January 2013

Translation, Adaptation and Invariance Testing of the Teaching Perspectives Inventory: Comparing Faculty of Malaysia and the United States

Jecky Misieng
University of South Florida, mjecky@cls.unimas.my

Translation, Adaptation and Invariance Testing of the Teaching Perspectives Inventory:

Comparing Faculty of Malaysia and the United States

by

Jecky Misieng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Measurement and Research
College of Education
University of South Florida

Major Professor: Robert F. Dedrick, Ph.D.
Yi-hsin Chen, Ph.D.
John M. Ferron, Ph.D.
James A. Eison, Ph.D.

Date of Approval:
November 15, 2013

Keywords: Teaching Beliefs, Cross-cultural Measurement, Translations and Adaptations of Research Instruments, Invariance Testing, Higher Education, Faculty Development

Dedication

To my mom, Noed Mudak, may she rest in peace.

Acknowledgments

It is with great humility that I express my gratitude to both of my major professors, Dr. Robert Dedrick and Dr. Yi-hsin Chen for their patience and enthusiasm in providing encouragement, guidance, and support to endure in this endeavor. I am also extremely grateful to my committee members, Dr. John Ferron and Dr. James Eison, for willingly sharing their wisdom and experience through insightful suggestions and feedback. I am proud to say that I have worked with all of them.

To my wife and children, thank you for bearing with me in my journey of learning as a lifelong process. I am indebted to Dr. Mike Churton, Dr. Betsy Doone, Dr. Victor Peppard, Rev. Theron Chewning, and Rev. Freddy Liwang for making our sojourn a memorable and blessed one.

Thank you, also, to my colleagues in UNIMAS who have helped me translate the TPI into Bahasa Malaysia. I cannot mention your name to protect your privacy but you know who you are. A special mention for Dr. Soubakeavathi Rethinasamy and Mr. Chuah Kee Man for their much needed assistance when I least expected it. I am greatly indebted to Ms. Mitty Chai and Universiti Malaysia Sarawak, especially Dr. Shahren Ahmad Zaidi Addruce for putting their trust in me to pursue my dream.

I am indebted to all my colleagues in the 20 public universities in Malaysia who graciously allowed me and made the time to complete my surveys. A sincere thanks to the officers at the Ministry of Higher Education of Malaysia for giving their approval to carry out this research. I really appreciate the kindness of Dr. Dan Pratt and Dr. John Collins for allowing me to use the Teaching Perspectives Inventory (TPI) as well as providing the data from their database for my US sample.

Lastly, I thank the Lord, my God for giving me this race to run and for faithfully staying at my side throughout this journey. To Him be the glory and honor forever. Amen.

Table of Contents

List of Tables

List of Figures

Abstract

As a result of growing attention in cross-cultural research, existing measurement instruments developed in one language are being translated and adapted for use in other languages and cultural contexts. The benefits of having the same instrument across cultures can only be realized if the process of translation and adaptation of the measurement instruments produces measurement operations that function similarly across national and cultural boundaries. Producing invariant measurement instruments that assess educational and psychological constructs provide a way of testing the cross-cultural generality of theories that include these constructs.

The major purposes of the study were to translate and adapt the Teaching Perspectives Inventory (Pratt, 1992, 1990) from English to Bahasa Malaysia and compare the psychometric properties of the two versions. The TPI is an instrument developed by Pratt (1992) to ascertain the different conceptions that teachers in higher education have about teaching. The TPI has 45 items, which are divided into five subscales or perspectives referred to as Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform. The first phase of this study translated and adapted the TPI from English into the Malay language of Malaysia or Bahasa Malaysia (BM) using multiple approaches as recommended by the International Test Commission. The approaches used to translate the TPI included forward and back translations, an expert panel review, a pilot study, and cognitive interviews. In the translation process, three initial translators, two back translators, and six expert panel members, including the researcher, came up with a pre-final version of the Malay TPI. During the translation process, two items

were found to contain expressions that had no exact equivalent forms in Malay: “virtuoso performers” and “higher ideals.” Overall, translating the TPI was a challenging task due to the relatively large number of items in the instrument (45) as well as the complexity and very abstract nature of the constructs. Many of the words and expressions that were brief and concise in the English version became longer and more verbose when translated in Malay. As a result, the translated TPI version appeared longer than the original version. Pilot testing with 25 native speakers of Malay who were faculty members from a number of public universities in Malaysia revealed nine items that needed modification. Cognitive interviewing with five participants from the pilot group revealed one item requiring a change by adding a borrowed word “novis” in brackets next to the Malay expressions, which refers to the original word *novice*. Due to the confusion with the words referring to ‘people’ in many of the items, additional instructions were added at the beginning of the survey to ensure that the participants responded according to the original intention of the items, which focuses on learners in the faculty’s specific classroom context instead of people in the society in general. Following changes to the TPI, this instrument was administered in phase two to a Malaysian sample of 561 faculty.

In the second phase, the study assessed the psychometric properties of the original English version of the TPI with 605 faculty in the U. S. and the translated TPI version of the TPI with the Malaysian sample. The overall internal consistency reliability of both the English (α=.88) and the Malay TPI (α=.93) appeared to be adequate. At the subscale level, the internal consistency reliabilities of all the scales were on the lower side considering the large number of items (9) for each subscale (range = .67 to .83 for the U. S. and .59 to .81 for Malaysia). It was found that three out of the five subscales of the U. S. and Malay TPI had similar alpha reliabilities (Apprenticeship, Nurturing, Social Reform). To assess the cross-cultural factorial

validity and measurement invariance of the TPI, confirmatory factor analysis (CFA) was carried out for both the original and the Malay TPI. The sample size for the U. S. group was 605 and the Malay group was 561. The fit for both the U. S. and the Malay correlated five-factor models was less than adequate with the Malay model showing a much worse fit. Correlated errors were found between 64 item pairs in the U. S. model and 389 item pairs in the Malay model. The correlations between the five perspectives in the Malay sample were much higher than those in the U. S. sample suggesting that the perspectives had limited discriminant validity. For example, the correlations between the Nurturing and Developmental perspectives and Nurturing and Social Reform perspectives were 1.0. The inadequate fit of the five-factor correlated model in the Malaysian sample and the minimally acceptable fit in the U. S. sample led to the decision to carry out analyses and compare the groups one subscale at a time. Model modifications for each subscale of both samples were carried out to improve the fit by adding one or more parameters (i.e., correlated errors) for each subscale model to obtain acceptable baseline models. The results of the invariance testing for each subscale did not support the existence of measurement invariance. Overall, the results indicate that the Malay version of the TPI is not ready for use and additional translation and adaptation work is recommended. Future efforts could incorporate improvements in the translation process in the form of recruiting a larger number of certified translators who have in-depth knowledge of teaching in higher education as well as a deep knowledge of the philosophy and purposes behind the TPI. Additional cognitive interviews before and after pretesting and pilot testing of the pre-final version are recommended. Finally, adding a large sample of bilingual educators who would complete both the Malay and English versions of the TPI would provide important psychometric data on the equivalence of the TPI items.

Chapter One:

Introduction

The challenges of understanding how teachers conceptualize the act of teaching and how these conceptualizations influence teachers' actions are complex, particularly in the context of adult and higher education. Many educationists adopt a view that there is a theory that best captures what learning is all about and offer the type of teaching methods that will effectively promote learning. As a result of this commonly accepted wisdom, three major philosophical approaches-- behaviorism, cognitivism, and constructivism—have competed for a place in the hearts and mind of the educationist. However, there are other experts who argue against this idea. It is their contention that there are many equally effective teaching methods depending on the context and the situation. One of the proponents of this view is Pratt (2002) who claims that there is no one best method of teaching and suggested that beliefs about teaching are influenced by both personal and external factors (Pratt, 1992). He suggests that the personal domain of meaning, values, beliefs, and intentions entwine with socio-cultural and historical standpoints that influence the way teaching is conceptualized in the consciousness. Many studies have investigated teachers' perceptions about teaching and their effects on student learning and have done so by measuring both teacher behaviors and underlying teacher beliefs and values (Gow & Kember, 1993; Kember & Gow, 1994; Pratt, 1992; Trigwell & Prosser, 1996a; Trigwell & Prosser, 1996b). Other studies have shown that teachers' beliefs about teaching and the teaching process are related to their practices (Borrich, 1999; Clark, 1986; Clark & Yinger, 1979; Fang,

1996; Kagan, 1992; Pajares, 1992; Thompson, 1992). Clark and Yinger (1979) describe teacher beliefs or implicit theories as the perceptions of the teacher when defining the elements of the classroom situation that are most important, the relationship between them, and the order in which they should be considered. In discussing the notion of teacher beliefs, Shavelson and Stern (1981) reported a number of studies that suggest that beliefs play a role in teachers' decisions, judgments, and behavior. They make a distinction between knowledge and beliefs by stating that when information is unavailable, teachers will rely on beliefs to direct them. In order words, beliefs form the basis for teachers' decision-making in the classroom when guidance is inaccessible. As a result of their observation, research must place an emphasis on the beliefs, attitudes, expectations, and perceptions about teaching and learning in order to understand why teachers do what they do in the classroom (Shavelson & Stern, 1981). Understanding what teachers' beliefs are and their importance to teaching and learning must be explored thoroughly. Research instruments used in assessing and measuring teachers' beliefs need to be examined to ensure that they faithfully explain the constructs being measured. The benefits of this undertaking are indispensable if the validity of the findings from closely-related studies can be shared with all educational practitioners world-wide.

Teachers' beliefs are important because of their indirect impact on students' learning. Most teachers' beliefs about teaching and learning come from their experiences as students themselves (Pajares, 1992). Observations and opinions about their own teachers' classroom behaviors help aspiring teachers in forming their early perceptions about teaching. By the time most of these students enter college many of their beliefs about teaching are already well established and are deeply entrenched in their schemata. Therefore, if they were to take up

teaching as a career of choice, their teaching styles would be influenced by these beliefs and as a consequence, many will exhibit behaviors that are very similar to what their former teachers did regardless of their effectiveness. This is important to note because many studies have shown that teachers' beliefs have an impact on learners' progress (Borrich, 1999; Clark, 1986; Kagan, 1992; Pajares, 1992; Thompson, 1992). Misguided notions about teaching may lead to ill-informed practices that may negatively impact student performance as well as development. This is further reiterated by Fang's (1996) study, which revealed that teacher beliefs, practices, and actions have a positive relationship with students' learning. In addition, beliefs not only shape how people behave but what they perceive in their environment. According to Menges (1990), not only do beliefs influence the likelihood of particular behaviors, they also influence perceptions, acting as filters that can distort otherwise objective data. Teachers with misguided assumptions about teaching may interpret classroom events erroneously and that will subsequently have a negative effect on their teaching as a whole. For example, the assumption that mass lecture is the best method in getting students to learn a foreign language quickly is flawed and this notion may stem from the belief that language learning is just another subject to be learned by rote by students rather than a skill that needs hands-on practice. In an exploratory study that examined the impact of several variables on the scholarship of teaching, Lueddeke (2003) reported teaching conceptualization as one of the factors that had the strongest influence on teaching practice. Therefore, teaching conceptions not only affect student learning but also influence the effectiveness of the professional growth of the instructors themselves.

Attempts to conceptualize the abstract notion of teachers' beliefs have been in progress for decades. Among the many attempts to capture teachers' beliefs and teaching practices were

those done by Gow and Kember (1993; 1994) who developed an instrument to ascertain teaching beliefs among faculty in institutions of higher learning. Two teaching conceptions emerged from their study, which they labeled as "learning facilitation" and "knowledge transmission" (Kember & Gow, 1994, p. 61). According to Gow and Kember (1993) these orientations of teaching are related to three kinds of study approaches referred to as surface learning, deep learning, and achieving learning. Surface learning is often associated with rote learning without much understanding involved while deep learning is more about thoughtful reflection about what is being learned. Achieving learning, on the other hand, is learning with a goal of accomplishing something in mind. Kember and Gow (1994) also discovered that these orientations to teaching influenced classroom practice, and learning facilitation orientation encouraged a deep learning approach. A more recent qualitative study by Kember and Kwan (2000) confirmed the existence of these two orientations of teaching. Among the findings in this study was the confirmation that the instructors' conceptions of teaching are best captured by the two main orientations of learning: facilitation versus knowledge transmission. Another finding from the study was that instructors who held the belief that teaching is knowledge transmission were more inclined to adopt approaches that focus on content while those who believe teaching is facilitating learning focus more on learning-centered approaches .

Another attempt to understand teachers' beliefs that built on the work by Gow and Kember (1993, 1994) was Prosser and Trigwell's Approaches to Teaching Inventory (ATI). Trigwell and Prosser (2004) view teachers' conceptions about teaching and learning as a two-dimensional model consisting of intentions to teach and strategies employed to achieve those intentions. In 1999, they developed an instrument called the ATI to explore teachers' approaches

to teaching and reported interesting findings about a teacher-centered strategy versus a student-centered teaching strategy. In the study that employed this instrument for the first time, a strong positive relationship between the teaching strategies used as portrayed by university instructors and the techniques utilized by the students in their lessons was discovered (Prosser & Trigwell, 1999). In their findings, when student-centered approaches were claimed to have been used by the instructors, students reciprocated by adopting a deep-learning approach to deal with the tasks given. In other words, the learning approaches employed by students were dependent upon the teaching methods used by their instructors and this relationship was very desirable as students are not just passive and quiet learners but architects of their own learning process.

The Teaching Perspectives

The groundwork for this study is drawn from another survey instrument that was constructed by Pratt (1992) to ascertain teachers' underlying conceptions of teaching in general. Trigwell and Prosser's notion about teacher beliefs led Pratt (1992) to argue that a teacher's conception about teaching is influenced by a set of beliefs, which then determines his or her intentions and actions that are tied to knowledge, learning, and the teacher's role. Influenced also by Kember's research (Kember, 1997), Pratt (1998) argued that even though there are many variations in the teaching styles of faculty teaching in higher education there appears to be only a few ways to perceive teaching.

Based on these premises, an instrument called the Teaching Perspectives Inventory (TPI) was developed in 1992 and later validated for research purposes with the help of Collins (Pratt & Collins, 2001). In this context, Pratt introduced the notion of teacher beliefs specifically targeted at teaching adult learners as well as teaching in higher education. Pratt's model differs from

Trigwell and Prosser's model in one aspect. Unlike the two-dimensional ATI, the TPI is based on the conceptualization of teaching that encompasses five common views that are made up of three dimensions of what the teacher believes about teaching, what goals of teaching are to be achieved, and what tasks are to be carried out in order to achieve these teaching goals (Pratt, 1998). However, both models argue against the notion that there is one best method of teaching and propose that each belief or conception of teaching has its own strengths and weaknesses (Pratt, 2002) and their effectiveness is dependent upon the context of what is to be learned.

As a survey instrument, the TPI contains 45 items that have been translated into at least eight languages but only the English, Spanish, and Chinese versions are available online. In the most recent article, Collins and Pratt (2011) summarized information about the development and validation of the TPI along with a number of the most current research findings about teaching beliefs based on the instrument. In the same article, Collins and Pratt report that over 100, 000 educators from as many as 100 countries have taken the survey As for the instrument itself, the items are grouped under five common perspectives as theorized by the developers: Transmission (lecture and teacher-centered); Apprenticeship (experiential and coaching-oriented); Developmental (facilitation and learning-centered); Nurturing (focused on building learners' self-esteem); and Social Reform (change the status quo orientation). For each of the five perspectives, the items are further divided into the three subcategories of teachers' classroom practice, their organization of the learning situations, and their beliefs about teaching and learning. These subcategories are called Actions, Intentions, and Beliefs. Beliefs pertain to conceptions that determine what is to be taught and what evidence will be accepted that the knowledge has been taught successfully. Actions, on the other hand, are defined as those

activities that are described as routines and techniques used to engage people in the content of the teaching. Meanwhile, intentions are viewed as general statements that point toward an overall agenda or purpose about teaching. Questions for Actions typically ask about what is done when instructing or teaching. Each of the five perspectives of the TPI produces a numerical score and the perspective that has the highest score means that it is the dominant teaching perspective being espoused by the respondent. Each of the three subscales of beliefs, intentions, and actions within each perspective also yields its own numerical score.

With the recognition of the connection between teachers' perspectives and practices, researchers have initiated an increasing number of studies examining these constructs and their relationship. This research has not been confined to the United States or North America in general but rather has been conducted across several countries. To facilitate this research it has often been necessary to take measurement instruments developed in one language and translate and adapt these instruments for use in other countries. An important benefit of having a common instrument in multiple languages is that it is possible to examine the generalizability of the results related to constructs, such as teacher perspectives, and test the cross-cultural generality of the theories underlying these constructs. While this benefit is important from a scientific standpoint for building a body of generalized knowledge related to teachers' beliefs and practices, these benefits can only be realized if the process of translation and adaptation of measurement instruments produces measurement operations that are invariant cross-nationally. In other words, in order to conduct cross-cultural research using surveys to study between group differences, members of different groups must ascribe the same meanings to the survey items (Rensvold & Cheung, 1998). Horn (1991) pointed out that without evidence of measurement

invariance, the conclusions of a study would be weak. We need to know if the instrument can be used in other cultural contexts besides the one that it was intended for and according to AERA/APA/NCME Standards 13.4 "When a test is translated from one language or dialect to another, its reliability and validity for the uses intended in the linguistic groups to be tested should be established" (AERA, APA, & NCME, 1999).

There are many different types of measurement invariance namely configural, metric, and scalar invariance. Along with these different types of invariance, a number of ways to test for measurement invariance have also been developed based on parametric and non-parametric statistics. Invariance is essentially a condition which supports the notion that measures across groups are considered to be on the same scale if relationships between the indicators or items used to measure the latent trait are the same across groups (Meredith, 1993). Such a definition of measurement invariance requires equality in terms of the structure of the construct as revealed by identical factor loadings and equality in the psychometric properties such as intercepts, residuals, and factor variances and covariances. In other words, constructs such as teacher beliefs must have the same basic structure and share the same psychometric properties regardless of samples or groups. Groups can be in the form of personal differences such as gender, age and personality or even larger groupings such as those bounded by ethnicity and country. The relationships between the construct and the items must be similar as shown by their equal factor loadings **(**Cheung & Rensvold, 1998). In other words, they must be invariant across cultures.

