzis and Economides (2011a). It was adapted to investigate if there were gender and country developed status effects on the acceptance model. The proposed model was called computerized ToEFL acceptance model (CTAM) and it contained nine latent variables with 30 observed variables. Based on the CBAAM, the 9 latent variables in the CTAM had specific relationships with each other. This study tried to examine group differences on the nine latent variables. The first research question

focused on potential differences in acceptance for computerized ToEFL between participants from developed countries and participants from developing. The second research question asked if there was a difference between female international students' and male international students' behavioral intention to use ToEFL IBT.

Research Design

Although numerous studies on technology acceptance have been conducted, few have performed a multigroup invariance analysis. For example, Deng, Doll, Hendrickson, and Scazzero (2005) analyzed the structural invariance of the TAM across four subgroups. Data from software application users was collected to examine the relationship among perceived usefulness, perceived ease of use, and behavioral intention to use technology. In another multigroup study, Lai and Li (2005) examined the Internet banking usage. This study used TAM to investigate whether there was invariance across different subgroups: age, gender, and IT competence.

The computerized ToEFL acceptance model (CTAM) proposed in the current study incorporated eight latent variables from previous studies in order to predict Behavioral Intention to Use a computerized examination system. Specifically, it adopted perceived playfulness (PP), perceived usefulness (PU), perceived ease of use (PEOU), computer self-efficacy (CSE), facilitating conditions (FC), social influence (SI), goal expectancy (GE), and content of examination (CE) from the CBAAM (Terzis & Economides, 2011a). However, the data were not sufficient to estimate the full CTAM and modifications were needed. In order to achieve an interpretable solution, the CTAM was reduced to a four-factor model. In this new CTAM, the latent variables were perceived playfulness, perceived usefulness, perceived ease of use, and behavioral intention to use.

Discussion of Findings

Imposing equality constraints on the factor loadings across gender and developing/developed country status did cause a significant decrement in the data-model fit. Hence, I could not retain the weak invariance model and interpret difference in factor variances and covariances. Therefore, a Multiple Indicator and Multiple Cause (MIMIC) model was examined to detect possible gender differences in latent factor means. The results of MIMIC model for gender revealed that there was no statistically significant difference between female and male computerized ToEFL acceptance. These results were consistent with some empirical research in the information technology literature (e.g., Cuadrado-García, Ruiz-Molina, & Montoro-Pons, 2010; Hung, Chou, Chen, & Own, 2010). Cuadrado-García, Ruiz-Molina, and Montoro-Pons examined college students' interaction with e-learning. They demonstrated that female and male students used e-learning in same way and their motivations to use such technology was similar. Furthermore, Hung, Chou, Chen, and Own conducted a confirmatory factor analysis to explore a multidimensional instrument for online learning readiness. The results revealed that there was no significant gender effect on the factors of online learning readiness.

The results of hypothesis testing are summarized in Table 14. In this table n/a symbolized the test of the hypothesis was not able to run. The table shows that H1, H3, H5, H8, H11, and H15 hypotheses are rejected after the conclusions of results. H1 indicated that perceived playfulness would be higher for men than for women; however, the results of MIMIC model for gender analysis did not provide results to accept the hypothesis. The third hypothesis, H3, referred that perceived playfulness would be higher for persons from developed countries than for those from developing countries and I could not find the results to support the hypothesis.

Moreover, the results of the study did not accept the H5 hypothesis which was perceived usefulness would be higher for men than for women. H8 hypothesized that perceived usefulness would be higher for developed countries than for developing countries and the results did not support to accept it. Another hypothesis, H11, was about perceived ease of use and gender difference. H11 indicated that perceived ease of use would be higher for men than for women. Finally, H15, perceived ease of use would be higher for developed countries than for developing countries, was not supported by the data of the current study.

Table 14

Summary of Hypothesis Testing Results

Hypothesis Number	Variable or Relationship	Hypothesized Direction of Effect	Decision
H1	Perceived Playfulness	Men > Women	Rejected
H3	Perceived Playfulness	Developed > Developing	Rejected
H5	Perceived Usefulness	Men > Women	Rejected
H8	Perceived Usefulness	Developed > Developing	Rejected
H11	Perceived Ease of Use	Men > Women	Rejected
H15	Perceived Ease of Use	Developed > Developing	Rejected

Relationships

H2	PP → BI	Men > Women	n/a
H4	PP → BI	Developed > Developing	n/a
H6	PU → BI	Men > Women	n/a
H7	PU → PP	Men > Women	n/a
H9	PU → BI	Developed > Developing	n/a
H10	PU → PP	Developed > Developing	n/a
H12	PEOU → BI	Women > Men	n/a
H13	PEOU → PU	Women > Men	n/a
H14	PEOU → PP	Women > Men	n/a
H16	PEOU → BI	Developed > Developing	n/a
H17	PEOU → PU	Developed > Developing	n/a
H18	PEOU → PP	Developed > Developing	n/a

Note. PP = perceived playfulness. PU = perceived usefulness. PEOU = perceived ease of use. CSE = computer self-efficacy. SI = social influence. FC = facilitating conditions. GE = goal expectancy. CE = content of examination. BI = behavioral intention to use. n/a = no statistical analysis conducted for this hypothesis.

Second, a multigroup SEM was estimated to analyze developed status effect on acceptance. Because of the small sample size for developed country group, the model could not be estimated. Thus, the MIMIC model was selected to investigate differences between developed countries and developing countries.

The MIMIC model for developed status was conducted and the fit of this model was acceptable. Specifically, the findings showed that, relative to the students from developed countries, the students from developing countries students were had statistically significantly higher factor means, on average, on perceived usefulness, perceived ease of use, perceived playfulness, and behavioral intention to use (see Table 13). Unfortunately, all the hypotheses about the developed status that were made in the study were not supported by the data. In other words, perceived playfulness, perceived usefulness, perceived ease of use, and behavioral intention to use were higher for students from developing countries than developed countries. Since there have been no previous research studies that addressed these hypotheses, it was difficult to conclude that these results were consistent with the literature.