Purpose of the Study

The purposes of the study were to translate and adapt Pratt's (1992, 2001) Teaching Perspectives Inventory (TPI) from English into Bahasa Malaysia (BM) and to evaluate the

success of the translation in achieving measurement invariance. Two groups of faculties teaching in universities from the United States of America and Malaysia were selected for comparison. In order to make cross-cultural comparisons, an invariant measurement system has to be constructed. One way to do that is to translate and adapt the instrument into a language that is familiar to respondents. By being able to share their perceptions unimpeded by language barriers, a more accurate measure can be achieved. However, simply translating from one language into another does not ensure accurate cultural and linguistic equivalence because the translation itself may produce differences in the measurement properties of the instrument. Extraneous differences in interpretation of the TPI need to be kept at a minimum by ensuring that both versions share as much similarity as possible in terms of formatting, instructions, and response options. In fact, Johnson (1998) pointed out that the importance of equivalence of survey questions rivals that of their reliability and validity. Procedures and guidelines for translating and adapting instruments as proposed by experts (van de Vijver & Hambleton, 1996; McGorry, 2000) served as a starting point when translating the TPI from English into BM. This process involved forward translation, back-translation, an expert panel review, and pilot testing. Following this process, measurement invariance of the TPI across the U.S. and Bahasa Malaysia faculty groups was examined to determine if the TPI indicators and the underlying constructs were the same across these two groups. To evaluate measurement invariance of the TPI, this study addressed two major questions:

1a. How well does the correlated five-factor structure of the TPI fit the data of college faculty from the U.S.?

1b. How well does the correlated five-factor structure of the TPI fit the data of university faculty from the Malaysia?

2. Is the correlated five-factor structure of the TPI invariant across the U.S. and Malaysian samples of university faculty?

Definition of Terms

The following definitions of terms were used in this study.

Beliefs. Beliefs are mental representations that influence how one views the physical and psychological world (Rokeach, 1968) and help one to define and understand the world by screening, filtering, and reorganizing new ideas so that they fit with our prior knowledge (Pajares, 1992).

Teacher beliefs. The perceptions of the teacher when defining the elements of the classroom situation that are most important, the relationship between them, and the order in which they should be considered (Clark & Yinger, 1979).

Teacher conceptions. According to Brown and Lake (2006), a conception is a mental construct or representation of reality containing beliefs, meanings, preferences, and attitudes that explains complex and difficult categories of experience. Teacher conceptions about the nature of teaching and learning are used synonymously with belief systems concerning teaching and learning as a whole. Teacher conception is used synonymously for teacher orientation, the beliefs, values, and perspectives of a teacher that underlie teaching.

Teaching perspectives. Pratt (1992) argued that a teacher's conception about teaching is influenced by a set of beliefs, which then determines his or her intentions and actions that are tied to knowledge, learning, and the teacher's role, which he refers to as teaching perspectives.

Transmission. This perspective of teaching refers to the teacher as a provider of knowledge through systematic presentations such as the lecture method. The focus is more on the mastery of the subject matter.

Apprenticeship. In this perspective, teachers are good practitioners who endeavor to impart their knowledge and skills by differentially guiding their learners from dependent individuals to independent practitioners themselves. Apprenticeship teachers are aware of what their learners can or cannot do and provide guidance where appropriate.

Developmental. Developmental teachers design their teaching approaches based on understanding who their learners are and their level of learning as a starting point. Then, the teacher uses effective questioning and scaffolding techniques to help learners grow in their level of understanding and learning.

Nurturing. The nurturing perspective stresses the caring nature of the teacher who makes an effort to provide a safe and a trusting atmosphere for learning. The nurturing teacher helps learners do their best through encouragement and support with clear expectation of what the learner has to achieve.

Social Reform. The goal of a Social Reform teacher is to encourage learners to take a more active role in building and maintaining a just society. The teaching approach uses class discussions to analyze and scrutinize common practices of society and suggests ways for change once a situation is deemed unacceptable.

Cross-cultural research. As stated by Byrne et al. (2009), research that compares groups from different cultures or nationalities can be considered as cross-cultural research.

Forward translation. When a document is converted from one language into another, the process is called forward translation.

Back translation. Back translation is a process of verifying the accuracy of a language translation procedure by getting the new language version converted back into its original language (Chapman & Carter, 1979).

Source language. The original language used as a starting point in any translation process is the source language. This is sometimes referred to as the first language of the translation process.

Target language. The second language in a translation process is also referred to as the target language.

Adaptation. In this study, adaptation is considered complete when the translated version is made appropriate for use in the new context and situation without altering its original intention.

Decentering. Decentering is a translation method that allows both language versions of an instrument to be modified during the translation process (Brislin et al.,1973). This method allows both the source and the target language to contribute to the final product of both language versions (Brislin, 1970).

TRAPD. TRAPD is an acronym for Translation, Review, Adjudication, Pretesting, and Documentation, which is a team approach to doing translation. The TRAPD approach was first developed by Harkness (2007) and employs a five-step process to translate an instrument. The first four are consecutive steps to convert the instrument from the source language into the target language beginning with a forward translation, followed by a review by an expert, whose work is

then viewed by the adjudicator to decide on the final version. Pretesting is then carried out to assess the translation outcome and the adjudicator may use the results of the pretesting to further modify the translation until it is considered ready for administration. Detailed documentations are carried out throughout the whole four steps, which is the strength of this approach.

Cognitive interviewing. Cognitive interviewing is a method to test and improve survey items during instrument construction. Cognitive interviewing attempts to make accessible the thinking processes that survey participants use to come to a decision to answer a particular item (Willis, 1999, 2005). This is usually carried out by asking participants to think out loud as they try to respond to an item. The interviewer can also ask probing questions to delve deeper into the cognitive processes of the participants to seek out the actual reasons behind the decision that was made.

Content equivalence. One of the five major notions of cross-cultural invariance as proposed by Flaherty et al. (1988) is content equivalence, which states that the items of a translated instrument remain appropriate for the target culture.

Semantic equivalence. Semantic equivalence means that the translated items maintain the same meaning in both the original and the target cultures (Flaherty et al., 1988).

Technical equivalence. The data collection method used must be the same for the original version and the translated version (Flaherty et al., 1988).

Conceptual equivalence. The instrument should be able to assess the same theoretical construct in both cultural groups (Flaherty et al., 1988).

Confirmatory factor analysis. According to Brown (2006), confirmatory factor analysis is a special type of structural equation modeling (SEM) used to test a measurement model based

on a theoretical foundation (Stevens, 1996) to ascertain the relations of variables to factors and between factors.

Measurement equivalence. Flaherty et al. (1988) view measurement equivalence as existing in five stages and suggest a five-stage approach to validating cross-cultural instrument equivalence. The five types of measurement equivalence are content equivalence, which confirms the consistency of the items to exhibit cultural applicability in both groups being measured; semantic equivalence, which supports the assertion that all items carry the same denotative or connotative meaning; technical equivalence entails that similar data gathering techniques produce analogous data for making comparisons; criterion equivalence guarantees the establishment of sameness in the way the variable is being interpreted based on the norms of both groups; and conceptual equivalence establishes that the same hypothesized concept is being assessed in each group.

Measurement invariance. Measures across groups are considered to be on the same scale if there is equality in terms of the structure of the construct as revealed by identical factor loadings, intercepts, residuals, and factor variances and covariances (Meredith, 1993).

Configural invariance. In measurement invariance testing, the step to assess whether the same basic factor structure is maintained in both groups under investigation is called configural invariance (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998). In other words, the model maintains the same number of factors and the same items remain relevant for exactly the same factor for both groups.

Metric invariance. On the other hand, metric invariance assesses whether the relationship between factors and items are the same for the groups being compared (Campbell, Barry, Jilliam,

& Finney, 2008). Metric invariance holds that the factor loadings are the same across the groups (Widaman & Reise 1997).

Scalar invariance. Scalar invariance tests whether groups understand the items in a similar manner (Byrne, 1998) where, according to Widaman and Reise (1997), the regressions of items on the latent construct have equal intercepts across groups.

Differential item functioning (DIF). When the same item works in different ways for different groups of people, it is said to exhibit differential item functioning (DIF). According to Zumbo (1999, p. 12), "DIF occurs when examinees from different groups show differing probabilities of success on (or endorsing) the item after matching on the underlying ability that the item is intended to measure."

Delimitations

Generalizations made from the findings have to take into account that samples are faculty members who are nested in their institutions and their individual departments. Further, due to the different locations of the institutions, some of which were in areas of the country (Malaysia) that are beyond the reach of the researcher, it was not possible to administer the surveys personally and therefore the researcher had to rely on the Internet to carry out the survey. Respondents had to be citizens of Malaysia and not expatriats working as staff of a university. This decision was made to ensure that participants from Malaysia provided data that would be representative of the views and beliefs of those who were native-born citizens of the country. This is crucial as beliefs about teaching are influenced by both personal and external factors (Pratt, 1992) and the sample must at least reflect a population that shares similar if not identical personal and collective experiences.

Limitations

As with other quantitative survey methods, this study has a number of limitations. The biggest concern using the online survey method was low rates of return. Samples from Malaysia were severely limited due to the different locations of the institutions and some were not easily accesible due to the distances. The researcher did not administer the survey personally. Even though a stratified random sampling was carried out, the limited sample size (n = 561) does not portray a representative sampling of the popluation. Interpretation of the data must be done with due care so as to avoid making sweeping generalizations from this study. In addition, Malaysians are usually exposed to multi-cultural environments, which may make it difficult to pinpoint exact causes of differences if found in the sample even after controlling for a number of personal and demographical variables. As a result, a closely-matched sample with identical characteristics and backgrounds with that of the U.S. sample was a challenge to obtain in order to facilitate a cross-cultural invariance testing of the TPI. This is another factor that has to be taken into consideration when discussing the findings.

Significance of the Study

Cross-cultural studies have seen unprecedented growth in recent years (Willis et al., 2010) and it is fast becoming a field in its own right. One of the benefits of doing cross-cultural research is that certain psychological theories and educational practices can be assessed to determine the extent some of the traits being investigated are universal or unique to a certain group of people. According to van Widenfelt et al. (2005), using established measures further allows for cross-cultural comparison of findings. Therefore, in order to make a statement about cross-cultural relationships, the instrument being employed in making that judgment must

function similarly across the two cultures of interest. At the time of this writing, studies on measurement equivalence of adapted research instruments used by researchers in Malaysia particularly in educational settings are few and far between. Cross-cultural research in Malaysia was found to be scarce as well (Fontaine & Richardson, 2003). By translating and adapting an instrument for use across cultures this study provides a tool that can extend our knowledge of beliefs about teaching across cultural backgrounds. This knowledge is particularly critical to educators in a globalized world such as the current times. This study also provides methodological insights into the methods and procedures used to translate and adapt the TPI and discusses if these methods worked well for a culturaly and racialy diverse country like Malaysia.

Chapter Two:

Review of the Literature

This chapter will begin by introducing the issue of improving practices in higher education and how cross-cultural sharing of ideas can be beneficial for all. This chapter will also describe what beliefs are and how they influence individuals' interactions with their environment. It will also discuss how systems of belief about teaching shape the way teachers conceptualize what they do in the classroom and how beliefs influence what happens in their classroom. Attempts to capture these conceptualizations of teaching are exemplified in three instruments that purport to measure beliefs about teaching and how they are realized in the classroom. The instruments are Gow and Kember's (1994) Orientations to Teaching Survey (OTS), Approaches to Teaching Inventory (ATI) developed by Prosser and Trigwell (1999), and the focus of this study, Pratt's Teaching Perspectives Inventory (TPI). Since the TPI is used to measure teachers' conceptualizations of teaching across different cultural and linguistic groups, it is deemed necessary to make the instrument more accessible by translating the instrument's items into a language that the Malaysian faculty know. This brings us to the discussion of translation and adaptation of instruments from one language to another and a set of procedures designed to come up with a cross-culturally equivalent instrument. To ensure that the instrument is functioning similarly across cultures, psychometric analyses have to be carried out and the last part of this chapter will discuss in detail the different types and levels of measurement invariance and the procedures required to evaluate measurement equivalence.

Teaching in Higher Education

In the last few decades, there has been renewed interest in different approaches to teaching and learning in higher education (Bowden & Marton, 1998; Laurillard, 2002; Ljubojevic & Laurillard, 2010; Merril, 2002). As early as in the 1990s, Barr and Tag (1995) observed that the long-established pattern of a higher education environment with strong emphasis on supporting teaching and instruction has shifted more towards providing support for student learning instead. The main focus has shifted from a teacher-centered approach to more learner-centered approaches (Fink, 2003; Kember, 2009; Reynolds, 2000). Many theories and assumptions have been put forth to describe and explain these different approaches, which include active learning (Bonwell & Eison, 1991), case studies (Merriam, 1998), the use of problem-based learning (Robinson, 1993), and other forms of teaching and learning. One impetus behind these activities has been pressure on improving the quality of teaching in higher education. Brancato (2003) pointed out that demands from society, organizations as well as from students have put pressure on institutions of higher learning to find ways to improve the quality and effectiveness of their instruction. This situation is not unique for a certain place or culture as many nations are struggling to improve the quality of their education system. This is especially so in a developing country like Malaysia, a country that aims to be the center of educational excellence in the Southeast Asian region with the enactment of the Malaysian Education Act of 1996 (Rahimah, 1998). The Act was ratified to establish some quality control on tertiary education in the country. Furthermore, the government has expressed keen interest that universities contribute more in the economic and social development and a study by a team of researchers from one of the more prominent public universities in Malaysia reported that the two

factors of faculty competence and teaching methodology employed were key predictors of business students' entrepreneurial potential (Zaidatul Akmaliah, Jamaliah, & Rahil, 2007). This underscores the importance of one of the key roles mentioned in the government document to meet the manpower demands of the nation with greater emphasis on science and technology. Though many of the universities are relatively new compared to those from developed countries, their contribution through research has been realized as of the utmost importance (Sharom, 1980). A desirable outcome of this surge of interest in the improvement of teaching practices in higher education is that knowledge gained from current research in other countries can be shared and learned by practitioners in the nation of Malaysia.

Teacher beliefs. Since research has shown that teachers' beliefs about teaching and the teaching process are related to their practices (Borrich, 1999; Clark, 1986; Clark & Yinger, 1979; Fang, 1996; Kagan, 1992; Pajares, 1992; Thompson, 1992), it is beneficial to explore in depth what these beliefs are and also how universal are they when comparing teachers across cultural boundaries. Belief systems in general influence how one views the physical and psychological reality (Rokeach, 1968). Beliefs help one to define and understand the world by screening, filtering, and reorganizing new ideas so that they fit with our prior knowledge (Pajares, 1992). Since attitudes and beliefs are usually formed over time, it is difficult and takes a considerable effort to change especially if time is of the essence. According to Rokeach's Belief System model, the more central a belief is in the central-peripheral dimension, the harder it is to change (Rokeach, 1968). Central beliefs are those perceptions that are usually formed early in life and which have stabilized over a long period of time. When these types of belief are changed, the more widespread the repercussions are in the rest of the individual's belief systems. Therefore, if

beliefs about teaching are similar in nature to other belief systems, most teachers' beliefs would remain quite stable over time and thus would be accessible for research.

Teachers possess perceptions of teaching and conceptions of what it means to teach (Pratt, 1992). These conceptions of teaching are deeply embedded in the personal schemata of the teacher and have been found to be extremely influential in the actual approach a teacher employs in the classroom such as choosing the kinds of materials and methodologies to be used to teach a particular subject (Kember & Gow, 1994; Trigwell & Prosser, 1996). Kagan (1992) provided a similar general description by stating that teacher beliefs are implicit and contain unconsciously held assumptions about learners, classrooms, and the learning points to be covered. Meanwhile, Collins et al. (2001) define teaching beliefs as a perspective, and state that a perspective is a set of beliefs and intentions related to knowledge, learning, and teaching.

With these different articulations about the nature of teacher beliefs, researchers have over the years developed different ways to measure these beliefs. Kagan (1990), in describing five alternative approaches to measuring teachers' cognitions, mentioned that one of the most direct methods used to assess teachers' beliefs are the short answer tests based on Likert-type self-report scales. Three of the most current self-report measures of teachers' beliefs are Gow and Kember's Orientations to Teaching Survey (OTS), Trigwell and Prosser's Approaches to Teaching Inventory (ATI), and Pratt's Teaching Perspectives Inventory (TPI).

One of the most influential attempts to summarize teachers' beliefs was carried out by Kember and Gow (1994) who developed an instrument, Orientations to Teaching Survey (OTS), which was designed to identify conceptions of teaching among teachers in higher education. The OTS is a questionnaire that consists of 46 items derived from interviews and is used to determine

orientations to teaching and their implications for the quality of learning that students experience. A nine factor model was posited and the 46 items were constructed to measure these factors, which were categorized into two broad categories or orientations to teaching referred to as "learning facilitation" and "knowledge transmission" (Kember & Gow, p. 61). The learning facilitation orientation includes subscales of problem solving, interactive teaching, facilitative teaching, pastoral interest, and motivator of students, while the knowledge transmission orientation comprises training for specific jobs, use of media, imparting information, and knowledge of subject. According to Gow and Kember (1993) these orientations of learning are related to three kinds of study approaches referred to as deep learning, surface learning, and achieving learning. The first study approach is deep learning, which involves intrinsic motivation on the part of the student. The second approach, which is surface learning, engages extrinsic motivation. The third study approach or achieving learning exceeds the engagement of motivation to include enthusiasm and a will to succeed as part of the approach. Kember and Gow (1994) also discovered that these orientations to teaching influence classroom practices. In addition to their own study, Kember (1997) also reviewed the accumulated findings of research on this subject matter and found that there were some variations in the use of terminology to describe the different conceptions. He observed that most studies seemed to share five common conceptions of teaching, which could be located on a continuum from a totally teacher-centered, content-orientated conception of teaching to a totally student-centered and learning-oriented conception of teaching. The fact that the different learning approaches were not mutually exclusive but exist on a continuum was further reinforced in a more recent study to debunk the myth that Asian students prefer passive learning and avoid more active learning styles (Kember,

2000). The study showed that rote-learning, which is a characteristic of surface learning, forms the basis of better understanding during the deep learning stage and this provided part of the evidence why Asian students outperform their peers in many academic fields.