Implications

The results of this research yield several implications for educators and administrators. First, as perceived usefulness and perceived ease of use do not remain static and are subject to situational influences, examinees who perceive this type of technology to be useful and easy to use may soon experience limitations if they do not keep abreast with advances in the relevant technologies. For international examinees, computer-based testing training programs

could perhaps design and develop more relevant courses or training modules in order for these students to be able to continue developing their knowledge and skills in the use of emerging technologies for test taking.

In addition, the current study provided the practitioners (i.e., instructors or academic institutions) a salient guideline on the design and implementation of an Internet based ToEFL exam. Because there was a significant difference in the decision to use a computerized ToEFL between students from developed countries and students from developing countries, practitioners should pay attention to the external factors for the two groups during the design process. To encourage the adoption and use of a ToEFL IBT, the developer of the ToEFL (i.e., ETS) should emphasize the unique features of a computerized test system in facilitating the testing process. They could create differently to make it more appealing and engaging to the students.

Limitations

One of the major limitations in this study was small sample size. The sample size was relatively small ($n = 237$) to analyze a multigroup structural equation model. A larger sample with participation of more universities can provide different picture of the results.

Other limitation to this study was the subjectivity in measuring acceptance. Because the data were self-reported, it was difficult to make inferences (Davis, 1989). Davis (1989) emphasized the problems regarding how accurately self-reports reflect actual behavior. He claimed that the accuracy of self-reports were not proved and halo effects could make problem. Another limitation was the use of a questionnaire as the main data collection instrument. The length of the questionnaire might have been too long for some participants. As a result, some participants might have attempted to complete the survey quickly and thus produced superficial

responses. Also, some of the participants decided not to complete the survey, affecting the response rate.

The measure of gender as a dichotomous variable in this study is consistent with biological sex. However, some prior studies have suggested that gender may also be considered as a psychological construct with men and women not considered a biological concept (Bem, 1981). Future studies could investigate gender differences in computerized ToEFL acceptance based on femininity and masculinity to understand how students make decision to use a ToEFL IBT. The generalizability of the results of this study might be limited in some ways. The results might not be generalized beyond the present time because access to technology, especially computers, changes over time. Hence, the results might vary in the future.

This study was based on only quantitative data. Additional qualitative analyses could provide additional insights into the gender and developed status effects on computerized ToEFL acceptance. As a follow up to this research, some of the participants could be selected and interviewed to explore their perspective on ToEFL IBT. Finally, because the questionnaire used in this study was in a web-based format, the sample was likely to include primarily students who were familiar with computers. In other words, people who were access to computers and internet were more willing to participate in a web-based questionnaire. People who had low computer-self-efficacy tend to avoid extra work on computer which might lead to withdraw the questionnaire.

Future Directions

There are some recommendations for further research. First of all, the instrument used in the current research to assess students' behavioral intention to use was based on self-reported data. Whereas one study (Barnett, Kellemanns, Pearson, & Pearson, 2006) found a strong

correlation between the self-reported and computer-recorded usage, another study conducted by Straub, Limayem, and Karahanna (1995) found a low relationship between self-reported and computer-recorded measurements. Therefore, to get more concrete data on the acceptance of computerized ToEFL, more advanced computer-recorded measurements should be developed.

In addition, future research is needed to collect larger samples from different countries. Testing the model in this way provides evidence as to whether or not it is helpful in terms of generalizability. The purpose and goal of this research is to seek and aid the decision makers to find additional tools that support their effort to procure the most effective computer-based testing. The aim, again, would be to introduce improvements and upgrades that would support and enhance the testing and learning experience of a student. Finally, initial acceptance is only the first step toward students' overall success using a computerized ToEFL implementation. It may also be interesting to examine the students' behavior while they are using the computer-based assessment.

Summary

The purpose and goal of this research is to examine the acceptance differences on computerized ToEFL exam amongst the examinees that are from different backgrounds. The results showed that there were no gender or developed status differences on test takers' perceived playfulness, perceived usefulness, perceived ease of use and behavioral intention to use a computerized ToEFL. In other words, both female and male international students perceived playfulness, usefulness, and ease of use the ToEFL IBT in the same way. Moreover, international students from developed countries and developing countries reacted to the questionnaire similarly and there were no significant difference between them. Additional research on this

topic is warranted, and it is only through continued investigation that the research questions addressed here can be further clarified and the results more completely interpreted.

APPENDIX: COMPUTERIZED ToEFL ACCEPTANCE QUESTIONNAIRE

Information

I am Kubra Karakaya, a graduate student at the University of North Carolina-Chapel Hill, and the purpose of this note is to invite you to participate in an important study. As an international student in the U.S., I took the ToEFL exam several times and I became interested in how other international students' opinions about the ToEFL. The purpose of this study is to collect information about differences in international students' perceptions and attitudes about the internet-based (IBT) ToEFL exam.

Procedures

If you choose to participate in this research, you will be asked to complete some demographic questions and then a short questionnaire about your IBT ToEFL experiences. The general demographic information (for example, age, gender, nationality) will be collected so that we can accurately describe the general traits of the group of international college students who participate in the study. The questionnaire consists of 30 questions and will take approximately 15 minutes or less to complete. The questions are designed to determine why you choose to use internet-based ToEFL exam as well as your perceptions about internet based ToEFL.

Risks/Discomforts

No risks or discomfort are anticipated with taking part in this study. If you feel uncomfortable with a question, you can skip that question or withdraw from the study altogether. If you decide to quit at any time before you have finished the questionnaire, your answers will NOT be recorded.

Benefits/ Costs

You will be contributing to knowledge about international students' attitudes toward the IBT ToEFL exam and finding any potential differences between groups. If you choose to participate, you will be entered in a random drawing for one of two \$10.00 Amazon.com gift certificates (we anticipate that between 300 to 500 international students will participate in the study). After we have finished data collection, we will conduct the drawing. Winners will receive the gift certificate via e-mail.