Another attempt to capture teacher beliefs and related teaching practices was done by Trigwell and Prosser (2004) who proposed that teachers' conceptions about teaching and learning fall into a two-dimensional model consisting of intentions to teach and strategies employed to achieve the intentions. Influenced by Gow and Kember's (1991) study, Trigwell and Prosser developed a 16-item instrument in 1999 called the Approaches to Teaching Inventory (ATI) to explore university teachers' approaches to teaching based on the model. Eight of the items measure a conceptual change/student focus (CSSF) approach in which four items refer to the motive behind the approach while the other four to the strategy. The other eight items form a measure of information transmission/teacher-focused (ITTF) approach with four items specifically targeting intentions to transmit information and four to the use of a teacher-focused strategy to achieve that intention. The ATI has been used to collect data in more than 15 countries and across most disciplines normally offered at universities. Based on this instrument, they discovered that learner-centered teaching aimed at changing students' conceptions about the subject-matter led to higher quality of student learning and greater teacher satisfaction compared to a teacher-centered strategy with the purpose of transmitting information to students. While these findings based on the ATI suggested that there is one best way to teach, Pratt (1992), on the other hand, argued against that notion.

Teaching perspective. Similar to the argument proposed by Gow and Kember (1994) and Trigwell and Prosser (2004), Pratt introduced an idea that states that a teacher's conception about

teaching is influenced by a set of beliefs, which then determines his or her intentions and actions that are tied to knowledge, learning, and the teacher's role (Pratt, 1992). Pratt agreed with the findings in Kember's research, which showed that even though there are many variations in the teaching styles of faculty teaching in higher education there appears to be a limited number of ways to perceive teaching. He also believed that each belief or conception of teaching has its own strengths and weaknesses (Pratt, 2002).

Teaching Perspectives Inventory (TPI). Based on the premise mentioned earlier, an instrument called the TPI was developed by Pratt and later validated for research purposes with the help of Collins (Pratt & Collins, 2001). Pratt introduced the notion of teacher perspectives specifically for teaching adult learners. Unlike Trigwell and Prosser's two-dimensional model, Pratt's model conceptualized teaching to generally fall into five common perspectives that are dependent upon three dimensions of actions, intentions, and beliefs. Pratt (1992) developed a general theory of teaching that was based on the premise that learning and teaching usually occur in a particular context. This model of teaching, as explained by Pratt (2005), specifies that learning usually occurs as the teacher, the learners, and the subject area or content interact with each other and this takes place within a specific context influenced by the beliefs and values of both the teacher and the learners. The type of teaching perspective and learning that occurs are influenced by how much emphasis is placed on the three different components in the learning process. If the belief is that the teacher-content relationship is most important and students' understanding of content is the goal, then the transmission perspective is endorsed.

This model of teaching has been validated by years of observations and interviews from teachers in adult and higher education that provided data on how teachers perceive the act of

teaching (Pratt, 1992).With the assistance of his graduate students, Pratt analyzed a large amount of data that have been gathered and based on their analysis, they identified five distinctly different perspectives or views of what teachers do and why they do what they do. These perspectives were labeled Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform. Each perspective is composed of a set of beliefs, intentions, and actions. According to Pratt and Collins (2001), these eight features (i.e., five perspectives, beliefs, intentions, and actions) will jointly help researchers and practitioners organize and classify narratives about how teachers differ in approach and justification of their teaching. The features are also claimed by them to provide a means by which educators could articulate their approach in order to reflect meaningfully on their teaching and ponder upon possible improvements.

The Teaching Perspectives Inventory (TPI) is an online survey instrument that contains 45 items. According to the instrument developers, faculty can use the inventory as a self-evaluation tool of their teaching skills and style and assist them to reflect on their personal beliefs and values about teaching (Pratt et al., 1998). The items are grouped under five perspectives: Transmission (lecture and teacher-centered); Apprenticeship (experiential and coaching-oriented); Developmental (facilitation and learning-centered); Nurturing (focused on building learners' self-esteem); and Social Reform (change the status quo oriented). Figure 1 presents the factor model underlying the TPI. For each of the five perspectives, the items are further divided into the three subcategories of teachers' classroom practice, their organization of the learning situations, and their beliefs about teaching and learning. These subcategories are called Beliefs, Intentions, and Actions. Beliefs pertain to conceptions that determine what is to be taught and what evidence will be accepted that the knowledge has been taught successfully.

Actions, on the other hand, are defined as those activities that are described as routines and techniques used to engage people in the content of the teaching. Meanwhile, intentions are viewed as general statements that point toward an overall agenda or purpose about teaching. As shown in Table 1, questions for Actions typically ask for what is done when instructing or teaching. For Intentions, the questions focus on what is being accomplished in the instruction or teaching. Belief questions address issues related to beliefs about instructing or teaching. The TPI yields numerical scores on each of the five perspectives, as well as three sub-scores within each of these perspectives that describe respondents' Beliefs, Intentions, and Actions.

Development of the TPI. The TPI emerged out of a phenomenological study where qualitative descriptions of teachers' views on what teaching is all about were gathered by Pratt (1992); interviews and observations on groups of educators from various teaching institutions were used to obtain these descriptions. From the data that were collected, Pratt and his colleagues categorized these concepts into perspectives and labeled them Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform. Later on, the TPI was developed as a self-administering inventory that was eventually put online and automatically scored; results of respondents' individual profiles of their conceptions about teaching were reported via email. From the wealth of qualitative data collected about the instrument, Pratt et al. (2001) later delved into ways of refining and rephrasing teachers' endorsements of different statements that reflected their dominant teaching perspectives and distinguished them from non-dominant or recessive viewpoints. The study also focused on validating the scores from the instrument to ascertain whether the inventory demonstrated acceptable standards of reliability and validity. From this streamlining process, more than 200 items were constructed. After further refining, the items

were reduced to 120 items. In the final process, the items were trimmed down to 75 items to reflect a simultaneously balanced representations of Beliefs/Intentions/Actions and learner-teacher, learner-content, and teacher-content relationships (Chan, 1994).

Psychometric analyses were carried out by Chan (1994) on the 75 items that resulted in a further revision that reduced the number of items to the current 45-item version with nine defining statements per perspective (see Appendix A). Each perspective is represented by three Belief statements, three Intention items, and three Action declarations. The sample items in terms of actions, intentions and beliefs are shown in Table 1.

Currently, the TPI has been translated into eight languages including Spanish, French, German, Portuguese, Mandarin, Korean, Japanese, and Indonesian Malay. To date, more than 125,000 people are reported to have taken the TPI and nearly 1000 of the respondents were from Malaysia who responded using the English version (Collins, email to author, February 22, 2010). Research on beliefs of teachers from Malaysia can add to the cross-cultural research on teachers' beliefs and practices but in order to conduct this research, it is necessary to have an instrument written in Bahasa Malaysia (BM). In addition, in order to make cross-cultural comparisons it is critical that the measurement properties of the instrument used to make comparisons are invariant across countries.

The instrument employs a 5-point Likert scale where the belief items are measured with options ranging from strongly disagree to strongly agree (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree) while the intention and action items are scales ranging from never to always (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always). Collins

Table 1

Teaching Perspectives Inventory Sample Items

Section	**Focus**	**Examples**
Actions	What do you do when instructing or teaching?	1. I cover the required content accurately and in the allotted time. 2. I link the subject matter with real settings of practice or application. 3. I ask a lot of questions while teaching.
Intentions	What do you try to accomplish in your instruction or teaching?	17. My goal is to demonstrate how to perform or work in real situations. 21. I expect people to master a lot of information related to the subject. 30. I want to make apparent what people take for granted about society.
Beliefs	What do you believe about instructing or teaching?	32. To be an effective teacher, one must be an effective practitioner. 36. Teachers should be virtuoso performers of their subject matter. 38. Teaching should focus on developing qualitative changes in thinking

(Pratt et al., 2001, p. 2)

and Pratt (2011) reported that the means of the summary scores from the individual scales (potential range for the nine item scale is 9 to 45) varied from scale to scale with Transmission averaging at 33.1 (*SD* = 4.6), Apprenticeship 36.2 (*SD* = 4.2), Developmental 34.5 (*SD* = 4.3), Nurturing 36.7 (*SD* = 5.0), and Social Reform 28.8 (*SD* = 6.0). The higher means obtained for the Apprenticeship and Nurturing scales prompted the instrument developers to claim that items of the two scales seemed to be more attractive to respondents as compared to those items of the other three scales.

Psychometric properties of the TPI. A measurement instrument must yield reliable and valid scores if it is to be used by practitioners and researchers. Estimating reliability is possible

through many different methods and the results obtained from these methods will yield varying estimates of reliability. Test-retest reliability is one such method, which refers to the extent that the scores on the same measurement correlate with each other on two different administrations. Correlation coefficient between scores on the two occasions can be calculated to obtain data on the stability of the test scores or observations over a period of time (Crocker & Algina, 1986). Cronbach's alpha can also be obtained to see if the item scores within a measure are internally consistent. For Collins and Pratt, opportunities to carry out a test-retest reliability study were made possible due to the fact that many respondents attempted the TPI survey more than once. They reported that some people came back some time later during the day to retake the survey, while others did it a few weeks or even two years apart (Collins & Pratt, 2011). As a result, they were able to gather test-retest data from 500 respondents and discovered that the overall reliability was .67 with individual scale scores ranging from .62 for Developmental to .71 for Social Reform. These results indicate reasonable stability over time. An even greater stability in scores was reported for a much longer period of time in the second and third administrations with a sample of 63 people showing an average correlation of .73 with an individual subscale correlation of .65 for Nurturing, which was the lowest, and .87 for Social Reform which was the highest.

Internal consistency measures of reliability pertain to methods that are concerned with the consistency of scores within the test itself (Crocker & Algina, 1986). The most common method of assessing internal consistency reliability estimates is through the use of coefficient alpha and the most widely used measure is Cronbach's coefficient alpha. In terms of consistency of the TPI scores, Collins and Pratt (2011) reported relatively high reliability coefficient for the five scales

where Developmental was the lowest showing Cronbach's alpha of .70 and the highest being Social Reform with .83 while the average for all the five scales was .76. Correlations between subscales were found to be quite low at .15 between Transmission and Nurturing to moderate at .58 between Apprenticeship and Developmental while the average correlation for all the five subscales combined was .41. The reported alpha reliabilities for the TPI's components of Beliefs, Intentions, and Actions were .72, .78, and .80, respectively, with an average of .77 overall.

Scores obtained from a measurement instrument must not only be consistent but must also accurately measure what it is supposed to measure. Since content validity is concerned with whether or not the items on a given instrument accurately reflect the theoretical domain of the latent construct it claims to measure, the items must effectively demonstrate that they are representative of all the possible questions that could have been derived from the construct (Crocker & Algina, 1986). The way that the TPI was developed, starting from the phenomenological study via interviews and observations (Pratt, 1992) all the way through the refining and rephrasing of items and then two years later a streamlining process pruning 200 items down to the current 45 by Chan (1994), supports the representativeness and content validity of the TPI items.

Construct validity, on the other hand, is concerned with the ability of the measurement instrument to actually assess the conceptual variable that it is meant to measure. Information for this type of validity is gained from many sources of evidence including evidence based on the internal structure of the instrument and through description of its relation to other variables (Crocker & Algina, 1986). Some of the more common ways that researchers can obtain evidence related to the internal structure of an instrument is by carrying out exploratory factor analysis

(EFA) or confirmatory factor analysis (CFA) techniques. EFA techniques provide information about the factor structure of the instrument by statistically demonstrating how items load on particular factors while confirmatory factor analysis techniques assess how well a theorized model fits the data. A principal component analysis (PCA) was carried out on the TPI instead of an EFA. Findings supported a five-factor model as the optimal model and each rotated factor accounted for roughly the same fraction of variance (i.e., one factor was not more dominant than another) (Collins & Pratt, 2011). It was reported that each item was correctly assigned to its proper scale and factor scores correlated highly with scale scores. Other than that, none of the items were reported to have communalities of less than .30 indicating that all 45 items contributed meaningfully to the defining of one or another of the perspectives. Nine factors with eigenvalues greater than 1.00 which accounted for 50.7% of the common factor variance were extracted through principal component analysis. It was reported that extractions of 9, 8, 7, 6, 5, and 4 factors were examined and rotational strategies involving both oblique and orthogonal rotations were tested over the course of the inventory development. Collins and Pratt (2011) discovered that the most reasonable rotational strategy was a quartimax rotation incorporating all 45 items that loaded on one of the five factors; there were no items that loaded on more than one factor.

Collins and Pratt (2011) reported that there were strong positive correlations between the scale and factor scores. The average correlation between scale and factor scores was .83. Correlations of each perspective's scale with factor scores for Transmission were .90, .66 for Apprenticeship, .77 for Developmental, .94 for Nurturing, and .88 for Social Reform. Each TPI item also showed a stronger positive relationship with its parent scale than with any of the scales,

which led the TPI developers to conclude "that scale items converge on their respective underlying concepts, but diverge from the latent continua of the other perspectives since the rotated factors are both orthogonal and roughly equal-sized" (Collins & Pratt, 2011, p. 12). Therefore, the construct validity of the TPI scores is supported based on these sources of evidence.

Even though the TPI has been used in many studies and validation work has been carried out and reported by Pratt and Collins (2011), so far no confirmatory factor analysis (CFA) of the full factor structure of the instrument has been reported to assess model fit. Based on the description of the instrument by Pratt (1992), the measurement model that seems to capture the relationships of the items and their factors involves a five factor structure as shown in Figure 1, even though there is mention of other underlying constructs such as beliefs, intentions, and actions for each perspective and composite scores are also available for each one. However, when a full CFA of the TPI was carried out, it was discovered by Brown and Lake (2006) that the five-factor model was not an acceptable solution. They carried out an analysis on a four-factor model with only 11 selected items instead. The first factor was renamed Apprenticeship-Developmental which had three items (I link the subject matter with real settings of practice or application; My intent is to help people develop more complex ways of reasoning; Teaching should focus on developing qualitative changes in thinking). Nurturing was the second factor which also had three items (I encourage expressions of feeling and emotion; My intent is to build people's self-confidence and self-esteem of learners; In my teaching, building self-confidence in learners is priority). The third factor, Social Reform, also had three items (I help people see the need for change in society; Individual learning without social change is not enough; I expect

people to be committed to changing society). The last factor in the Brown and Lake's model was Transmission with two items (My intent is to prepare people for examinations; Effective teachers must first be experts in their own subject). Brown and Lake reported that the inter-correlated four factor model revealed a good fit, $\chi 2$ (76) =541.1 (TLI = .88; CFI = .91; RMSEA =.066), and similar patterns were found for three groups of teachers with one group from New Zealand (*n*=235) and the other two were primary (*n*=784) and secondary teachers (*n*=614) from Queensland, Australia. The current study, however, tested the five-factor model with 45 items (see Figure 1).

Cross-cultural Research, Translation and Adaptation of Instruments

Validity of an instrument developed in one country or culture must also be established for another culture before it can be used for making any score comparisons. Cross-cultural validity is

extremely crucial when doing research that attempts to compare results from two countries (Wu, Li, & Zumbo, 2007). Cross-cultural research may be viewed as flawed (Chapman & Carter, 1979; Douglas & Nijssen, 2003) as a result of borrowing instruments for research without checking their relevance and equivalence in other countries and contexts. Many worthwhile studies were deemed to be flawed as researchers in most developing countries tended to borrow instruments outright due to lack of funding (Chapman & Carter, 1979). Even when attempts are made to adapt the instrument, the changes made were simply in the form of translation from the original language to the language of the target population (Swaine-Verdier et al., 2004). An adapted or translated instrument does not guarantee that the adapted or translated version measures the same constructs as the original one does due to cultural and linguistic

differences. Test adaptation is not simply literally translating item content from one language to another (Geisinger, 1994). Besides, translation for cross-cultural research is a complex process

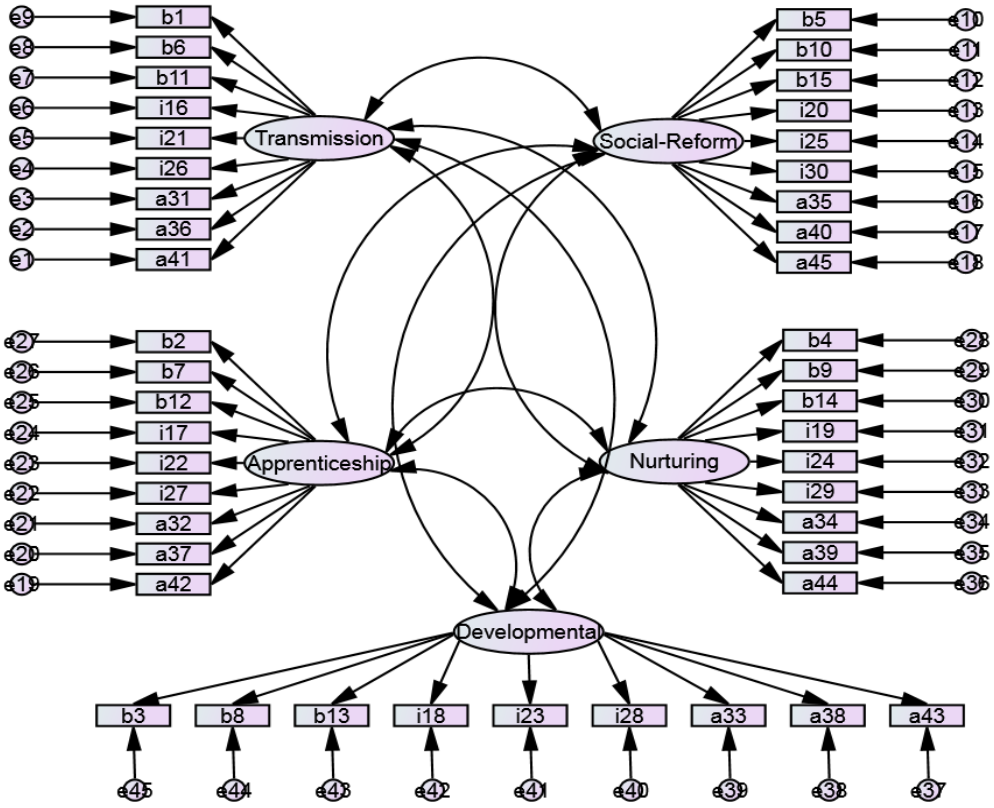


Figure 1. Five-Factor Structure of the TPI.

and translation errors can introduce measurement errors, which may result in conveying different meanings across cultures (Brislin, 1970; Heine et al., 2002). There are several issues that must be

addressed before undertaking the task of adapting an instrument from one socio-cultural and linguistic context into another.