How the Findings Will Be Used

The results of the study will be used for scholarly purposes only. The results will be presented in educational settings and at professional conferences, and may be published in a professional journal in the field of educational measurement.

Confidentiality

All data obtained from participants will be kept confidential and will only be reported in aggregate. All questionnaires will be kept secure and no one other than the primary investigator listed below will have access to them. The data collected will be stored in the computer with a password protected, secure database until it has been deleted by the primary investigator. The e-mails for Amazon gift card drawing will be stored separately from the survey data by directing participants to another survey web-link.

Compensation

There is no compensation or cost to participants for participation in this study.

Participation

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your academic status, GPA or standing with the university. If you desire to withdraw, please close your Internet browser and

notify the principal investigator at this email: (karakaya@live.unc.edu). Or, if you prefer, inform the principal investigator as you leave.

Questions about the Research

If you have questions regarding this study, you may contact the principal investigator, Kubra Karakaya, at 919-360-3492, karakaya@live.unc.edu.

Questions about your Rights as Research Participants

The University of North Carolina-Chapel Hill's Institutional Review Board has reviewed my request to conduct this project. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Office of Human Research Ethics by calling (919) 966-3113

Or by email: IRB_Subjects@unc.edu

Or access their website at <https://research.unc.edu/offices/human-research-ethics/index.htm>.

Routine questions about scheduling, explanation of procedures, or similar matters about your particular study should be addressed to the research investigator, Kubra Karakaya.

Date of IRB exemption: 10/19/2012

IRB no: 12-1981

I have read, understood, and printed a copy of, the above consent form and desire of my own free will to participate in this study.

Yes

No

Demographics

1. Please indicate your age.

2. Please indicate your gender.

- Male
- Female
- Other

3. Please indicate your nationality.

4. Which of the following best describes the area you lived in before you came to the USA?

- Urban
- Suburban
- Rural

5. How many years have you been living in the U.S.?

6. Please indicate your degree that you currently working on or recently completed.

- Bachelor's degree
- Master's degree
- Doctoral degree
- Professional degree (M.D., J.D. etc.)
- Post-doc

7. Please indicate your major/department that you are studying or recently studied.

8. Please indicate the university that you are currently enrolled in or earned your last degree.

9. Please indicate your last cumulative grade point average (if you have it).

10. Have you ever taken ToEFL IBT? (Required)

- Yes
- No

11. Please indicate your first scores on each of the test sections indicated.

_____ Reading

_____ Listening

_____ Speaking

_____ Writing

12. Have you taken other type of ToEFL before (Paper Based Test (PBT) or Computer Based Test (CBT))?

Yes

No

13. Please indicate the year, country in which you took the ToEFL, the type of ToEFL administration (PBT= Paper-based; CBT= Computer-based; IBT= Internet-based) and the total score you earned on that administration.

Year

Country

Type

Total score

14. On average, about how many hours per week and how long did you spend preparing for the internet-based ToEFL?

15. Have you ever taken any kind of computer based assessment in addition to ToEFL IBT?

Yes

No

16. Which of the following computer-based assessments did you take? (Check all that apply.)

GRE

SAT

GMAT

Course assessment

Computer-based IELTS

Others _____

Instructions: Below are statements that you may agree or disagree with. Using the –one to seven scale below, indicate your agreement with each item by placing the appropriate number. Please be open and honest in your responding.

Computerized ToEFL Acceptance Questionnaire

Assume that you had/ have options to choose internet-based ToEFL or paper-based ToEFL

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Using the internet-based ToEFL will improve my work on learning English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the internet- based ToEFL will enhance my effectiveness on learning English.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the internet-based ToEFL will increase my productivity on learning English.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My interaction with the system (ToEFL IBT testing system) is clear and understandable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is easy for me to become professional in using the	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

system (ToEFL IBT testing system)							
I find the system (ToEFL IBT testing system) easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can complete a job or task using the computer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can complete a job or task using the computer if someone showed me how to do it first	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can navigate easily through the Web to find any information I need	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was fully able to use computer and Internet before I began using the internet- based ToEFL	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
People who influence my behavior think that I should use the internet-based ToEFL	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
People who are important to me think that I should	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<p>use internet-based ToEFL</p> <p>The seniors in my school have been helpful in the use of internet-based ToEFL</p> <p>In general, my prior education institute supported the use of the internet-based ToEFL</p> <p>When I need help to use the internet-based ToEFL, someone is there to help me</p> <p>When I need help to learn to use the internet-based ToEFL, (ToEFL IBT testing system) responsible people and computer help are there to teach me</p> <p>Internet-based ToEFL's questions were clear and understandable</p> <p>Internet-based ToEFL's questions were easy to answer</p>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
---	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Internet based ToEFL's questions were relative to the course's that I took in the university.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Internet based ToEFL's questions were useful for my higher education (undergraduate or graduate education)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Test preparation was sufficient for the internet-based ToEFL	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My personal preparation was sufficient for the internet-based ToEFL	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My performance expectations were sufficient for the internet-based ToEFL	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the internet-based ToEFL keeps me happy for my task	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the internet-based ToEFL gives me enjoyment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

for my learning							
Using the internet-based ToEFL, my curiosity is stimulated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the internet-based ToEFL will lead to my exploration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I intend to use the internet-based ToEFL in the future	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I predict I would use the internet-based ToEFL in the future if I needed a ToEFL score.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I plan to use the internet-based ToEFL in the future	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