Measurement biases. According to van de Vijver and Hambleton (1996), there are three types of bias that may affect the performance of a measurement instrument, namely construct bias, method bias, and item bias. Construct bias is essentially the issue of non-equivalence of constructs across cultural groups. Some constructs that are the norm in one culture may not exist in others. Even if they do, they might be perceived differently particularly in language terms as declared by the Sapir-Whorf hypothesis. The hypothesis claims that it is language that shapes thoughts about the world and how the experiences in it are interpreted and the ways that a language is used to organize thoughts vary to a certain extent from culture to culture (Whorf, 1956). A study by Burns and Brady (1992) provided support for this claim when they reported that Malaysian college business students' need for uniqueness in expressing innovative behavior was not equivalent to the way their counterparts in the U. S. perceived it. Method bias, on the other hand, is attributed to the administration procedures of the measurement. It may stem from the socio-cultural forces that influence response patterns or the physical conditions in which the instrument is administered. Meanwhile, item bias is often a result of inadequate translations. An item may be biased if it contains item content or language that is differentially familiar to subgroups of examinees, or if the item structure or format is differentially difficult for subgroups of examinees (Hambleton & Rogers, 1995).

Many experts have argued that when tests are adapted from the language and culture in which they were developed to another language, the measurement equivalence of the adapted instrument should be assessed (Budgell et al., 1995; Geisinger, 1994; Swaine-Verdier et al.,

2004; Sireci et al., 2005; van Widenfelt et al., 2005). They reasoned that the original and adapted instruments may not be equivalent because the meaning intended by the original test items may have been accidentally lost in the translation process and/or the test items may not have the exact equivalent form in the target language. In some cases, the construct being measured may not be perceived as equivalent at all by the target group as it was originally intended. This may happen in psychological research where some forms of attitudes or behaviors are viewed positively in one culture but negatively in others. For example, giving an intimate kiss to a spouse in public is acceptable in most western countries but is frowned at in Asian countries like Malaysia. In other research situations, the construct being measured may involve elements that do not exist in a particular cultural setting. It is hard to extract a response from an ordinary Malaysian adult about most technological terms because equivalent forms have yet to be developed. In this case, we have to establish whether the instrument is amendable to adaptation or whether the construct is culture-specific or more universal in nature. We need to know if the instrument can be used in other cultural contexts beside the one that it is intended for. According to AERA/APA/NCME Standard 13.4 "When a test is translated from one language or dialect to another, its reliability and validity for the uses intended in the linguistic groups to be tested should be established" (AERA, APA, & NCME, 1999).

In the case of a multi-racial and multilingual country like Malaysia, comparability of results across languages within the same border is another issue not to be taken lightly. Standard 9 (AERA, APA, & NCME, 1999) discusses the topic of test takers who are of non-mainstream linguistic backgrounds and consequently, tests should be written to be equivalent across linguistically diverse populations. This principle is clearly evident in Standard 9.2, which

requires that, when possible, test developers should study the application of their tests in different linguistic subgroups. By the same token, researchers must establish that their instrument behaves equivalently when attempted by different language groups. Collins and Pratt (2012) made no mention of the equivalence of the TPI across language groups but mentioned briefly that native speakers of English in their study scored slightly lower than non-native speakers of English on the Social Reform scale. Their study also showed small inter-group differences but they were construed to be not strong enough "to suggest the presence of scale bias" (Collins & Pratt, 2011, p. 13).

Procedures for instrument translation and adaptation. There are many suggestions and guidelines as to how best to translate instruments for cross-cultural research (e.g., van de Vijver & Hambleton, 1996; Hambleton & Patsula, 1999; Hambleton, 2001; Harkness et al., 2004; McGorry, 2000). Developers of the TPI (Collins, email to author, Feb 22, 2010) recommend the use of their 24-step process to translate the instrument into other languages (see Appendix B). Van de Vijver and Hambleton (1996) provided a list of 22 guidelines describing recommended practices in test translations formulated by the International Test Commission. The guidelines cover four major areas. The first one is context that spells out the basic principles of multilingual studies. The second aspect is development that includes recommended practices in developing multilingual instruments. The third domain is administration that describes issues in instrument administrations. The fourth domain is documentation and score interpretation, which is related to interpretation and cross-cultural comparisons of scores. This study used these guidelines in the translation and adaptation of the TPI.

According to McGorry (2000), based on the observations from their study, it is recommended that the following six steps be employed to ensure accurate cultural and linguistic revision of a survey:

1. Use a blend of at least two or more translation methods with an emphasis on the decentering method.
2. A minimum of two translators must be employed.
3. If the researcher lacks the necessary proficiency in the target language, a translator must be present alongside while collecting data.
4. To enable identification of difficulties or challenges with the back translation, acquiring immediate feedback during and after data collection is critical.
5. Randomly investigating surveys after data collection can also assist in identifying issues such as misinterpretations or presence of missing data.
6. Obtaining and scrutinizing basic statistical data such as distribution patterns or item analyses must be carried out before any advanced data analyses can proceed.

Effectiveness of translation and adaptation procedures. There have been translations in numerous countries world-wide with several conducted in Malaysia. The questionnaires translated into Bahasa Malaysia (BM) or Malay have measured various constructs including self-concept (Khoo et al., 2008; Mohammad Aziz Shah et al., 2013, Musa, Fadzil, & Zain, 2007; Nur Fazidah, 2012; ; Quek et al., 2002; Swami, 2012). Reports of the translation procedures in most of these studies were rather sketchy and some studies reported the researcher as one of the few translators involved in the process; no substantial evidence was given to show credibility of the final version. Besides, relatively few translation studies (Tan, 2005; Watkins & Ismail, 1994)

have been carried out in the field of education particularly for a developing country like Malaysia which has a multicultural and a multiracial population that spends the biggest portion of its national budget on education (Khader, 2012). According to the United States Department of State website (2010), Malaysia has a population of 28.3 million people where 53.3% are Malays, 26% Chinese, 7.7% Indian, and the remaining 1.2% belong to other minority groups. The Malays speak a variety of Malay dialects while the Chinese and the Indians speak a number of their own respective regional dialects. Many of the minority groups are natives of the states of Sabah and Sarawak, which are situated across the South China Sea on the island of Borneo, and speak a multitude of languages and dialects. Originally, English was the language of communication across these multi-racial boundaries but the government of the day has replaced it with Bahasa Malaysia even though English is still widely used among the older generations of Malaysians. For the purpose of this study, only native speakers of Malay were chosen. This meant selecting only those who were from the Malay race or other ethnic groups that use Malay as their first language.

Evaluation of Measurement Equivalence Across Cultures

The TPI is an instrument that contains several scales that are intended to measure teaching beliefs or perspectives. Each specific scale is made up of multiple items or subscales. Researchers have used the TPI in samples that vary by gender, culture, race as well as age. According to Steenkamp and Baumgartner (1998) evaluating the appropriateness of abstract notions developed in one country and extending them to other countries is an essential step in establishing the generalizability of these notions. However, if we wish to make a generalization about teaching beliefs across different cultural groups, it is imperative that we fulfill the critical

assumption that the instrument functions the same way regardless of the difference between groups. In other words, the instrument has the quality of measurement invariance (MI). If that assumption holds, then comparisons and analyses of those scores are acceptable and yield meaningful interpretations. But if that assumption is not supported, then such comparisons and analyses do not yield meaningful results.

Types of measurement invariance (MI). There are many different types of MI and along with that, several ways of testing for invariance have been developed. A main concern of measurement invariance is that measures across groups are considered to be on the same scale if relationships between the indicators or items used to measure the latent trait are the same across groups (Meredith, 1993). Such a definition of measurement invariance would require the equality of item factor loadings, item intercepts, and item residual variances. What this essentially means for this study is that teacher beliefs must be associated with the same set of items in each culture. As in the case of the TPI, all items associated with each scale must be the same across both the U.S. sample and the Malaysian sample. Furthermore, Cheung and Rensvold (1998) reiterated that the relationships between the construct and the items, as represented by factor loadings, must not be significantly different or must be invariant across cultures. In other words, the factor loadings for all items in the Transmission scale must be similar for both countries and the same goes for the other four scales of Apprenticeship, Developmental, Nurturing, and Social Reform.

Measurement invariance testing process. Multigroup confirmatory factor analysis (MCFA) is one of several statistical approaches that has been used to evaluate measurement invariance. If the confirmatory factor analysis (CFA) model fits well for both samples, more invariance testing is carried out to ascertain if the measure is functioning similarly for both

groups. Based on the theoretical conceptualizations of the TPI, we would expect the five-factor model to fit each population. Model fit can be assessed using a variety of measures including the X^2 statistics and descriptive fit indices recommended by Hu and Bentler (1999), such as the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Byrne, Shavelson and Muthén (1989) distinguish two types of invariance namely, measurement invariance, which is invariance of item intercepts, item factor loadings, and item error variances, and structural invariance, which includes invariance of the variances and covariances of the latent variables. Widaman and Reise (1997), on the other hand, put forth the idea of four levels of measurement invariance, which are classified as configural invariance, metric invariance, scalar invariance, and strict factorial invariance. Their notion of invariance levels, which was employed in the current study, is basically forcing progressively more stringent forms of equality constraints on parameters in the measurement model to observe if the parameters are indeed equal.

Configural invariance. The first level of measurement invariance is configural invariance or pattern invariance, which states that the pattern of salient and nonsalient loadings is the same across groups (Horn & McArdle, 1992; Widaman & Reise, 1997). It entails that each group being compared has the same number of factors with the same pattern of fixed and free parameters. However, at this stage no equality constraints are imposed on the model. The model is deemed to exhibit patterns that are similar but not identical. This measurement invariance level is a prerequisite for the other invariance tests. Should a model display a non-invariant pattern, cross-groups comparisons are pointless as the latent traits may be viewed differently by different groups based on the dissimilar endorsements of the observed variables (Cheung & Rensvold,

2002; Horn & McArdle, 1992; Little, 1997; Vandenberg & Lance, 2000). In this case, the same number of factors and items are forced to load on the same factors but the parameter estimates are free to be different across the two countries (Steenkamp & Baumgartner, 1998) as shown in Figure 1.

Metric invariance. The second level, according to Widaman and Reise (1997), assumes that the loadings are equal across groups. It is also referred to as metric invariance. Not only are the same items forced to load on the same factors for both countries but the factor loadings are constrained to be equal across the two groups. According to Steinmetz et al. (2008), metric invariance is not only concerned with construct comparability as pointed out by Steenkamp and Baumgartner (1998) but also with similarity of construct meaning across groups. They proposed that metric invariance is essential to make inferences that the construct has the same meaning regardless of group differences primarily due to the fact that it can provide evidence about the equality of validity coefficients. Metric invariance tests whether the United States and Malaysian university faculties interpret the items of the TPI in the same way (Byrne, 1998) and the conceptions of teaching beliefs carries the same meaning for people in both groups. Once evidence of configural invariance has been established, metric invariance testing can begin. In this procedure, a referent item is usually chosen to set the metric for each factor. This referent item must be invariant across the two samples. The whole process is completed by using all the other items on the subscale as a temporary referent item so that the target item remains invariant across samples (Cheung & Rensvold, 1999). In other words, besides having the same number of factors and items loading on the same factors, the factor loadings are constrained to be equal across the two groups. If the factor loadings are equal then there is evidence of measurement

equivalence in terms of metric invariance. Factor loadings are established to be unequal when the regression lines are not the same and the slopes are different, which essentially means that there is a lack of metric invariance as shown in Figure 2. No further invariance testing is recommended beyond this point if the two groups responded differently to the item regarding that particular construct.

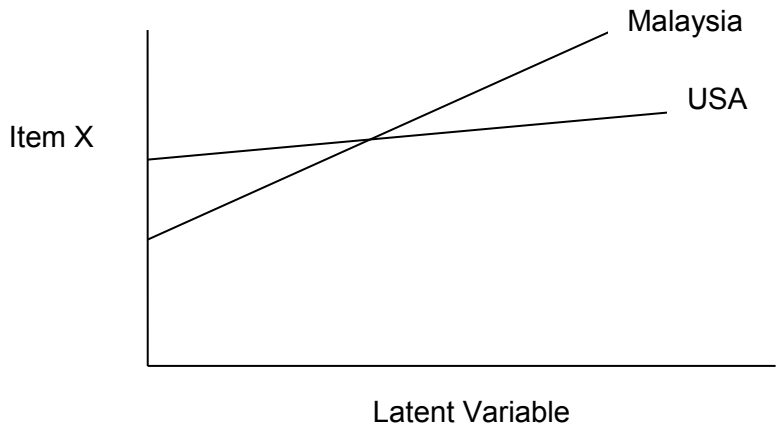


Figure 2: Unequal Factor Loadings

Scalar invariance. If the metric invariance is supported, the next hypothesis to be tested is scalar invariance. Scalar invariance is the third level that tests for equality of item intercepts of the regression of items on the latent construct for both groups (Widaman & Reise, 1997). Essentially, it points to invariance of the item intercepts in the regression equations that link the indicators to their latent construct. Experts like Hayduk (1989) observed that item intercepts can be interpreted as systematic biases in the responses of a group to an item. Scalar invariance is only evident if the item intercepts are not significantly different across groups.

When both the factor loadings and the intercepts are constrained to be equal across the countries and they are found to be the same for both groups, evidence for strong factorial or

scalar invariance has been successfully established. Figure 3 shows scalar invariance where the intercepts are not significantly different for the two countries.

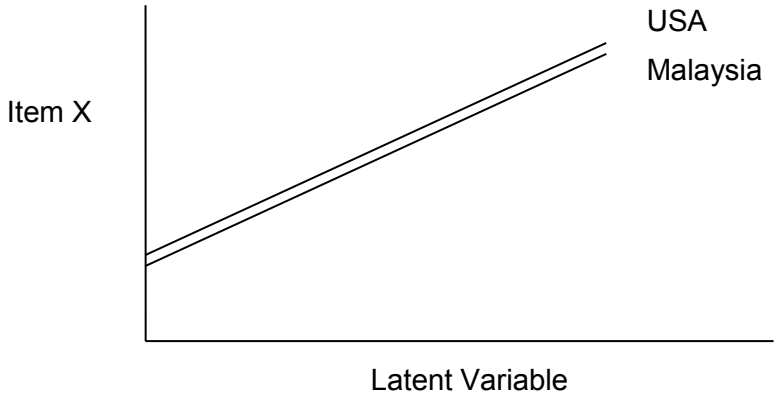


Figure 3: Equal Intercepts

On the other hand, Figure 4 demonstrates a lack of equivalence where the item intercepts of the regression of items on the latent trait are unequal for the two countries. This can be interpreted as differential item functioning (DIF) which violates measurement invariance as described by Meredith (1993).

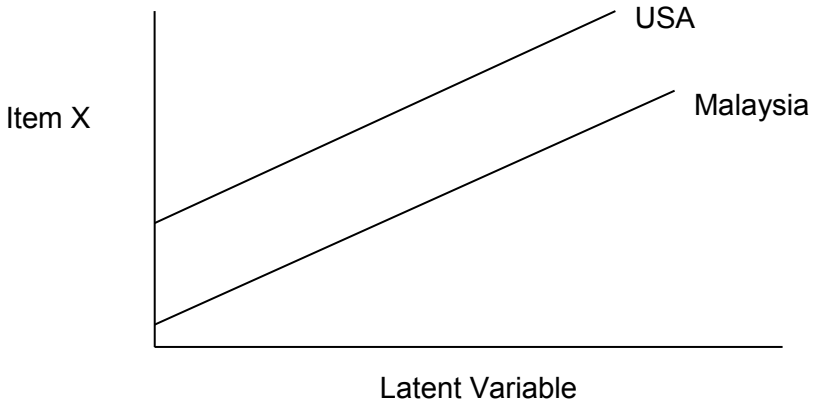


Figure 4: Unequal Intercepts

Strict factorial invariance. Widaman and Reise's (1997) fourth level is called strict factorial invariance. It extends the previous models by bringing into play additional constraints. Strict factorial invariance is a highly constrained model that includes invariance of item error variances, invariance of factor variances/covariances, and invariance of latent means.

Invariance of item error variances. Invariance testing of error variances of the items on the TPI involves additional constraints on the measurement model. Here the factor loadings and variances of the latent variables have to be equal across groups. If this is so, then the error variances can be interpreted as equivalent to the reliability of the indicators. According to experts, the test of invariant error variances checks to see that the measurement error in the construct is the same in all groups (Cole & Maxwell, 1985; Steenkamp & Baumgartner, 1998). If the factor loadings and variances of the latent variables have been shown to be equal, then the error variances can be interpreted as equivalent to the reliability of the indicators.

Invariance of factor variances. This type of measurement invariance is present when groups have the same variances in their respective latent constructs (Steenkamp & Baumgartner, 1998). In the present study, this test of invariance involves evaluating the equality of each of the five latent variable's variances (e.g., Transmission) across the two countries.

Invariance of factor covariances. The equality of the associations (covariances) among the latent variables across groups can be tested following the tests of the five latent variances. Since the TPI has five latent variables there are 10 covariances involving these latent variables (e.g., covariance between Transmission and Apprenticeship). According to Cronbach and Meehl (1955), covariances among constructs have implications for the constructs' meaning or validity.

Unequal covariances have raised concerns among experts about the equality of construct meanings (Cole & Maxwell, 1985). Little (1997) noted that equal factor covariances have implications for the comparability of constructs when viewed by groups that are different in many ways.

Invariance of latent means. Analysis of the latent means for the five latent variables requires scalar invariance, in which the factor loadings and intercepts of the manifest variables are invariant across groups (Meredith, 1993). To test for invariance of latent means, the factor means for one group are fixed to zero for the purpose of achieving identification of the model. The group whose means is constrained to a value of zero serves as the reference group.

Full versus Partial Invariance. Full measurement invariance and especially strict factorial invariance may often be difficult to obtain and sometimes can be proven to be impractical in real life situations. As a cautionary note, Yoo (2002) pointed out that in cross-national or cross-cultural research, the failure of any level of factorial invariance might occur because of the instrument, population, or both. In other words, both the TPI as an instrument and the differences between the U.S. and the Malaysian groups can contribute sources of failure to achieve measurement invariance. Furthermore, a perfectly invariant instrument is an elusive goal (Cheung & Reinsvold, 1999) due in part to the possible different interpretation of the meanings of standard scales across nations (Riordan & Vandenberg, 1994). Meaning is often embedded in a variety of contexts and situations such that the same item in a scale can be viewed in many different ways due to different backgrounds and prior knowledge. Besides, Yoo (2002) aptly stated that a matched sample with identical characteristics and backgrounds is hard to get across nations. The U.S. is comprised of people from diverse ethnic and racial backgrounds while the

Malaysian sample used in this study is from one group of people who speak Malay as their mother tongue. However, it must be made clear that closely-matched samples are important in controlling extraneous variables and will help in identifying which source is more responsible for the failure of invariance. In view of this, some researchers have suggested that partial invariance is a more reasonable compromise when testing an instrument for measurement equivalence (Byrne et al., 1989). Partial invariance is achieved when some model parameters are invariant while others are allowed to vary across groups (Byrne et al., 1989).

Summary

Regardless of the many different views about what teaching is, a fundamental goal for research is to define conceptions of teaching in order to better understand and, therefore, evaluate teaching with the ultimate goal of improving teacher performance. Faculty often adopt certain types of teaching styles, which to them are the best in helping them be effective as educators, and carry with them a set of teaching repertoires to mirror this fact. According to Pratt (1992), there are many perspectives to teaching and there is not one best way to teach. However, each perspective can be made to work so that teaching is effective after considering contexts and content of teaching. According to Kember and Kwan (2000), unless faculty challenge their beliefs about teaching, critical transformations to the quality of their teaching and student learning may not be possible. Therefore, it is imperative for us to learn more about what these beliefs and assumptions are so that we can use the knowledge to develop a more effective training and development regime to help faculty improve their practice. In a globalized world where physical and geo-political boundaries are fast disappearing, such knowledge can be shared among people all over the world quickly and effectively if socio-cultural and linguistic

boundaries can be overcome as well. The rekindled interest in different approaches to teaching and learning in higher education that has left its mark in many western nations will ultimately reach the shores of other countries and the lessons learned will definitely benefit them if they can be absorbed into the culture easily. This can only happen if we can evaluate the cross-cultural generality of our theories and assumptions. Research on the process of translating and adapting instrument has identified a number of challenges to successfully implement this process (Solano-Flores et al., 2009). Experts have voiced concern that even following the rigorous standards of the International Test Commission, establishment of full measurement invariance may be difficult (Byrne et al., 2009; Poortinga, 1995; van de Vijver & Leung, 2000).

Chapter Three:

Methods

For this study, a two-stage process was employed to achieve the purposes of the study. The first stage was the translation and the adaptation process of the Teaching Perspectives Inventory (TPI) using a combination of procedures suggested by van de Vijver and Hambleton (1996) and McGorry (2000) to come up with a matching instrument in Bahasa Malaysia (BM), the standard Malay dialect used in Malaysia. The second stage was the administration of the adapted instrument to a large sample of faculty teaching at all 20 government-funded universities of Malaysia. The results from the Malaysian sample were compared to those obtained from the data from a similar number of faculty in the U.S. who took the original English version of the TPI. This chapter begins with a review of the purposes of the study, followed by a description of the procedures involved in the two stages of this study. The presentation includes an explanation of the translation and adaptation procedures and a description of the translators involved for the first stage. For the second stage there is a description of the participants from the two countries along with a discussion of the original and the adapted versions of the TPI. The last part of the chapter is a description of the data collection procedures and statistical analyses that were carried out.

Purposes

The purposes of the present study were to translate and adapt Pratt's (1992, 2001) Teaching Perspectives Inventory (TPI) from English into Bahasa Malaysia (BM) and evaluate the measurement invariance of the TPI between Malaysian and U.S. faculties. The two groups being investigated were faculty teaching in universities from the United States and Malaysia. Extraneous differences in interpretation of the TPI were kept at a minimum by ensuring that both versions shared as much similarity as possible in terms of formatting, instructions, and response options. However, simply translating from one language into another does not ensure accurate cultural and linguistic equivalence because the translation itself may threaten invariance. To ensure that this did not happen and to guarantee accurate cultural and linguistic equivalence, procedures and guidelines for translating and adapting instruments as proposed by van de Vijver and Hambleton (1996) and McGorry (2000) were followed. This included (a) an initial (forward) translation, (b) a backward translation, (c) an expert panel review, (d) a pilot study that involved administering the instrument and computing test-retest reliability, and (e) cognitive interviews. The next stage was to address the issue of measurement invariance, which means making sure that the measures across groups are considered to be on the same scale (i.e., relationships between the items and the constructs are the same across groups). The following two research questions guided the measurement invariance testing of the TPI:

1a. How well does the correlated five-factor structure of the TPI fit the data from the faculty from the U.S.?

1b. How well does the correlated five-factor structure of the TPI fit the data from the faculty from the Malaysia?

2. Is the correlated five-factor structure of the TPI invariant across the U.S. and Malaysian faculty samples?

Stage I: Adaptation and Translation

Initial translation. The initial translation (forward translation) of the TPI instrument from English to Bahasa Malaysia (BM) or Malay was carried out by three translators who were native speakers of Malay and who had been identified as competent users of the English language. Brislin (1970) reported using only one forward translator while another study employed two forward translators (Wang, Lee, & Fetzer, 2006); both studies, however, reported inadequacies in the outcome of the translation. Harkness (2003), who proposed the use of the Translation, Review, Adjudication, Pretesting, and Documentation (TRAPD) approach to improve translation procedures, has recommended two qualified translators. Based on this information as well as recommendations by McGorry (2000), the researcher chose to use three forward translators. All three translators were faculty members from the Language Studies Center of a public university in Malaysia. They were recruited after recommendations from their department head and were chosen based on their academic and professional qualifications. All had experience doing translation work for the university even though some were more experienced than others. The translators were not paid but an incentive in the form of refreshments was provided at the discussion meeting with fellow translators and the researcher. More details regarding the qualifications and characteristics of the forward translators are provided in Chapter Four. A letter of recruitment was sent to all the translators to seek their permission to be the translators for this study (see Appendix C).

These forward translators translated the TPI independently and any attempts to communicate with each other were strongly discouraged. After they had completed the translation of the instrument, they met once for two hours to discuss the best possible translation that conformed to the original intent of the instrument. The researcher was present as a facilitator to ensure that the discussions were on task and the goal of a Malay version of the TPI was achieved. The consensus to accept an item as being faithful to the original English version was based on the fact that everyone agreed that it was equivalent; when there was a disagreement with an item, the item was scrutinized further to look for the source of the contention. Many items were accepted this way. For items that were difficult to agree on, the researcher chose what was deemed the best among all the possible choices.

Back translation. The Malay version of the instrument, which represented the combined form from the three forward translators, was translated back into English by two bilingual translators who were not part of the initial translation team. The same method of obtaining the forward translators from the department head's recommendations was used to recruit the back translators but the criterion for selection was relaxed to include individuals who were fluent in both languages but they did not need to be native speakers of Malay. While Brislin (1970) and Wang et al. (2006) both employed only one back translator, this study employed two to ensure precision of translation in the final product. One was a non-native speaker of Malay but both were competent users of the language and had an almost native-like competency in using English. Like the forward translators, the back translators were not paid but were given a similar incentive. The back translators did not see the original instrument to ensure that that they were not influenced by it nor were they informed about another back-translator working on the same

instrument. This was to avoid any attempts to communicate with each other. The two translations were put in a table along with the original English version to be used as a checklist for evaluation of the translated items.

Expert panel review. At this stage, an expert panel met to discuss and evaluate the translated instrument based on the adaptation and evaluation checklist. More details regarding the checklist are provided in Chapter Four. The expert panel consisted of five members along with the researcher who met as a group to discuss the translation for four hours with an hour break in between two, two- hour sessions. The same procedures in selecting the translators were used to obtain the five panel members. The five expert panel members were made up of four native speakers of Malay while one was a non-native speaker but fluent in both Malay and English. They were all academics teaching languages and linguistics in a public university of Malaysia and each one had more than six years of teaching experience. Rubio et al. (2003) recommended that the number of panel experts should be around 6 to 20 participants to be adequate. The more experts there are, the more information is generated about the measures. This study decided to use the minimum number possible (five panel members plus the researcher) so that greater interaction and more in depth feedback from the panel could be efficiently achieved. During the expert panel review meeting, members evaluated the instrument by providing their suggestions and revisions with the researcher acting as a facilitator as well. The panel compared the original items in English and the back-translated versions in order to validate the accuracy of the translation in the BM version. This was to ensure that there were no mistranslations, missing texts, and other translation errors. Flaherty et al. (1988) proposed five-criteria in validating cross-cultural instrument equivalence. Criteria included: (1) *content*

equivalence to ensure that the content of each item in the instrument has consistent cultural relevance, (2) *semantic equivalence*: to ensure that the meaning of each item remains conceptually and idiomatically the same, (3) *technical equivalence*: to ensure that the methods of data collection (interviews, observation, or self-report) elicit comparable data, (4) *criterion equivalence*: to establish the normative interpretation of the item, and (5) *conceptual equivalence*: to ensure that the same theoretical construct is being measured in each culture. For this study, the panel assessed only content, semantic, criterion, and conceptual equivalences. According to Solarno-Flores et al. (2009), translation error is multidimensional because the task of translating involves broad categories as mentioned by Flaherty et al. (1988) above and the fact that languages encode meaning in different ways. A perfect one to one correspondence in translation is a lofty goal to achieve between languages especially if the languages are from two very different cultures. Solarno-Flores et al. (2009) introduced the idea that a translated item is either acceptable or objectionable. An item can be viewed as objectionable if it has a few but severe errors, too many mild errors, or many severe errors. If the errors or discrepancies are not too severe, the item is acceptable with revisions. As such, each panel member was given an adaptation and evaluation checklist based on the four criteria to evaluate each item as acceptable or objectionable.

Pilot testing. The next step in the adaptation process of the TPI from English to BM was a pilot study to assess reliability based internal consistency and the test-retest method. Initially, all lecturers from the 20 Malaysian public universities were contacted via email with a letter of consent (see Appendix D) indicating that their participation in this phase of the study was designed to examine the test-retest reliability of the TPI. A link to access the adapted TPI was

also given to them to complete the survey online. A period of two weeks or more in between administration was used for this study. The means all the items were calculated and correlation coefficients were computed based on the administrations at the two time points. Reliability measures for each of the five TPI perspectives were also obtained and used as a basis for comparing the test and retest scores. This was to establish whether the items were performing equivalently during the two administrations. Internal consistency reliability was also assessed for the five perspectives at each time point.

Cognitive interviews. The next step to ensure quality in the translation process was conducting cognitive interviews. Cognitive interviewing is a technique originally developed during the 1980's by survey methodologists and psychologists for testing and improving items during the questionnaire-design process of a survey project (Willis, 1999, 2005). The overall goal of cognitive interviewing is to make explicit the cognitive processes that respondents use to answer questions, which normally are hidden and unobservable to public view. The aim is to reduce misinterpretation and confusion created by misbehaving items included on the survey instrument, which will then improve the quality of the data. The two major methods of cognitive interviewing are think-aloud interviewing and verbal probing techniques. The think-aloud method of interviewing is carried out by the interviewer who reads each question to the respondent, and then keeps a record of the way the respondent arrives at an answer to the item (Willis, 1999, 2005). The verbal probing method, on the other hand, involves the interviewer exploring deeper into the respondent's answers with specific questions to uncover the reasons behind the response given (Willis, 1999, 2005). After the respondent attempts the question or item, the interviewer then proceeds with a series of questions pertaining to the answer given until

a sufficient amount of data is obtained regarding the decision making process that is involved in the production of the response. Further revisions to items are made when the cognitive interviews reveal that items are not behaving similarly across individuals participating in the interviews or that they are not functioning the same way as the original instrument. For this study, the participants involved in the cognitive interviews were chosen from the pilot study group. A purposeful sampling method was used to choose five from the group (three female faculty members and two male faculty members from Universiti Malaysia Sarawak). First contact with the interviewees was by electronic mail (see Appendix E) by the investigator to obtain their permission before making an appointment to meet face to face. Since the data for the test-retest were confidential, all identification codes of the other participants were deleted. For this study, the researcher as the interviewer conducted each cognitive interview by starting each meeting with a short training session to demonstrate the steps in the cognitive interviewing process.

Stage II: Measurement Invariance Testing

Participants. The two groups being investigated in this study were faculty members teaching in higher education in the United Sates and Malaysia.

U.S. sample. Data from 605 respondents from the United States were provided to the present researcher by the TPI developers. The TPI developers provided item responses from their existing database from 605 respondents who were similar to the Malaysia sample in terms of faculty gender, percent time teaching as part of their work assignment, types of students taught (e.g., undergraduates), and years of experience teaching. All cases were de-identified (no names, no e-mails, no institutions and no cities or states mentioned) so respondents' anonymity

could be guaranteed. Information about the characteristics of the U. S. sample is provided in Chapter Four.

Malaysian sample. A total of 565 Malaysian faculty members made up the sample drawn from 20 public universities, which closely-matched the number of faculty in the U.S. sample. Participants were only those who were citizens born in the country and who were native speakers of the Malay language. Once the total number of respondents reached above 500 people, the researcher stopped the data collection. A stratified sampling approach was not possible due to the lack of responses from the major universities like University of Malaya and University Science Malaysia despite additional reminders and requests for participation. As a result, the sample gathered did not represent the true distribution pattern of faculty members among all public universities in Malaysia. Information gathered included item level scores and demographic information such as faculty gender, age, workload, years of teaching, primary role, educational level, and usual learners taught. More demographic variables compared to the original TPI were gathered from the Malaysian sample, such as highest academic degree earned, academic rank, tenure status, academic college or school, ethnicity and language background, so that it was possible to explore if these demographic variables were related to the five TPI teaching perspectives. Information about the characteristics of the Malaysian sample is provided in Chapter Four.

The Instrument

The Original English version. The Teaching Perspectives Inventory (TPI) is an online survey instrument that contains 45 items. Faculty can use the inventory as a self-evaluation tool of their teaching skills and style to assist them to reflect on their personal beliefs and values

about teaching (Pratt et al., 1998). The items are grouped under five perspectives: Transmission (lecture and teacher-centered); Apprenticeship (experiential and coaching-oriented); Developmental (facilitation and learning-centered); Nurturing (focused on building learners' self-esteem); and Social Reform (change the status quo oriented). For each of the five perspectives, the items are further divided into the three subcategories of teachers' classroom practice, their organization of the learning situations, and their beliefs about teaching and learning. These subcategories or 'manifestations of commitments' are labeled as Beliefs, Intentions, and Actions (Pratt et al., 1998). As shown in Table 2, questions for Actions typically ask for what is done when instructing or teaching. For Intentions, the questions focus on what is being accomplished in the instruction or teaching. Belief questions address issues related to beliefs about instructing or teaching. The TPI yields numerical scores on each of the five perspectives, as well as three subscale scores within each of these perspectives that describe respondents' Beliefs, Intentions, and Actions. The TPI uses a 5-point Likert scale: 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Neutral*, 4 = *Agree*, 5 = *Strongly Agree* for the belief items and 1 = *Never*, 2 = *Rarely*, 3 = *Sometimes*, 4 = *Often*, 5 = *Always* for the Intentions and Actions items (Pratt & Collins, 2010). The scores for each perspective range from 9 up to 45. As mentioned earlier, each perspective is made up of three commitment categories and each commitment category across the five perspectives has 15 items with scores that can range from 15 to 75. An inventory total across all 45 items for each respondent could vary from 45 to 225.

The online TPI format was recently upgraded by the authors and is divided into five sections. The first section is a welcoming page where the contents in the section are displayed.

Table 2

Sample Items for the Transmission Perspective

Section	Focus	Examples
BELIEFS	What do you believe about instructing or teaching?	6. Teachers should be virtuoso performers of their subject matter. 11. Effective teachers must first be experts in their own subject.
INTENTIONS	What do you try to accomplish in your instruction or teaching?	16. My goal is to prepare people for content-related examinations. 26. I want people to score well in the exams as a result of my teaching.
ACTIONS	What do you do when instructing or teaching?	36. My teaching is governed by the course objectives. 41. I make it very clear to people what they are to learn.

On this page, respondents are also given the option to take the inventory in English, Spanish, or Chinese. The second section is a webpage that shows instructions that explain the procedures for taking the inventory as well as a section for respondents to provide contact details to receive feedback on their responses. Another subsection is provided to get respondents' affiliation detail because some institutions use the TPI to assess their own personnel. Also, the final subsection was added to ensure that the respondents stayed focused on the specific course or group of learners they were teaching and not the general public at large. This was in response to the query made by the researcher during the pilot study which showed that participants were inconsistent with their response as they shifted focus on the subject area and students they taught. This is discussed in detail in Chapter Four. The actual instrument page only begins when the "next" button is clicked to show the 15 items about different educational beliefs followed by 15

more items assessing intentions, then, another 15 items regarding actions and finally, 10 questions about demographic details of the respondents. On the last page of the instrument, respondents have to click submit to get their scores calculated and their dominant perspectives identified. Details about how to interpret scores are provided in the third section of the TPI website. Here, a brief summary of the five teaching perspectives and ways to understand the scores obtained from the instrument are specified. For the purpose of this study, the focus was on the five perspectives to see if their factor structures and item properties were equivalent across the two groups being studied.

Data Collection Procedures

Malaysian sample. Data collection for the Malaysian sample was conducted using an online survey where respondents were asked to complete the adapted Teaching Perspectives Inventory (TPI) via the Internet and at the same time respond to questions regarding demographic variables. Prior to that, a letter was obtained from the Malaysian Ministry of Higher Education to begin the research and carry out data collection among the faculty members of the 20 public universities (see Appendix F) stressing that the participants must only be those faculty members who are native speakers of Malay. This was also done to ensure an optimum response rate. All participants were contacted through their university administration office by mail (see Appendix G). As requested in the letter, the administration personnel in charge would then, forward an email to all their native speakers of the Malay faculty inviting them to be a voluntary participant in this investigation (see Appendix H). In the same email, participants were informed of the study's purposes as well as to provide them with the required "Informed Consent" information. Individuals who agreed to participate provided their "Informed Consent" in

compliance with the IRB approval, accessed the TPI by clicking on the easy web-link to the survey, and responded to the items. As a preparation to address any concerns from the participants during the data collection period, researcher contact information was also provided for the participants.