REFERENCES

- Agarwal, R. P. (2000). *Difference equations and inequalities: Theory, methods, and applications*. New York, NY: Marcel Dekker.
- Agarwal, R. P., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24(4), 665-694.
- Agarwal, R. P., Sambamurthy, V., & Stair, R. M. (2000). Research report: The evolving relationship between general and specific computer self-efficacy: An empirical assessment. *Information Systems Research*, 11(4), 418-430.
- Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior*. Berlin: Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
- Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22(5), 453-474.
- Alderson, J. C. (2009). Test of English as a foreign language: Internet-based test (ToEFL IBT). *Language Testing*, 26(4), 621-631.
- Anandarajan, M., Igarria, M., & Anakwe, U. P. (2000). Technology acceptance in the banking industry: A perspective from a less developed country. *Information Technology and People*, 13(4), 298-312.
- Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet based MBA courses. *Journal of Management Education*, 24(1), 32-54.
- Arenas-Gaitán, J., Ramírez-Correa, P. E., & Javier Rondán-Cataluña, F. (2011). Cross cultural analysis of the use and perceptions of web based learning systems. *Computers & Education*, 57(2), 1762-1774.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, 8(4), 243-254.
- Bagozzi, R. P., & Lee, K. H. (2002). Multiple routes for social influence: The role of

- compliance, internalization, and social identity. *Social Psychology Quarterly*, 65(3), 226-247.
- Bandura, A. (1977). Self-efficacy toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191.
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.
- Barclay, D., Thompson, R., & Higgins, C. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(2), 285-309.
- Barnett, T., Kellermanns, F. W., Pearson, A. W., & Pearson, R. A. (2006). Measuring information system usage: Replication and extensions. *Journal of Computer Information Systems*, 47(2), 76-85.
- Bem, S. L. (1981). *Sex-role inventory: Professional manual*. Palo Alto, CA: Consulting Psychologists Press.
- Benbasat, I., & Taylor, R. N. (1978). The impact of cognitive styles on information system design. *MIS Quarterly*, 2(2), 43-54.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588.
- Birch, D., & Burnett, B. (2009). Bringing academics on board: Encouraging institution-wide diffusion of e-learning environments. *Australasian Journal of Educational Technology*, 25(1), 117-134.
- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90-98.
- Bocij, P., & Greasley A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 1(1), 3-12.
- Bollen, K. A. (1989a). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Thousand Oaks, CA: Sage.
- Bonanno, P., & Kommers, P. A. M. (2008). Exploring the influence of gender and gaming

- competence on attitudes towards using instructional games. *British Journal of Educational Technology*, 39(1), 97-109.
- Boudreau, M. C., Gefen, D., & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. *MIS Quarterly*, 25(1), 1-16.
- Breland, H., Lee, Y., Najarian, M., & Muraki, E. (2004). *An analysis of ToEFL-CBT writing prompt difficulty and comparability for different gender groups* (ETS Research Report 4). Retrieved from Educational Testing Service Website: www.ets.org.libproxy.lib.unc.edu/Media/Research/pdf/RR-04-05.pdf
- Broadfoot, P. M., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education*, 11(1), 7-26.
- Broadfoot, P. M. (2005). Dark alleys and blind bends: Testing the language of learning. *Language Testing*, 22(2), 123-141.
- Brown, G. T., & Hirschfeld, G. H. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3-17.
- Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28(3), 282-299.
- Burton-Jones, A., & Straub Jr., D. W. (2006). Reconceptualizing system usage: An approach and empirical test. *Information Systems Research*, 17(3), 228-246.
- Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, 1(1), 55-86.
- Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456.
- Cassady, J. C., & Gridley, B. E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning, and Assessment*, 4(1), 1-14.
- Cassel, C., Hackl, P., & Westlund, A. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, 26(4), 435-446.
- Chen, M. (1986). Gender and computers: The beneficial effects of experience on attitudes. *Journal of Educational Computing Research*, 2(3), 265-282.

- Cheung, C. M., & Lee, M. K. (2010). A theoretical model of intentional social action in online social networks. *Decision Support Systems*, 49(1), 24-30.
- Cheung, C. M., Lee, M. K., & Chen, Z. (2002, January). Using the internet as a learning medium: An exploration of gender difference in the adoption of FaBWeb. *Proceedings of the 35th Hawaii International Conference on System Sciences, Hawaii, FL*, 475-483. doi: 10.1109/HICSS.2002.993901
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Chin, W. W. (1998b). The partial least squares approach for structural equation modeling. In G. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chin, W. W., & Todd, P. A. (1995). On the use, usefulness, and ease of use of structural equation modeling in MIS research: A note of caution. *Management Information Systems Quarterly*, 19, 237-237.
- Chou, C. P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald test. *Multivariate Behavioral Research*, 25(1), 115-136.
- Cohen, L., Manison, L., & Morrison, K. (2011). *Research methods in education*. New York, NY: Routledge.
- Comber, C., Colley, A., Hargreaves, D. J., & Dorn, L. (1997). The effects of age, gender and computer experience upon computer attitudes. *Educational Research*, 39(2), 123-133.
- Compeau, D. R., & Higgins, C. A. (1995a). Application of social cognitive theory to training for computer skills. *Information Systems Research*, 6(2), 118-143.
- Compeau, D. R., & Higgins, C. A. (1995b). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189-211. Retrieved from <http://www.jstor.org/stable/249688>
- Creswell, J. W. (2009). Editorial: Mapping the field of mixed methods research. *Journal of Mixed Methods Research*, 3(2), 95-108.
- Croft, A. C., Danson, M., Dawson, B. R., & Ward, J. P. (2001). Experiences of using computer assisted assessment in engineering mathematics. *Computers and Education*, 27(1), 53-66.
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Josey-Bass.
- Cuadrado-Garcia, M., Ruiz-Molina, M., & Montoro-Pons, J. D. (2010). Are there gender

- differences in e-learning use and assessment? Evidence from an interuniversity online project in Europe. *Social and Behavioral Sciences*, 2(2), 367-371.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. Retrieved from <http://www.jstor.org/stable/249008>
- Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3), 475-487.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982-1003.
- Davis, F. D., & Venkatesh, V. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, 27(3), 451-481.
- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, 19(2), 109-134.
- Deng, X., Doll, W. J., Hendrickson, A. R., & Scazzero, J. A. (2005). A multigroup analysis of structural invariance: An illustration using the technology acceptance model. *Journal of Marketing*, 42(5), 745-759.
- DeSanctis, G. (1983). Expectancy theory as an explanation of voluntary use of a decision support system. *Psychological Reports*, 52(1), 247-260.
- Deutsch, T., Herrmann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer based assessment: A web-based mock examination changes attitudes. *Computers & Education*, 58(4), 1068-1075. doi:10.1016/j.compedu.2011.11.013
- Dishaw, M. T., & Strong, D. M. (1999). Extending the technology acceptance model with task-technology fit constructs. *Information & Management*, 36(1), 9-21.
- Doll, W. J., & Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS Quarterly*, 12(2), 259-274.
- Drasgow, F. (1999). *Innovations in computerized assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Drennan, J., Kennedy, J., & Pisarski, A. (2005). Factors affecting student attitudes toward flexible online learning in management education. *The Journal of Educational Research*, 98(6), 331-338.
- Durndell, A., & Haag, Z. (2002). Computer self-efficacy, computer anxiety, attitudes towards

- the internet and reported experience with the internet, by gender, in an east European sample. *Computers in Human Behavior*, 18(5), 521-535.
- Durndell, A., Haag, Z., & Laithwaite, H. (2000). Computer self-efficacy and gender: A cross cultural study of Scotland and Romania. *Personality and Individual Differences*, 28(6), 1037-1044.
- Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. *Psychology of Women Quarterly*, 15(2), 203-216.
- Eglesz, D., Fekete, I., Kiss, O. E., & Izso, L. (2005). Computer games are fun? On professional games and players' motivations. *Educational Media International*, 42(2), 117-124.
- Epitropaki, O., & Martin, R. (2005). From ideal to real: A longitudinal study of the role of implicit leadership theories on leader-member exchanges and employee outcomes. *Journal of Applied Psychology*, 90(4), 659-676.
- Escudier, M. P., Newton, T. J., Cox, M. J., Reynolds, P. A., & Odell, E. W. (2011). University students' attainment and perceptions of computer delivered assessment: A comparison between computer-based and traditional tests in a high-stakes examination. *Journal of Computer Assisted Learning*, 27(5), 440-447. doi:10.1111/j.1365-2729.2011.00409.x
- Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs. another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage*, 15(3), 596-603.
- Fischer, M., & Kopp, V. (2006). Computer-based preclinical assessment: Does the embedding of multiple choice questions in a clinical context change performance. *GMS Z Med Ausbild*, 23(3), 26-43.
- Fishbein, M. (1967). *Readings in attitude theory and measurement*. New York, NY: Wiley.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Australia: Addison-Wesley Pub. Co.
- Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology*, 25(4), 509-523.
- Fornell, C. (1982). *A second generation of multivariate analysis*. New York, NY: Praeger Publishers.
- Frank, S. (2011). IT organization assessment-using COBIT and BSC. *COBIT Focus*, 1(1), 1-6.

- Gall, M. D., & Borg, W. R. (1996). *Educational research: An introduction (6th ed.)*. New York, NY: Longman.
- Gannon, M. J. (1994). *Understanding global cultures: Metaphorical journeys through 17 countries*. Thousand Oaks, CA: Sage.
- Gao, Y. (2005). Applying the technology acceptance model to educational hypermedia: A field study. *Journal of Educational Multimedia and Hypermedia*, 14(3), 237-247.
- Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly*, 21(4), 389-400. Retrieved from <http://www.jstor.org/stable/249720>
- Ginther, A. (2001). *Effects of the presence and absence of visuals on performance on ToEFL CBT listening-comprehensive stimuli* (ETS Research Report 66). Retrieved from Educational Testing Service Website: http://www.ets.org/research/policy_research_reports/publications/report/2001/hsfh
- Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 94(2), 377-381.
- Goodhue, D., Lewis, W., & Thompson, R. (2006). PLS, small sample size and statistical power in MIS research. *Proceedings of the 39th Hawaii International Conference on System Sciences, Los Alamitos, CA*, 381, 202-212. doi: 10.1109/HICSS.2006.381
- Grandon, E. E., Alshare, K., & Kwun, O. (2005). Factors influencing student intention to adopt online classes: A cross-cultural study. *Journal of Computing Sciences in Colleges*, 20(4), 46-56.
- Grant, R. A. (1991). Building and testing a causal model of an information technology's impact. *Journal of Information Technology Management*, 2(1), 11-23.
- Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. *Simulation & Gaming*, 41(2), 238-259.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate data analysis*. New York, NY: Maxwell Macmillan.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis (5th ed.)*. Upper Saddle River, NJ: Prentice Hall.
- Harrison, A. W., & Rainer Jr, R. K. (1992). The influence of individual differences on skill in end-user computing. *Journal of Management Information Systems*, 9(1), 93-112.
- Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication*, 11(4), 910-931.