Approximately one month after the first email, a follow-up email was sent thanking participants and politely providing others with a reminder of the approaching deadline for participation. Additional reminders (see Appendix I) were sent to the administrators in order to increase the number of respondents, particularly to those universities that lacked the number of respondents required. Even though one university replied with a letter sending a whole list of emails of their Malay faculty members, direct contact between the researcher and the respondents was kept to a minimum by reminding the university that the survey was anonymous and participants' identity must not be revealed to the researcher. Three universities emailed a list of the faculty email addresses but were also informed of the importance of protecting the privacy of faculty members during the data collection. There were, however, some direct contacts with participants who responded to the survey but were still receiving reminders and wished to be taken off the mailing list. These requests were forwarded to their respective university administrators.

U.S. sample. Among the more than 100,000 who took the TPI online over the past 10 years, the researcher requested an equivalent number of respondents from the TPI database maintained by Pratt and Collins (2001) to match the number gathered for the Malaysian sample. Respondents from the U.S. group were matched as closely as possible to their Malaysian counterparts based on gender, primary role or function, highest academic degree, academic

major, years of experience instructing, educating, or teaching, and subjects or specialties taught. Potential participants may have other duties and responsibilities but teaching had to be their central role. They also had to teach undergraduate courses in an institution of higher learning and be a full time faculty member in their institution similar to their peers in Malaysia. Permission was granted by Pratt for the researcher to use the data in the TPI database for the U. S. sample (see Appendix J). The researcher contacted Drs. Pratt and Collins to request that they randomly select the 600 individuals from all the U.S. respondents in their database who took the survey from 2010 to 2012 so that the participants were closely matched to the time that the adapted TPI was administered to the Malaysian group. Pratt and Collins were also informed that all names and other personal identifiers had to be removed from the data to protect their privacy before they were sent to the researcher.

Data Analyses

Treatment of missing data. The issue of missing data and how to deal with it has been a common problem in statistical analysis but it has become more important in recent years as researchers have become more aware of its impact on research findings (Acuna & Rodriguez, 2004; Little & Rubin, 2002; Rubin, 1976; Schafer, 1997). In order to understand the issue of missing data, we must begin with the question of why data are missing in the first place as some types of missingness are ignorable while others are non-ignorable (Rubin, 1976). Types of ignorable missingness are those omitted data that are known as missing completely at random (MCAR) and those that are missing at random (MAR). MCAR cases are instances where missing data appear to follow no discernible pattern in their missingness nor are they related to any of the other variables being studied (Acock, 2005). MAR data, on the other hand, are those missing

cases that are somehow related to a variable in the data set but they are not the focus of the study (Allison, 2001). Data that are missing not at random (MNAR) are non-ignorable because there is a pattern in which they occur and this pattern may have a bearing on the results of the study and influence the interpretation of the findings.

For most MCAR and MAR situations, the methods often used by researchers are deleting instances containing at least one missing value of a feature. This works well if the sample size is large enough to compensate for the lost. If sample size is an issue, then pairwise deletion is another course of action to take where the respondents will not be deleted from the whole analysis but for those variables that he or she is not responding to, they will not be included (Howell, 2012). Another method is by substitution or imputation where values are plugged in to replace missing data. This is especially useful in cases where missingness is non-ignorable like MNAR. The simplest method of imputation is carried out by substituting missing values with the mean but just like deletion methods, imputation methods have been reported to be inadequate in dealing with missingness (Schlomer, Bauman, & Card, 2010) and in fact, are claimed to be biased (von Hippel, 2004). Besides, more sophisticated imputation methods can be handled by most statistical software such as SPSS (IBM Corp., 2012) and Mplus (Muthén & Muthén, 1998-2012). Most of IBM's SPSS Base software use the deletion and imputation methods mentioned above besides the more refined method called expectation maximization (EM). However von Hippel (2004) cautioned that even though the EM produces unbiased estimates under some conditions it is limited to point estimates only. According to Schafer (1997) special features available in Mplus software can be utilized to examine multiple data sets and Muthén et al. (2003) have assured that non-ignorable missing data modeling is possible using maximum

likelihood estimation procedures. Since this study employed Mplus 7.1 (Muthén & Muthén, 1998-2012) to carry out measurement invariance, missing data were handled using full information maximum likelihood (FIML) where according to Muthén & Muthén (1998-2012), imputation of missing data values for each individual is only done after each parameter has been estimated directly. A simulation study done by Enders (2001) revealed that FIML was better at handling missing data assumed to be MAR (missing at random) as compared to deletion, and imputation methods by having less bias and less variability in its sampling.

Descriptive statistics. The data were analyzed to ascertain the distribution pattern, the measure of central tendency and the dispersion of the study variables from the samples from the two countries. Study variables included demographics such as gender, primary roles, percentage of teaching hours, usual learner groups, years of practicing area of expertise, and highest academic level. These variables were collected to compare the composition of the two samples and to assist in the investigation of how the instrument functions between the two countries. For the Malaysian sample, additional variables were examined such as current age, race and language groups, academic college, and institutional affiliation. These variables were useful in trying to understand the population distribution of the sample better. Distributions of item responses were analyzed and displayed in table form. Reliability analysis based on Cronbach's alpha and item-to-total correlation was used to examine the relationship of the items within their respective factors (i.e., perspectives).

Factor structure invariance. Based on the theoretical conceptualizations of the TPI, a five-factor model was the expected model to fit each population. Maximum likelihood estimation with robust standard errors (MLR) was employed in estimation, and the covariance matrix of the

45 items was used as input for the parameter estimation. The data from the Malaysian faculty represented a nested data structure with faculty responses nested within institutions. Intraclass correlation coefficients (ICC), indicators of the degree of dependence in the data within institutions, were calculated for the TPI items (see Appendix K). The ICCs ranged from .017 (i20) to .066 (b9) with a mean of .040 (median = .040). One approach that Mplus uses to handle nested data is to use Type = Complex, which computes standard errors using a sandwich estimator. This approach was used with the Malaysian sample and the results were compared to the same analyses, not taking into account the nested data structure. The results were very similar and therefore the decision was made not to use Type = Complex. One rationale behind this decision was that the data from the U.S. sample did not have identifiers for institutions and therefore it was not possible to determine if there were multiple faculty respondents from institutions. Without knowledge of the institutional affiliation of the U.S. faculty it was not possible to calculate the ICCs for the TPI items in the U. S. sample. In view of the comparative purpose of this study it was decided to use maximum likelihood estimation with robust standard errors for both groups, but not use the Type = Complex, to insure comparability of the analyses.

Model fit assessments were based on the X^2 statistics in conjunction with descriptive fit indices recommended by Hu and Bentler (1999) including the comparative fit index (CFI), standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). Values of approximately .08 or below for the SRMR, CFI values of approximately .95 or above, and RMSEA values of approximately .06 or below would suggest adequate model–data fit (Hu & Bentler, 1999).

Measurement invariance. The prerequisite of testing for measurement invariance is to evaluate whether the same general factor structure holds for both groups. CFA was conducted on each sample independently to ensure that the same basic factor structure fits the data for each sample. This was to evaluate if the same structure of the TPI still fits the data when the two groups are assessed separately. In this case, the five-factor model in Figure 1 must fit the data from both groups equally well and the fit was deemed acceptable based on the X^2 statistics and descriptive fit indices such as the CFI, RMSEA, and SRMR.

When the CFA model was shown to fit well for both samples, additional invariance tests were conducted. Steenkamp and Baumgartner (1998) have made a recommendation that the order for tests of invariance should start with configural invariance, followed by metric invariance, scalar invariance, and then the strict factorial invariance such as invariance of the item error variances, factor variances, factor covariances, and finally latent mean invariance.

Configural invariance testing was conducted to test whether the same basic factor structure holds for the two groups. It was a prerequisite for the other tests. At this level of invariance testing, evidence to support whether the same number factors and the same items are relevant for each factor across groups was examined. In this case, the same number of factors and items were forced to load on the same factors but the parameter estimates were free to be different across the U.S. and Malaysian samples.

Upon obtaining confirmation of configural invariance, metric invariance was tested. Metric invariance is concerned with construct comparability and similarity of meaning about the construct as viewed by two different groups. This was done to test whether United States and Malaysian university faculties interpret the items in the same way (Byrne, 1998) and whether the

construct carried the same meaning for them. In this procedure, a referent item was selected to set the metric for each factor. This referent item had to be invariant across the two samples. The whole process is completed by using all the other items on the subscale as a temporary referent item so that the target item remains invariant across samples (Cheung & Rensvold, 1999). This was done until all items in all factors were tested. In other words, besides having the same number of factors and items loading on the same factors, the factor loadings are constrained to be equal across the two groups. If the factor loadings were found to be equal then there is evidence of measurement equivalence in terms of metric invariance. If the factor loadings were unequal, there is proof of non-equivalence. Follow-up comparisons were conducted to identify which specific items were different between countries.

Once metric invariance is supported, the next hypothesis tested is scalar invariance. When both the factor loadings and the intercepts are constrained to be equal across the countries and they are shown to be equal, this essentially supports the evidence of strong factorial or scalar invariance.

Once there is evidence for metric and scalar invariance, the next step is to look at invariance of factor variances. To carry this out, the five factor variances of the TPI perspectives are constrained to be equal. The test assesses possible differences in homogeneity of variance of the latent variables in the groups (Steenkamp & Baumgartner, 1998). If the factor variances are found to be equal, then it is fair to say that the two groups of faculty have the same variances in their respective latent constructs.

Since Little (1997) has pointed out that unequal factor covariances may have implications for the comparability of the constructs, a test of invariance of the factor covariance has to be

carried out next. This is usually done by constraining the factor covariances to be equal. If they are found to be equal, then there is equality of construct meanings shared by the two countries.

To test for invariance of latent means, the factor intercepts for one group are fixed to zero for the purpose of achieving identification of the factors. The U.S. sample served as the reference group (latent means fixed to zero) and the estimated mean of the Malaysian group represented how much the Malaysian group's latent means deviated from the U.S. group's means.

The last procedure for the invariance testing was to test the hypothesis of invariance of item error variances. Here the factor loadings and variances of the latent variables have to be equal across groups. If this is so, then the error variances can be interpreted as equivalent to the reliability of the indicators. The results of the measurement invariance are displayed in a table format to show the different fit indices chosen.

Partial measurement invariance. Obtaining full measurement invariance based on the procedures described above is often not met in practice. Most comparisons of group differences rely on traditional analyses that assume full invariance of intercepts and loadings which are frequently unrealistic. This has often led to situations where minor violations of these assumptions increase the risk of drawing erroneous conclusions. Byrne et al. (1989) propose the use of partial measurement invariance as a sufficient requirement for assessing measurement equivalence. Their argument for this position is that partial invariance is a compromise between full measurement invariance and complete lack of invariance. In general, partial invariance, unlike full invariance, allows some factor loadings or intercepts to differ across groups. In view of these arguments, this study was open to the option of employing partial measurement invariance to assess the TPI's performance across the two groups.

Summary

This chapter reviewed the purposes of the study and described the methods and procedures used in this study. The description of the methodology encompasses two stages beginning with the translation and adaptation procedures where each step of the process was explained along with the participants involved. The participants were translators, expert panel members, and the pilot study group. Five of the pilot study group members were shortlisted to participate in the cognitive interviews that followed. After the presentation of the first stage, a thorough account of the psychometric evaluation procedures of the TPI was given. Details regarding participants from the U.S. and Malaysia were provided along with a description of the original and the adapted versions of the TPI followed by a report of the data collection procedures and statistical analyses.

Chapter Four:

Results

The two main purposes of the study were to translate and adapt Pratt's (1992, 2001) Teaching Perspectives Inventory (TPI) from English into Bahasa Malaysia (BM) and to evaluate the success of the translation in achieving measurement invariance. The TPI is an instrument that looks at five differing perspectives of teaching and learning with each perspective consisting of teaching beliefs, intentions, and behaviors of teachers in higher and adult education. To accomplish these tasks, a two-stage process was utilized. The first stage was the adaptation of Pratt's Teaching Perspectives Inventory (TPI) based on the suggestions proposed by van de Vijver and Hambleton (1996) and McGorry (2000) to produce an instrument that is similar but in Bahasa Malaysia (BM). The adaptation process involved initial translations, back-translations, an expert panel review, a pilot study, and cognitive interviews. After the translation process, the second stage, which was a psychometric investigation of the TPI, was initiated to address the issue of measurement invariance. This included a confirmatory factor analysis of both the English and the Bahasa Malaysia versions that were used to address the first research question of how well the correlated five-factor structure of the TPI fit the data from both the U.S. and the Malaysian faculty samples. The assessment of the cross-cultural equivalence of the two versions by means of invariance testing was also performed to answer the second question of whether the correlated five-factor structure of the TPI was invariant across the U.S. and Malaysian faculty samples. A large sample of faculty teaching at the 20 government-funded universities of

Malaysia was obtained by administering the adapted version of the TPI via an online survey. A sample of similar size was obtained from the TPI's database looking at faculty in the U.S. who took the original English version. This chapter is organized into two main sections whereby each section presents results obtained from each of the two stages mentioned above.

Stage I: Adaptation and Translation

Initial translation. The initial translation (forward translation) of the TPI instrument from English to BM was carried out by three translators who were native speakers of Malay and who had been identified by the researcher as competent users of the English language based on their qualification as shown in Table 3. These forward translators translated the TPI independently and any attempts to communicate with each other were strongly discouraged. After they had completed the translation of the instrument, they met as a team together with the researcher to discuss the best possible translation that conformed to the original intent of the instrument.

Reflections of translators. Reflections on the translation process were gathered as part of the task that each translator had to carry out on the TPI. Based on the reflections, as reported by Initial Translator 1, the translation was accomplished in 25 minutes in only one sitting. The translator did not find any items that were too difficult or too challenging to translate. This was attributed to the fact that this particular translator was very experienced in doing translation work.

> "Translation was done based on my capability to communicate in both Malay and English languages." (Translator 1)

Table 3

Qualifications of Translators

Demographics and Qualifications	Forward Translator 1	Forward Translator 2	Forward Translator 3	Back Translator 1	Back Translator 2
Gender	Female	Female	Female	Female	Male
Race	Malay	Malay	Malay	Malay	Bidayuh/ Chinese
Native language	Malay	Malay	Malay	Malay	Bidayuh/ Malay
Bahasa Malaysia	College Level	College Level	College Level	Malaysian Certificate Exam (SPM)	Malaysian Certificate Exam (SPM)
English Language Proficiency Level	College Level	College Level	Malaysian Certificate Exam (SPM)	College	College
Educational level	Ph.D.	Ph.D.	MA	MA	Ph.D.
Field/Major	Socio-linguistics	Pragmatics	Malay Linguistics	TESL	English Literature
Training in psycho-metrics/Research Methods	Yes	Yes	Yes	Yes	Yes
Experience in instrument construction and development	Yes	Yes	Yes	Yes	Yes
Other relevant experience	Qualitative Research	Qualitative Research	Qualitative Research	-	Qualitative Research

Note. SPM = Malaysian Certificate of Education; TESL = Teaching English as Second Language.

The second Initial Translator, on the other hand, faced some difficulties in doing the translation. The translator noted that the task took about six hours on three separate occasions to complete the translation of all the items. The reason for this was lack of time due to teaching and

other workloads. In addition, several TPI items such as items b2, i20, i23, i28 and a32 were especially challenging when attempting to find words that were similar in terms of meaning as shown in the context of the sentences used. Lack of proficiency in English was cited as another source of this challenge even though in her qualification she stated she had college level English classes. Unlike the first initial translator, the second one had no formal training in doing translation. A third source of the problem was the unique differences present in both the languages, which made it difficult to find a direct one-to-one translation of the items. The fundamental differences that existed between English and BM prevented a word for word translation. There were many BM forms that fit the same English word while in other cases there were English words that did not have any equivalent form in BM at all. For these challenges, the second initial translator adopted a number of problem solving strategies to address these challenges. For those items containing words that had no equivalent in BM, the translator interpreted the meaning of the whole sentence in its context first before proceeding to search for suitable words to express the same meaning and context. There were times where suitable extra words were added to ensure that the sentences were grammatically correct.

A specific example put forth by the translator was item i20 as shown in Table 4. There were three alternatives to choose from as a way to translate the item.

i20. My intent is to challenge people to seriously reconsider their values.

Table 4

Alternative Translations to Item i20

Alternatives		Literal Translation	Meaning
i.	*Tujuan saya ialah untuk mencabar individu supaya mereka secara serius menghargai nilai-nilai yang ada.*	Aim my is to challenge individual so that they with seriousness appreciate the values that exist.	My aim is to challenge the individual so that they will seriously appreciate all the values that exist.
ii.	*Tujuan saya ialah untuk mencabar individu secara serius menghargai nilai mereka**	Aim my is to challenge individual with seriousness appreciate their values.	My aim is to challenge the individual seriously to appreciate their values.
iii.	*Tujuan saya ialah untuk mencabar individu menghargai nilai mereka secara serius.**	Aim my is to challenge individual to appreciate their values with seriousness.	My aim is to challenge the individual to appreciate their values seriously.

The decision to choose Alternative i was made based on the second initial translator's opinion that Alternatives ii and iii were inaccurate and the sentences were not grammatically correct. So, the decision to add a few extra phrases was made to complete the sentence structure. The rationalization for this course of action was that in doing a good translation, one cannot depend solely on equivalent words available but at times will need to add other relevant words because a concept expressed in one word in a language cannot be translated as a single word in another language and still maintain the same meaning.

The third translator did not submit any reflections but did provide input to resolve issues brought up by the other two translators when they all met and completed the translated Malay version as shown in column 3 of Table 7.

Back translation. The Malay version of the instrument was translated back into English by two bilingual translators who were not part of the initial translation team (see Table 2). The first back translator was a native speaker of Malay while the other was a non-native speaker of Malay; both were competent users of the language. The first back translator had a post-graduate degree in Linguistics from an Australian university and had an almost native-like competency in English while the second back translator had a Ph.D. in English Literature from a British university. The back translators were not shown the original instrument so that they were not influenced by it nor were they informed about another back-translator working on the same instrument. This was to avoid any attempts to communicate with each other. The two translations were put in a table along with the original English and the Malay translation to be used as a checklist for evaluation of the translated items (see Table 7).

Reflections of the back translators. The two back translators worked at getting the translated items back into English and the process was carried independently from each other. Even then, they each faced similar challenges. The meaning of certain items in the Malay version was quite difficult to ascertain and looking for an equivalent word in English was even harder. The first back translator resorted to guessing as part of her strategy (see Table 5 for an example, Item a38).

The second back translator left many items incomplete whenever faced with difficulties in doing the back translation. The translator put forward three main explanations for the difficulties. The first one was the challenge to maintain the original meaning and sentence structure as compared to the original form where items such as b1, b3, b6, and b15 that had problematic words or phrases were left as blank lines in the sentences as shown in Table 7.

Table 5

Translation of Item a38

Original Version	Malay Version	Back translation
I challenge familiar ways of understanding the subject matter.	*Saya mencabar pendekatan pengajaran yang lazim untuk memahami sesuatu bahan pengajaran.*	I challenge the common teaching approaches to understand teaching materials.

Secondly, the translator could not make a decision as to which word or phrase to use on items such as items b13 and b15. Lastly, there was the uncertainty that the Malay words used were the right words or the most equivalent words as used by the English version for items b6 and a35 so two options were given for each case. The translator also reported employing guessing as a strategy to overcome these challenges.

Once the initial and the back translations were carried out, the original items along with their Malay and the back translated versions with incomplete forms for some of the items by the second back translator were compiled in table form to be used as a checklist in the expert panel review (see Table 7). The incomplete items provided indications of difficulties in translating the items back into English and were used by the panel to decide whether the items were adequately translated into Malay in the first place. The checklist also included columns for the expert panel to verify whether the translation was acceptable, needed revision, or was not acceptable. The last column in the checklist was made available for the panelists to provide their corrected versions for those items that needed revision or for those that were deemed unacceptable.

Expert panel review. Expert panel reviews to help develop multilingual versions of an instrument have been used in many different ways by different researchers with expert panels of

varying sizes (Daouk-Oyry & McDowal, 2012; Hyrkas, Appelqvist-Schmidlechner, & Oksa, 2003). For the next step in the translation process of this study, an expert panel of five members (see Table 6) along with the researcher convened as a team to evaluate the translation done by the initial translators. Four of the panel members were Malays who were also native speakers of the Malay Language while one was a non-native speaker who was able to speak both English and Malay fluently. They all had a minimum of six years of university teaching at the time of the review.

Each panel member was given the checklist very similar to Table 7 but with the last four columns blank. They were all given a week to complete the checklist individually. They were to examine each item to see if there were discrepancies between the two back translations and the original items. Any discrepancy was an indication that the translated version was not faithfully conveying the intended meaning of the original version.

Following that, all the members met at an appointed time to review the items together in two sessions of an hour each. Since there were 45 items to be examined with a number requiring corrections, fatigue was an issue and the review sessions had to be short to maintain focus and accuracy. During each session, an item by item review was carried out to see if they were acceptable, needed revision, or were unacceptable. For those items that all panel members agreed were acceptable and thus considered to be equivalent to the English version, no further discussion was required except an *A* was placed in the sixth column of Table 7 to show that it was accepted. Items that were acceptable but needed revision were marked with an *R* and the item was discussed in detail to identify the source of the disagreement. Then, the panel worked together to provide an acceptable alternative translation until 5 out of 6 of the panel members

Table 6

Panel Members' Qualifications

Demographics and Qualifications	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Gender	Female	Female	Female	Female	Male
Race	Malay	Malay	Malay	Eurasian	Malay
Native language	Malay	Malay	Malay	Bidayuh	Malay
Bahasa Malaysia Proficiency Level	College Level	College Level	College Level	College Level	College Level
English Language Proficiency Level	Malaysian Certificate of Education (SPM)	Malaysian Certificate of Education (SPM)	College Level	College Level	College Level
Educational level	MA	MA	MA	MA	MA
Field/Major	Malay Language	Linguistics	Psycho-linguistics and Neuro-linguistics	Literary Linguistics	English Literature
Training in psychometrics/ Research Methods	Yes	Yes	Yes	Yes	Yes
Experience in instrument construction/ development	Yes	Yes	Yes	Yes	Yes
Other relevant experience	Qualitative Research	Qualitative Research	Qualitative Research	Qualitative Research	Qualitative Research

agreed to accept it. Then, the new revised version was added in the last column of the checklist.

For any unacceptable items, the panel was asked if they could also provide a better alternative but in one particular case, the researcher decided to retain the initial translation as the panel could not come up with a better translation. The product of the panel review was compiled and shown in Table 7.

As shown in Table 7, when the back-translated versions were compared to the original English version, 16 out of the 45 TPI items were rated by the panel as acceptably equivalent. The other 27 items were deemed acceptable but needed revision and two were rated as unacceptable items. The items that contained minor errors were easily corrected such as b13 which had the word "teaching" in it but was translated as "pembelajaran" meaning "learning" in English instead of the correct form "pengajaran"; the panel agreed to use the latter as it provided a much more faithful interpretation of the item. Items b6 and a35 proved to be quite challenging for the panel members. The expression "virtuoso performer" in b6 was a difficult concept to translate into BM as discovered by the panel. They objected to the Malay expression "pengamal yang luarbiasa" meaning "extraordinary practitioners", which the panel found to be quite different from "virtuoso performers". The phrase "virtuoso performers" is an English usage that has no exact equivalent in BM. Similarly, a35, with the expression "higher ideals" proved to be quite a challenge for the panel to decide. The panel members all agreed that the initial translation of "kesempurnaan yang lebih tinggi" failed to capture the essence of the original meaning of the English version as shown by the two back translations, which appeared as "higher perfection" instead. This is because the word "kesempurnaan" carries the meaning of "perfection" in BM. Even after much deliberation, the panel could not reach a consensus as to alternative statements

to replace the translation of both items. Therefore, items b6 and a35 were classified as unacceptable.

Table 7

Completed Adaptation and Evaluation Checklist

No	Original	Malay Version	Back Trans 1	Back Trans 2	Acceptable, Revise, Unacceptable	Revised Malay version (if needed)
b1	Learning is enhanced by having predetermined objectives.	*Pembelajaran dapat diperkukuh apabila mempunyai objektif yang telah ditentukan terlebih dahulu.*	Learning can be enhanced when objectives have been pre-determined.	Learning can be reinforced when _____.	R	*Pembelajaran diperkukuh apabila mempunyai objektif yang telah ditentukan terlebih dahulu.*
b2	To be an effective teacher, one must be an effective practitioner.	*Untuk menjadi seorang pengajar yang berkesan, seseorang itu mesti juga pengamal yang berkesan.*	An effective teacher must also be an effective practitioner.	To be an effective teacher, one must be an effective practitioner.	A	
b3	Most of all, learning depends on what one already knows.	*Yang paling pentingialah proses pembelajaran bergantung kepada asas pengetahuan sedia ada pada seseorang.*	Most importantly, the learning process must depend on the fundamental knowledge that a person has.	The most important is that learning process depends on _____.	A	
b4	It is important that I acknowledge learners' emotional reactions.	*Penting untuk saya mengambil kira reaksi emosi pelajar.*	It is important for me to take into con-sideration the emotional reaction of students.	It's important for me to consider students' emotional reactions.	A	

Table 7 (Continued)

b5	My teaching focuses on societal change, not the individual learner.	*Pengajaran saya berfokus kepada perubahan masyarakat, tidak pada pelajar tertentu*	My teaching is focused on changes in society, not on specific students.	My teaching focuses on the changes in society/ social change, not on a particular student.	R	*Pengajaran saya berfokus kepada perubahan masyarakat, tidak padase seorang pelajar .*
b6	Teachers should be virtuoso performers of their subject matter.	*Para pengajar sepatutnya menjadi pengamal yang luarbiasa terhadap subjek yang diajar*	Educators must be exceptional practitioners in the subject taught.	Instructors/Teachers should be an extra-ordinary practitioner of the subject taught .	U	*Para pengajar sepatutnya menjadi pengamal yang luarbiasa terhadap subjek yang diajar.**
b7	The best learning comes from working alongside good practitioners.	*Pembelajaran yang berkesan wujud daripada kerjasama dengan pengamal-pengamal yang baik.*	Effective learning takes place with the cooperation of good practices.	Effective learning comes from a partnership with the best practitioners.	R	*Pembelajaran terbaik wujud daripada kerjasama dengan pengamal-pengamal yang baik.*
b8	Teaching should focus on developing qualitative changes in thinking.	*Pengajaran harus memberi fokus kepada matlamat untuk membawa perubahan yang jelas dalam cara berfikir.*	Teaching has to focus on bringing clear changes in the ways of thinking.	Teaching should/must focus on the aim/goal to bring a clear change in the way of thinking.	R	*Pengajaran harus berfokus kepada membina perubahan kualitatif dalam pemikiran.*
b9	In my teaching, building self-confidence in learners is a priority.	*Dalam pengajaran saya, membina keyakinan diri dalam diri pelajar menjadi keutamaan.*	In my teaching, building self-confidence in students is a priority.	In my teaching, fostering self-confidence in students becomes/is a priority.	A	
b10	Individual learning without social change is not enough.	*Pembelajaran individu tanpa perubahan sosial adalah tidak memadai.*	Individual learning without social change is insufficient.	Individual learning without social change is not sufficient.	A	

Table 7 (Continued)

b11	Effective teachers must first be experts in their own subject areas.	*Pengajar yang berkesan mesti terlebih dahulu pakar dalam bidangnya.*	An effective teacher must first be an expert in his or her field.	An effective teacher must be an expert in his/her field.	A	
b12	Knowledge and its application cannot be separated.	*Ilmu pengetahuan dan aplikasinya tidak dapat dipisahkan.*	Knowledge and its applications are inseparable.	Knowledge and its application cannot be separated.	A	
b13	Teaching should build upon what people already know.	*Proses pembelajaran seharusnya bersandarkan pengetahuan sedia ada seseorang.*	The learning process should be based on a person's inherent knowledge.	Learning process should be based on the existing knowledge of an individual/an individual's existing knowledge.	R	*Proses pengajaran seharusnya berasaskan pengetahuan sedia ada seseorang.*
b14	In learning, people's effort should be rewarded as much as achievement.	*Dalam pembelajaran, usaha individu perlu diberi ganjaran setimpal dengan pencapaian-nya*	In learning, an individual's effort should be rewarded based on his or her achievement.	In learning, individual efforts need/must be awarded/given a reward (that is) com-men surate with his/her achievement.	A	
b15	For me, teaching is a moral act as much as an intellectual activity.	*Pada saya, mengajar ialah satu tindakan moral seperti aktiviti intelektual.*	To me, teaching is a moral act, similar to an intellectual activity.	To me, teaching is a moral action (just like/for instance) an intellectual activity.	A	*Pada saya, mengajar ialah satu tindakan moral yang juga aktiviti intelektual.*
i16	My intent is to prepare for examinations	*Hasrat saya adalah untuk menyediakan individu menduduki peperiksaan*	My aim is to prepare individuals to sit for examinations.	My desire/intention is to prepare an individual to sit for an exam.	R	*Hasrat saya adalah untuk mempersiap-kan individu untuk peperiksaan.*

Table 7 (Continued)

i17	My intent is to demonstrate how to perform or work in real situations.	*Hasrat saya adalah untuk menunjuk ajar cara melakukan sesuatu atau bekerja dalam situasi sebenar.*	My aim is to show how things are done or work in real situations.	My desire/intention is to show (through teaching) how to do something or to work in real situations.	A	
i18	My intent is to help people develop more complex ways of reasoning.	*Hasrat saya ialah untuk membantu individu mengembang-kan penaakulan yang lebih kompleks.*	My aim it to help individuals to develop complex understanding.	My desire/intention is to help individuals develop more complex reasoning.	R	*Hasrat saya adalah untuk membantu individu mengembang-kan penaakulan yang lebih kompleks.*
i19	My intent is to build people's self-confidence and self-esteem as learners.	*Hasrat saya adalah untuk membina keyakinan dan harga diri individu sebagai pelajar.*	My aim is to develop self-confidence and self-esteem in students.	My desire/intention is build confidence and self-esteem in individuals as learners.	A	
i20	My intent is to challenge people to seriously reconsider their values.	*Hasrat saya adalah Untuk mencabar individu menilai semula prinsip diri secara serius.*	My aim is to help individuals to seriously assess their principles.	My desire/intention is to challenge the individual to reevaluate (seriously) his/her self principles.	R	*Hasrat saya adalah untuk mencabar individu mempertimbang kan semula nilai diri secara serius.*
i21	I expect people to master a lot of information related to the subject.	*Saya berharap individu dapat menguasai banyak maklumat yang berkaitan dengan subjek yang diajar.*	I hope individuals will acquire a lot of information related to the subject taught.	I hope that individuals can master a lot of information related to the subject taught.	R	*Saya meng-kehendaki individu untuk menguasai banyak maklumat berkaitan subjek.*

Table 7 (Continued)

i22	I expect people to know how to apply the subject matter in real settings.	*Saya berharap individu berupaya mengaplikasi bahan pengajaran dalam situasi sebenar.*	I hope individuals will be able to apply what has been taught in real situations.	I hope that individuals are able to (apply/use) teaching materials in real situations.	R	*Saya meng-kehendaki individu untuk mengetahui cara mengaplikasi kandungan pelajaran dalam situasi sebenar.*
i23	I expect people to develop new ways of reasoning about the subject matter.	*Saya berharap individu dapat membangun-kan kaedah baru bagi menimbang-kan hal yang berkaitan dengan bahan pengajaran.*	I hope individuals will be able to develop new methods to assess issues related to the teaching materials.	I hope that individuals can develop new ways to consider matters related to teaching materials.	R	*Saya meng-kehendaki individu untuk membangun-kan kaedah baru dalam memper-timbangkan hal-hal berkaitan kandungan pelajaran.*
i24	I expect people to enhance their self-esteem through my teaching.	*Saya berharap individu dapat meningkatkan harga diri mereka melalui pengajaran saya.*	I hope individuals will grow in self-esteem through my teaching.	I hope that individuals can improve their self-esteem through my teaching.	A	
i25	I expect people to be committed to changing our society.	*Saya berharap individu komited untuk melakukan perubahan kepada masyarakat.*	I hope individuals will be committed to bringing about change in society.	I hope that individuals are committed to bring (about) change in the society.	R	*Saya meng-kehendaki individu untuk komited melakukan perubahan kepada masyarakat.*

Table 7 (Continued)

i26	I want people to score well on examinations as a result of my teaching.	*Saya mahu individu memperoleh keputusan yang baik dalam peperiksaan hasil daripada pengajaran saya.*	I would like individuals to obtain good grades in their examinations as a result of my teaching.	I want the individuals to obtain/get good results in the exam as a result of my teaching.	R	*Saya mahu individu memperoleh keputusan cemerlang dalam peperiksaan hasil daripada pengajaran saya.*
i27	I want people to understand the realities of working in the real world.	*Saya mahu individu memahami realiti bekerja dalam dunia yang sebenarnya.*	I would like individuals to understand the realities of working in the real world.	-	R	*Saya mahu individu memahami realiti bekerja dalam dunia sebenar.*
i28	I want people to see how complex and inter-related things really are.	*Saya mahu individu melihat betapa kompleks dan saling kebergantung-an sesuatu perkara itu sebenarnya.*	I would like individuals to see how complex and inter-connected things really are.	I want the individual to see how complex and inter-dependent _____.	R	*Saya mahu individu melihat betapa kompleks dan saling bergantungnya sesuatu perkara itu.*
i29	I want to provide a balance between caring and challenging as I teach.	*Saya mahu menyediakan keseimbangan antara mengambil berat dan mencabar kemampuan pelajar semasa saya mengajar.*	I would like to strike a balance between caring for and challenging my students when I teach.	I want to provide a good balance between caring and challenging students' ability when I teach.	A	
i30	I want to make apparent what people take for granted about society.	*Saya mahu mendedahkan sikap sambil lewa individu terhadap masyarakat.*	I would like to expose the laidback attitude of individuals towards society.	I want to expose _____ attitude of the individual to/toward (the) society.	R	*Saya mahu mendedahkan perihal masyarakat yang diambil mudah oleh individu.*

Table 7 (Continued)

a31	I cover the required content accurately and in the allotted time.	*Saya melengkapkan kandungan kursus dengan tepat dan dalam masa yang diperuntukkan.*	I complete the course content accurately within the specified time.	I want to complete the course content (correctly) and within the time allotted/ given/given time.	R	*Saya menyelesaikan keperluan kandungan kursus dengan tepat dan dalam masa yang diperuntukkan.*
a32	I link the subject matter with real settings of practice or application.	*Saya menghubung-kaitkan bahan pengajaran secara praktis dengan dunia sebenarnya atau aplikasinya.*	I relate the teaching materials in a practical way to its real world applications.	I relate teaching materials (practically) with (its) real world or (its) application.	R	*Saya meng-hubungkaitkan kandungan pelajaran secara praktis dengan dunia sebenar atau aplikasi.*
a33	I ask a lot of questions while teaching.	*Saya bertanya banyak soalan semasa mengajar.*	I ask a lot of questions while teaching.	I ask a lot (of questions) while teaching.	A	
a34	I find something to compliment in everyone's work or contribution.	*Saya menemui sesuatu untuk diberi pujian dalam setiap sumbangan seseorang.*	I look for something praise-worthy in every individual contribution.	I always find something to praise _____ .	R	*Saya mencari sesuatu untuk dipuji dalam setiap sumbangan seseorang.*
a35	I use the subject matter as a way to teach about higher ideals.	*Saya menggunakan bahan pengajaran sebagaicara untuk mengajar mencapai kesempurnaan yang lebih tinggi.*	I use teaching materials as a tool to achieve greater perfection.	I use teaching materials as a way to teach how to achieve a "higher perfection".	U	*Saya menggunakanb ahan pengajaran sebagai cara untuk mengajar mencapai kesempurnaan yang lebih tinggi**
a36	My teaching is governed by the course objectives.	*Pengajaran saya dikawalselia oleh objektif kursus.*	My teaching is directed by the course objectives.	My lessons are governed by the course objectives.	R	*Pengajaran saya berpandukan objektif kursus.*