- Hartwick, J., & Barki, H. (1994). Explaining the role of user participation in information system use. *Management Science*, 40(4), 440-465.
- He, J., & Freeman, L. A. (2010). Are men more technology-oriented than women? The role of gender on the development of general computer self-efficacy of college students. *Journal of Information Systems Education*, 21(2), 203-212.
- He, L., & Shi, L. (2008). ESL students' perceptions and experiences of standardized English writing tests. *Assessing Writing*, 13(2), 130-149.
- Heeks, R. (2002). I-development not e-development: Special issue on ICTs and development. *Journal of International Development*, 14(1), 1-11.
- Hill, T., Smith, N. D., & Mann, M. F. (1987). Role of efficacy expectations in predicting the decision to use advanced technologies: The case of computers. *Journal of Applied Psychology*, 72(2), 307.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Hung, M., Chou, C., Chen, C., & Own, Z. (2010). Learner readiness for online learning: Scale development and student perceptions. *Computers & Education*, 55(3), 1080-1090.
- Ismail, A., & Celikli, G. E. (2009). How does student ability and self-efficacy affect the usage of computer technology? *The Turkish Online Journal of Educational Technology*, 8(1), 20-26.
- Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of ToEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, 44(1), 1-21.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Lincolnwood, IL: Scientific Software International.
- Karahanna, E., & Straub, D. (1999). The psychological origins of perceived usefulness and ease-of-use. *Information & Management*, 35(4), 237-250.
doi:10.1016/j.physletb.2003.10.071
- Karsten, R., & Schmidt, D. (2008). Business student computer self-efficacy: Ten years later. *Journal of Information Systems Education*, 19(4), 445-453.
- Keasar, T., Baruch, R., & Grobgeld-Dahan, E. (2005). An evaluation of web enhanced

- instruction in college level biology courses. *Australasian Journal of Educational Technology*, 21(4), 533.
- Kelleher, T., & O'Malley, M. (2005). Applying the technology acceptance model to assess outcomes in a globally linked strategic communication project. *Journalism & Mass Communication Educator*, 60(4), 402-414.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Los Angeles, CA: Sage.
- Kennedy, G., Gray, K., & Tse, J. (2008). 'Net Generation' medical students: Technological experiences of preclinical and clinical students. *Medical Teacher*, 30(1), 10-16.
- Kennedy, G., Judd, T. S., Churchward, A., Gray, K., & Krause, K. L. (2008). First year students' experiences with technology: Are they really digital natives. *Australasian Journal of Educational Technology*, 24(1), 108-122.
- Kies, S. M., Williams, B. D., & Freund, G. G. (2006). Gender plays no role in student ability to perform on computer-based examinations. *BMC Medical Education*, 6(1), 57.
- King, C. G., Guyette Jr, R. W., & Piotrowski, C. (2009). Online exams and cheating: An empirical analysis of business students' views. *Journal of Educators Online*, 6(1), 15-32.
- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740-755.
- Kline, R. B. (1998). Software review: Software programs for structural equation modeling: Amos, EQS, and LISREL. *Journal of Psychoeducational Assessment*, 16(4), 343-364.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford.
- Klinger, D. A., & Luce-Kapler, R. (2007). Walking in their shoes: Students' perceptions of large-scale high-stakes testing. *The Canadian Journal of Program Evaluation*, 22(3), 29-52.
- Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th ed.). Los Angeles, CA: Sage.
- Lai, V. S., & Li, H. (2005). Technology acceptance model for internet banking: An invariance analysis. *Information & Management*, 42(2), 373-386.
- Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. New York, NY: Routledge.
- Lederer, A. L., Maupin, D. J., Sena, M. P., & Zhuang, Y. (2000). The technology acceptance

- model and the world wide web. *Decision Support Systems*, 29(3), 269-282.
- Lee, Y. C. (2008). The role of perceived resources in online learning adoption. *Computers & Education*, 50(4), 1423-1438.
- Lee, Y. W., Breland, H., & Muraki, E. (2005). Comparability of ToEFL CBT writing prompts for different native language groups. *International Journal of Testing*, 5(2), 131-158.
- Lester, L. H. (2008). *Measuring, modeling and monitoring the dynamics of labour market success and successful settlement of immigrants to Australia* (Unpublished doctoral dissertation). National Institute of Labour Studies, Flinders University, Adelaide.
- Li, N., & Kirkup, G. (2007). Gender and cultural differences in internet use: A study of China and the UK. *Computers & Education*, 48(2), 301-317.
- Link, T. M., & Marz, R. (2006). Computer literacy and attitudes towards e-learning among first year medical students. *BMC Medical Education*, 6(1), 34.
- Ma, W. W. K., Andersson, R., & Streith, K. O. (2005). Examining user acceptance of computer technology: An empirical study of student teachers. *Journal of Computer Assisted Learning*, 21(6), 387-395.
- Maldonado, U. P. T., Khan, G. F., Moon, J., & Rho, J. J. (2009, November). E-learning motivation, students' acceptance and use of educational portal in developing countries: A case study of Peru. *Proceeding of the Fourth International Conference on Computer Sciences and Convergence Information Technology, South Korea*, 77, 1431-1441. doi: 10.1109/ICCIT.2009.77
- Maldonado, U. P. T., Khan, G. F., Moon, J., & Rho, J. J. (2011). E-learning motivation and educational portal acceptance in developing countries. *Online Information Review*, 35(1), 66-85.
- Manning, M. (2009). The effects of subjective norms on behavior in the theory of planned behavior: A meta-analysis. *British Journal of Social Psychology*, 48(4), 649-705.
- Manochehr, N. N. (2006). The influence of learning styles on learners in e-learning environments: An empirical study. *Computers in Higher Education Economics Review*, 18(1), 10-14.
- Mao, E., Srite, M., Bennett Thatcher, J., & Yaprak, O. (2005). A research model for mobile phone service behaviors: Empirical validation in the US and Turkey. *Journal of Global Information Technology Management*, 8(4), 7-28.
- Marcoulides, G. A., & Saunders, C. (2006). Editor's comments: PLS: A silver bullet? *MIS Quarterly*, 30(2), iii-ix.

- Martinsons, M. G., & Davison, R. (2003). Guest editorial cultural issues and it management: Looking ahead. *IEEE Transactions on Engineering Management*, 50(1), 113-117.
- Mathieson, K., Peacock, E., & Chin, W. W. (2001). Extending the technology acceptance model: The influence of perceived user resources. *ACM SigMIS Database*, 32(3), 86-112.
- Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal*, 48(5), 874-888.
- McCoy, S., Everard, A., & Jones, B. M. (2005). An examination of the technology acceptance model in Uruguay and the US: A focus on culture. *Journal of Global Information Technology Management*, 8(2), 27-45.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55(1), 51.
- McSporrnan, M., & Young, S. (2001). Does gender matter in online learning? *Research in Learning Technology*, 9(2), 3-15.
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10(9), 9-20.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45(1), 35-44.
- Meyers, W., Bennett, S., & Lysaght, P. (2004). Asynchronous communication: Strategies for equitable e-learning. *Proceedings of the 21th Australian Society for Computers in Learning in Tertiary Education, Perth, Western Australia*, 655-662. Retrieved from <http://www.ascilite.org.au/conferences/perth04/procs/meyers.html>
- Mills, C. N., Potenza, M. T., Fremer, J. J., & Ward, W. C. (Eds.). (2005). *Computer-based testing □: Building the foundation for future assessments*. Mahwah, NJ: L. Erlbaum Associates.
- Mitra, S., Sambamurthy, V., & Westerman, G. (2011). Measuring IT performance and communicating value. *MIS Quarterly Executive*, 10(1), 47-59.
- Moni, K. B., Van Kraayenoord, C. E., & Baker, C. D. (2002). Students' perceptions of literacy assessment. *Assessment in Education: Principles, Policy & Practice*, 9(3), 319-342.
- Moon, J., & Kim, Y. (2001). Extending the TAM for a world-wide-web context. *Information & Management*, 38(4), 217-230.

- Muthen, B., & Muthen, L. (2012). Mplus (7.1) [Statistical modeling program]. Los Angeles, CA: Muthen & Muthen.
- Nair, I., & Das, V. M. (2012). Using technology acceptance model to assess teachers' attitude towards use of technology as teaching tool: A SEM approach. *International Journal of Computer Applications*, 42, 12-33.
- Ngai, E. W., Poon, J. K. L., & Chan, Y. H. C. (2007). Empirical examination of the adoption of WebCT using TAM. *Computers & Education*, 48(2), 250-267.
- Noyes, J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology*, 35(1), 111-113.
- Odedra, M., Bennett, M., Goodman, S., & Lawrie, M. (1993). Sub-Saharan Africa: A technological desert. *Communications of the ACM*, 36(2), 25-29.
- Olivier, T. A., & Shapiro, F. (1993). Self-efficacy and computers. *Journal of Computer Based Instruction*, 20(3), 81-85.
- Ong, C., & Lai, J. (2006). Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Computers in Human Behavior*, 22(5), 816-829. doi:10.1016/j.chb.2004.03.006
- Ozden, M. Y., Erturk, I., & Sanli, R. (2004). Students' perceptions of online assessment: A case study. *Journal of Distance Education*, 19(2), 77-92.
- Padilla-Meléndez, A., Garrido-Moreno, A., & Del Aguila-Obra, A. R. (2008). Factors affecting e-collaboration technology use among management students. *Computers & Education*, 51(2), 609-623
- Pan, C. C., Sivo, S., Gunter, G., & Cornell, R. (2005). Students' perceived ease of use of an e-learning management system: An exogenous or endogenous variable? *Journal of Educational Computing Research*, 33(3), 285-307.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York, NY: Springer.
- Pino-Silva, J. (2008). Student perceptions of computerized tests. *ELT Journal*, 62(2), 148-156.
- Pituch, K. A., & Lee, Y. K. (2006). The influence of system characteristics on e-learning use. *Computers & Education*, 47(2), 222-244.
- Png, I. P., Tan, B. C., & Wee, K. L. (2001). Dimensions of national culture and corporate

- adoption of IT infrastructure. *IEEE Transactions on Engineering Management*, 48(1), 36-45.
- Qureshi, I., & Compeau, D. (2009). Assessing between-group differences in information systems research: A comparison of covariance- and component-based SEM. *MIS Quarterly*, 33(1), 197-214.
- Ricketts, C., & Wilks, S. J. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & Evaluation in Higher Education*, 27(5), 475-479.
- Rogers, E. M. (1995). *Diffusion of innovations*. New York, NY: Simon and Schuster Inc.
- Rothbard, N. P. (2001). Enriching or depleting? The dynamics of engagement in work and family roles. *Administrative Science Quarterly*, 46(4), 655-684.
- Russell, M. (1999). Testing writing on computers: A follow-up study comparing performance on computer and on paper. *Educational Policy Analysis Archives*, 7(20), 115.
- Russell, M., & Plati, T. (2002). Does it matter with what I write? Comparing performance on paper, computer and portable writing devices. *Current Issues in Education*, 5(4), 24.
- Russell, M., Goldberg, A., & O'connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy & Practice*, 10(3), 279-293.
- Sambell, K., Sambell, A. & Sexton, G. (1999). Student perceptions of the learning benefits of computer-assisted assessment: A case study in electronic engineering. In S. Brown, P. Race & J. Bull (Eds.), *Computer Assisted Assessment in Higher Education* (pp. 179-191). London: Kogan Page Limited.
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the ToEFL internet based test. *Language Testing*, 26(1), 5-30. doi:10.1177/0265532208097335
- Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management*, 44(1), 90-103.
- Schneberger, S., Amoroso, D. L., & Durfee, A. (2008). Factors that influence the performance of computer-based assessments: An extension of the technology acceptance model. *Journal of Computer Information Systems*, 48(2), 74-90.
- Scott, C. R., & Rockwell, S. C. (1997). The effect of communication, writing, and technology apprehension on likelihood to use new communication technologies. *Communication Education*, 46(1), 44-62.

- Selim, H. M. (2003). An empirical investigation of student acceptance of course websites. *Computers & Education*, 40(4), 343-360.
- Shee, D. Y., & Wang, Y. S. (2008). Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Computers & Education*, 50(3), 894-905.
- Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research*, 15(3), 325-343.
- Shohamy, E. G. (2001a). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.
- Shohamy, E. G. (2001b). *The power of tests: A critical perspective on the uses of language tests*. Harlow, England: Longman.
- Shohamy, E. G. (2004). Assessment in multi-cultural societies: Applying democratic principles and practices to language testing. *Critical Pedagogies and Language Learning*, 5, 72-92.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173-180.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893-898.
- Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, 13(2), 147-169.
- Straub, D., Keil, M., & Brenner, W. (1997). Testing the technology acceptance model across cultures: A three country study. *Information & Management*, 33(1), 1-11.
- Straub, D., Limayem, M., & Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science*, 41(8), 1328-1342.
- Stricker L. J., & Attali Y. (2010). *Test takers attitudes about the ToEFL IBT* (ToEFL IBT Research Report 13). Retrieved from Educational Testing Service Website: <http://www.ets.org/Media/Research/pdf/RR-10-02.pdf>
- Sun, H., & Zhang, P. (2006). The role of moderating factors in user technology acceptance. *International Journal of Human-Computer Studies*, 64(2), 53-78.
doi:10.1016/j.ijhcs.2005.04.013
- Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Belmont, CA: Thomson/Brooks/Cole.

- Tan, B. C. Y., Watson, R. T., & Wei, K. K. (1995). National culture and group support systems: Filtering communication to dampen power differentials. *European Journal of Information Systems*, 4(2), 82-92.
- Tate, R. (1996). *An introduction to modeling outcomes in the behavioral and social sciences*. Edina, MN: Burgess International Group.
- Taylor, C., Kirsch, I., Jamieson, J., & Eignor, D. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning*, 49(2), 219-274.
- Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research*, 6(2), 144-176.
- Terzis, V., & Economides, A. A. (2011a). The acceptance and use of computer based assessment. *Computers & Education*, 56(4), 1032-1044.
doi:10.1016/j.compedu.2010.11.017
- Terzis, V., & Economides, A. A. (2011b). Computer based assessment: Gender differences in perceptions and acceptance. *Computers in Human Behavior*, 27(6), 2108-2122.
doi:10.1016/j.chb.2011.06.005
- Terzis, V., Moridis, C. N., & Economides, A. A. (2013). Measuring instant emotions based on facial expressions during computer-based assessment. *Personal and Ubiquitous Computing*, 17(1), 43-52.
- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly*, 15(1), 125-143.
- Tomte, C., & Hatlevik, O. E. (2011). Gender differences in self-efficacy ICT related to various ICT-user profiles in Finland and Norway: How do self-efficacy, gender and ICT-user profiles relate to findings from PISA 2006. *Computers & Education*, 57(1), 1416-1424.
- Triandis, H. C. (1971). *Attitude and attitude change*. New York, NY: Wiley.
- Triandis, H. C. (1977). *Interpersonal behavior*. Monterey, CA: Brooks/Cole Publishing Company.
- Triandis, H. C. (1979). Values, attitudes, and interpersonal behavior. *Nebraska symposium on motivation*. Symposium conducted at the meeting of the University of Nebraska, Lincoln, NE.
- Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high stakes testing. *Journal of Literacy Research*, 37(2), 273-260.

- Tsaia, Y., & Tsou, C. H. (2009). A standardized English language proficiency test as the graduation benchmark: Student perspectives in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319-330.
- Van Raaij, E. M., & Schepers, J. J. (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education*, 50(3), 838-852.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Veiga, J. F., Floyd, S., & Dechant, K. (2001). Towards modeling the effects of national culture on IT implementation and acceptance. *Journal of Information Technology*, 16(3), 145-158.
- Vekiri, I., & Chronaki, A. (2008). Gender issues in technology use: Perceived social support, computer self-efficacy and value beliefs, and computer use beyond school. *Computers & Education*, 51(3), 1392-1404.
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 11(4), 342-365.
- Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, 27(3), 451-481. doi:10.1111/j.1540-5915.1996.tb00860.x
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186-204.
- Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly*, 24(1), 115-139.
- Venkatesh, V., Morris, M. G., Davis G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478. Retrieved from <http://www.jstor.org/stable/30036540>
- Venkatesh, V., Speier, C., & Morris, M. G. (2002). User acceptance enablers in individual decision making about technology: Toward an integrated model. *Decision Sciences*, 33(2), 297-316. doi:10.1111/j.1540-5915.2002.tb01646.x
- Wang, Y. S. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management*, 41(1), 75-86.
- Wang, Y. S., Wu, M., & Wang, H. (2009). Investigating the determinants and age and gender

- differences in the acceptance of mobile learning. *British Journal of Educational Technology*, 40(1), 92-118. doi:10.1111/j.1467-8535.2007.00809.x
- Wheaton, D. E. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, 8, 84-136.
- White, S. A., & Duda, J. L. (1994). The relationship of gender, level of sport involvement, and participation motivation to task and ego orientation. *International Journal of Sport Psychology*, 25, 4-18.
- Whitelock, D. (2009). Editorial: E-assessment: Developing new dialogues for the digital age. *British Journal of Educational Technology*, 40(2), 199-202.
- Whitley, B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior*, 13(1), 1-22.
- Williams, J. E., Bennett, S. M., & Best, D. L. (1975). Awareness and expression of sex stereotypes in young children. *Developed Psychology*, 11(5), 635.
- World Bank. (2014). *World development indicators* [Databank]. Retrieved from <http://data.worldbank.org/indicator/NY.GNP.PCAP.CD>
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147-170.
- Yuen, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology* 30(1), 165-200.
- Yuen, H. K., & Ma, W. K. (2002). Gender differences in teacher computer acceptance. *Journal of Technology and Teacher Education*, 10(3), 365-382.
- Yun, J., & Ulrich, D. A. (2002). Estimating measurement validity: A tutorial. *Adapted Physical Activity Quarterly*, 19(1), 32-47.
- Zhao, J., & Tan, W. (2010, November). E-learning systems adoption across cultures: A comparison study. *Proceedings of the International Conference on E-Product E-Service and E-Entertainment, Henan, China*, 1-4. doi:10.1109/ICEEE.2010.5661033