Table 7 (Continued)

a37	I model the skills and methods of good practice.	*Saya mencontohi kemahiran dan kaedah pengajaran yang berkesan.*	I imitate effective teaching skills and methods.	I follow/model effective teaching skills and methods.	R	*Saya mencontohi kemahiran dan kaedah pengajaran yang baik*
a38	I challenge familiar ways of understanding the subject matter.	*Saya mencabar pendekatan pengajaran yang lazim untuk memahami sesuatu bahan pengajaran.*	I challenge the common teaching approaches to understand teaching materials.	I challenge the conven-tional teaching approach to understand a particular teaching material.	R	*Saya mencabar kaedah-kaedah lazim yang digunakan untuk memahami kandungan pelajaran*
a39	I encourage expressions of feeling and emotion.	*Saya menggalakkan ekspresi perasaan dan emosi.*	I encourage the expression of feelings and emotions.	I encourage expressions of (feelings?) and emotions.	A	
a40	I emphasize values more than knowledge in my teaching.	*Saya lebih memberi tumpuan kepada nilai-nilai murni dalam pengajaran saya berbanding dengan ilmu pengetahuan.*	I focus more on the moral values rather than knowledge in my teaching.	I focus more in moral values in my lesson/ teaching compared to knowledge.	R	*Saya lebih memberi penekanan kepada nilai-nilai murni dalam pengajaran saya berbanding ilmu pengetahuan.*
a41	I make it very clear to people what they are to learn	*Saya menerangkan dengan jelas kepada individu tentang sesuatu perkara yang akan mereka pelajari.*	I explain clearly to individuals things that they are learning.	I explain clearly to individuals on what they will learn.	R	*Saya menerangkan dengan jelas kepada individu tentang perkara yang akan mereka pelajari.*

Table 7 (Continued)

a42	I see to it that novices learn from more experienced people.	*Saya memastikan individu yang kurang ber-pengalaman mempelajari daripada mereka yang lebih ber-pengalaman.*	I ensure that less experienced individuals learn from those who are more experienced.	I ensure/make sure that less experienced students learn from more experienced students/ ones.	R	*Saya memastikan individu yang kurang ber-pengalaman belajar daripada mereka yang lebih ber-pengalaman.*
a43	I encourage people to challenge each others' thinking.	*Saya menggalakkan individu mencabar pemikiran masing-masing.*	I encourage individuals to challenge their own thinking.	I encourage individuals to challenge their own thinking.	R	*Saya menggalakkan individu mencabar pemikiran antara satu sama lain*
a44	I share my own feelings and expect my learners to do the same.	*Saya berkongsi perasaan saya dan mengharapkan pelajar saya juga berbuat demikian.*	I share my feelings and hope my students do the same.	I share my feelings and I expect/hope students (will) do the same/ likewise.	R	*Saya berkongsi perasaan saya dan meng-kehendaki pelajar saya juga berbuat demikian.*
a45	I link instructional goals to necessary changes in society	*Saya menghubung-kaitkan matlamat pengajaran dengan perubahan yang diperlukan dalam masyarakat.*	I relate the aims of my teaching to the changes needed in the society.	I relate the learning objectives with the changes needed in (the) society.	A	

As a recourse, the researcher made the decision to retain the translation of both items as proposed by the initial translators and test their equivalence psychometrically.

Pilot testing. Participants for the test-retest study were 25 lecturers from three Malaysian public universities. Initially, all lecturers from the 20 Malaysian public universities were contacted via email with a letter of consent indicating that their participation in this phase of the

study was designed to examine the test-retest reliability of the TPI. A link to access the adapted TPI was also given for them to complete the survey online. However, due to time constraints, the pretest was closed after the first 25 participants successfully completed it and the description of their profiles is shown in Table 8.

In order for the test-retest to be carried out, all the participants were asked to provide their own identification code on their electronic survey so that responses for the first and the second administrations could be linked. Emails were again sent out to all respondents after two weeks to complete the retest and the whole test-retest period took about four months to complete because many respondents were slow to complete the process. Furthermore, after more than four weeks had lapsed, only 16 respondents who completed the first administration came back to do the retest. Another group of respondents was contacted three months after the first email was sent out to make up for the missing nine respondents. Two weeks after the new group of respondents took the pilot survey, they were contacted to take the retest. This time, all nine of them completed the test-retest on time. After that, the means of all the items were calculated and compared to see if they were significantly different. This was to determine if the items were functioning similarly across the two administrations. Paired samples *t*-tests were carried using SPSS version 21 and correlations between the items across the two time points were calculated.

Overall, four items b6, i17, i19 and i29, as shown in Appendix L, were identified as showing significant difference in their means with 2-tailed *p* values of .032, .029, .050 and .038, respectively. These items were scrutinized in the cognitive interviews. The remaining items showed no significant difference between their means after two administrations. In addition to the items identified in the paired samples test-retest, items b1, b11, i26, and a42, which showed

low and non-significant correlation values ranging from .12 to .37, were also investigated in the cognitive interviews. Out of the 45 items in the TPI questionnaire, eight of the items were deemed to have shown irregular performance across the two administrations.

Table 8

Demographic Characteristics of the Pilot Study Group (*n* = 25)

Characteristic	Frequency	%
Gender		
Male	19	76.0
Female	5	20.0
Missing	1	4.0
Age		
51- 60	2	8.0
41 - 50	4	16.0
31 - 40	15	60.0
25 - 30	3	12.0
Missing	1	4.0
Years Teaching		
26-50	2	8.0
16-25	5	20.0
6-15	11	44.0
1-5	6	24.0
Missing	1	4.0
Academic Level		
Masters	17	68.0
Ph.D.	7	28.0
Missing	1	4.0
Academic Status		
Lecturer	18	72.0
Senior lecturer	5	20.0
Professor	1	4.0
Missing	1	4.0

A subscale by subscale analysis was also carried out to determine test-retest correlations. Table 9 shows the Cronbach's α for each subscale as well as the descriptive statistics, and the test-retest correlations for the subscales and their items. The Cronbach alphas were generally low in both the first and the second administrations. As shown in Table 9, in the first administration, the alpha values for the Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform scales were .37, .50, .67, .54, and .62, respectively. As for the second administration, the alphas were slightly lower for most of the subscales except for Apprenticeship; the alpha values for the Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform scales were .20, .50, .46, .24, and .28, respectively. Test-retest reliabilities were .68 for Transmission, .48 for Apprenticeship, .77 for Developmental, .81 for Nurturing, and .57 for Social Reform. Collins and Pratt (2011) reported test-retest reliability between the first and the second administrations of the individual scale scores that ranged from .62 (Developmental) to .71 (Social Reform). Collins and Pratt went on to report the test-retest reliabilities between the second and third administrations with a sample of 63 people showing an average correlation of .73 with individual scale correlations between .65 (Nurturing) and .87 (Social Reform).

Table 9

Descriptive Statistics and Results ofTest-Retest of the TPI Subscales and their Items (n = 25)

Subscale	Internal Consistency		Descriptive Statistics		Test-Retest Correlations	
	Cronbach's α		Mean (*SD*)		*r*	Sig.
	Test	Retest	Test	Retest		
Transmission	.37	.20	3.84 (0.34)	3.83 (0.33)	.68	.00
b1			4.56 (0.51)	4.32 (0.80)	.26	.21
b6			3.44 (0.96)	3.72 (1.02)	.81	.00
b11			4.32 (0.56	4.40 (0.58)	.10	.62
i16			3.16 (0.90)	3.24 (0.93)	.75	.00
i21			4.12 (0.88)	3.92 (1.04)	.56	.00
i26			4.20 (1.04)	4.12 (1.01)	.17	.41
a31			3.72 (0.94)	3.68 (0.95)	.98	.00
a36			4.00 (0.87)	3.92 (0.86)	.84	.00
a41			4.08 (0.64)	4.08 (0.70)	.91	.00
Apprenticeship	.50	.50	4.26 (0.36)	4.13 (0.38)	.48	.01
b2			4.32 (0.69)	4.32 (0.75)	.60	.00
b7			4.04 (0.61)	4.00 (0.71)	.77	.00
b12			4.24 (0.88)	4.08 (1.00)	.64	.00
i17*			4.68 (0.48)	4.36 (0.70)	.36	.08
i22			4.28 (0.89)	4.16 (0.94)	.79	.00
i27			3.92 (1.08)	3.88 (1.09)	.70	.00
a32			4.16 (0.75)	4.20 (0.71)	.96	.00
a37			4.04 (0.79)	4.00 (0.76)	.97	.00
a42			3.92 (0.91)	3.72 (1.06)	.32	.11
Developmental	.67	.46	4.04 (0.47)	3.85 (0.46)	.77	.00
b3			4.20 (1.08)	4.04 (0.94)	.65	.00
b8			4.16 (0.99)	4.04 (0.98)	.86	.00
b13			3.80 (0.87)	3.56 (1.12)	.51	.01
i18			4.48 (0.77)	4.20 (0.96)	.66	.00
i23			3.92 (0.86)	3.88 (0.93)	.77	.00
i28			3.92 (1.08)	3.80 (1.12)	.85	.00
a33			4.28 (0.68)	4.20 (0.91)	.92	.00
a38			3.64 (1.04)	3.48 (1.12)	.73	.00
a43			3.44 (1.16)	3.04 (1.24)	.57	.00
Nurturing	.54	.24	4.08	3.92	.81	.00
b4			4.40 (0.58)	4.36 (0.76)	.52	.01
b9			4.32 (0.69)	4.00 (0.96)	.57	.00
b14			4.28 (0.61)	4.28 (0.61)	.89	.00
i19			4.56 (0.58)	4.28 (0.89)	.65	.00
i24			4.24 (0.78)	4.08 (0.81)	.82	.00

Table 9 (Continued)

	i29		4.20 (0.71)	3.80 (1.00)	.47	.02
	a34		3.16 (1.14)	2.96 (1.21)	.88	.00
	a39		3.44 (1.04)	3.48 (1.01)	.98	.00
	a44		3.80 (0.96)	3.60 (1.12)	.86	.00
Social Reform	.62	.28	4.00 (0.47)	3.86 (.37)	.57	.00
	b5		3.60 (1.08)	3.44 (1.23)	.55	.01
	b10		4.36 (0.57)	4.28 (0.61)	.65	.00
	b15		4.56 (0.65)	4.52 (0.71)	.69	.00
	i20		4.20 (0.76)	3.96 (0.99)	.57	.00
	i25		4.00 (1.16)	3.84 (1.18)	.86	.00
	i30		3.76 (1.05)	3.60 (1.08)	.83	.00
	a35		3.60 (0.76)	3.56 (0.92)	.81	.00
	a40		3.80 (0.87)	3.72 (0.94)	.55	.01
	a45		3.64 (1.11)	3.48 (1.05)	.62	.00

Note. For the paired *t*-tests *df*=24

In addition to the items identified by the analyses of the responses by the 25 pilot study respondents, inconsistent responses between the initial test and the retest by each of the cognitive interview participants were also short-listed and scrutinized in the interview. This information was obtained from their responses after they were contacted to get permission to be interviewed individually. It was discovered that the participants for the cognitive interviews did not show any inconsistencies for some of the items that were identified from the pilot study participants. For example, cognitive interviewee 1 endorsed "Disagree" on both administrations for item b6 which was identified as one of the inconsistent items. However, there were instances where the interviewees showed inconsistencies in their responses to which the other respondents in the pilot study showed consistencies. For instance, cognitive interviewee 1 chose "Always' for item i26 in the first administration but selected "Rarely" in the second one. It was decided that a more constructive approach was to also focus on those items that showed conflicting responses by the interviewees and seek clarifications as to the source of the discrepancies even though in the overall analysis, the items did not show any inconsistencies. All in all, nine items were identified

(b1, b6, b7, i17, b11, i19, i26, i29, and a42) for all the interviewees to explore and discuss with the interviewer. As for the individual participant's inconsistent responses, the number of additional items discussed varied among the five interviewees ranging from an additional 2 items to 12 items.

Cognitive interviews. The next step to ensure quality in the translation process was for the researcher to conduct cognitive interviews with a combination of think-aloud and verbal probing techniques (Willis, 1999, 2005). The overall goal of cognitive interviewing was to uncover the thought processes that respondents employed to answer the TPI items. These thought processes are normally hidden and not revealed to outside observers. By exploring these thought processes through cognitive interviewing, information regarding the accuracy and appropriateness of the responses was ascertained. Identification of confusion and misinterpretation of certain items helped in making them more congruent to the requirements of the original items. Five participants from the pilot study group were selected by the researcher as interviewees in a series of cognitive interviews. A purposeful sampling method was used to choose the participants from the pilot study pool by identifying three female faculty members and two male faculty members from Universiti Malaysia Sarawak based on the demographic data gathered from the pilot study. They were contacted via email by the investigator to obtain their permission before making an appointment to meet face to face. Since the data for the test-retest were confidential, all identification codes of the other participants were deleted. For this study, the researcher as the interviewer conducted each cognitive interview by starting each meeting with a short training session to demonstrate the steps in the cognitive interviewing process. The specific techniques included the use of both a think-aloud process and verbal

probing. Respondents were shown a particular item from the questionnaire and instructed to think out loud as they answered the question. This was followed by the interviewer asking a series of spontaneous questions to probe for further information about why the respondent answered the question the way he or she did. After the training session, the researcher explained the rationale and the purpose of the think-aloud and the probing questions so as to alleviate any feelings of suspicion or anxiety on the part of the interviewees. Some of the questions asked were "What do you think the question is asking for?", "What do the words or phrases in the items mean to you?", What types of information do you need to recall in order to answer the question?", and "How did you arrive at that answer?"

Outcome of the cognitive interviews. Table 10 shows the responses of the five interviewees, three females and two males, gathered during the test-retest phase of the pilot study. Even though some items above were endorsed consistently by some of the interviewees, all of them were asked to think aloud how they responded to all these items followed by a series of probing questions to gain further information about the way they answered these items in light of the findings in Table 9. The responses to items that are in bold indicated inconsistencies of responses by individual participants which are similar to the rest of the participants of the pilot study. Items i17, i19 and i29 had 3 out of the 5 participants providing inconsistent responses, which indicates that these are items were challenging for the five participants. All of these items were from the Intention domain as opposed to the other domains of Beliefs or Actions of the TPI. Overall, the language in terms of content and reading level of the instrument was manageable according to all the respondents. In other words, the sentences could be understood well enough to make a decision about the items. However, looking at the responses in the two

administrations, some inconsistencies were also found in other items besides those identified in the test-retest. The reasons given were very similar to those discovered in Table 9. Some items needed to be clarified in terms of whether they addressed issues at a general level or specifically towards a class that the faculty member was currently teaching. There was tension between achieving the ideal goal versus classroom reality. This can be seen in items like i16 and i26 which focus on teaching to pass examinations. This is a common struggle in the Malaysian education scenario where teachers are torn between teaching students for learning or just for passing exams. Even though most faculty members believe in helping students learn, expediency often forced them to do otherwise. This is especially pronounced in item i29 where the sentence seems to contain two parts, caring as opposed to challenging students, which the participants had to consider. The word 'challenge" itself when translated into BM "mencabar" can portray a very aggressive posture, which may induce differing responses among participants. As shown in Table 10, three out of the five interviewees had inconsistent responses to it.

There were items like i17 where some interviewees were not clear whether the item was asking about specific classroom situations or asking about educational goals in general. Even though they responded to this item more consistently than the rest of the pilot study respondents, they mentioned that item i17 was dependent on the type of subject or students that they teach. One interviewee gave item b12 as an example of her Mathematics class which does not really require her students to see a demonstration of how Mathematics is applied in real work situations as compared to her Statistics class which has a more practical application.

Table 10

Interviewees' Responses to Items Identified During the Test-Retest Pilot Study

No	Item	Interviewee 1 Female	2 Female	3 Female	4 Male	5 Male	Reasons
b1	Learning is enhanced by having predetermined objectives.	Ag, Sag	Ag, SAg	SAg, SAg	SAg, SAg	SAg, SAg	Consistent response for all
b6	Teachers should be virtuoso performers of their subject matter.	**D, Ag**	D, D	Ag, SAg	Ag, SAg	Ag, SAg	Confusing words
b11	Effective teachers must first be experts in their own subject areas.	Ag, Ag	**Ag, Nu**	Ag, SAg	Ag, SAg	Ag, SAg	Similar to b6 What is expert?
i17	My intent is to demonstrate how to perform or work in real situations.	U, U	A, A	**A, U**	**U, A**	**U, A**	Depending upon type of students or subject taught
i19	My intent is to build people's self-confidence and self-esteem as learners.	A, A	A,A	**A, S**	**S, A**	**S, A**	Same as i17
i26	I want people to score well on examinations as a result of my teaching.	U, U	**R, A**	U, U	A, A	A, A	"People" in general, yes but 'students' no.
i29	I want to provide a balance between caring and challenging as I teach.	**R, U**	**R, A**	**A, R**	U, U	U, U	In general, yes, but for a specific group need to be more caring
a42	I see to it that novices learn from more experienced people.	S, S	A,A	**U, N**	U, U	U, U	From teachers, yes but from fellow students, no.

Belief items: SD=Strongly Disagree, D= Disagree, Nu=Neutral, Ag= Agree, SAg=Strongly Agree
Intention and Action items: N=Never, R=Rarely, S=Sometimes, U=Usually, A=Always

Similarly, there were also items like i16, i23, and a43 that had the word, "people" that was translated as "individu" (individuals), which was often misinterpreted as people in general

by respondents and not about their students. Based on the clarification by Pratt and his colleague, Collins (personal communication, August 10, 2013), the context of all the items like i16, i17, i23, and a43 was specifically aimed at the most recent class that the faculty was teaching and not to other forms of social interactions in general. As a result, additional reminders, as advised by the original developers, were added to the instructions for every subsection in the survey to ensure participants responded to items by reflecting on their latest class that they had taught rather than the general context of teaching.

Even though there were inconsistencies in the responses to different items by the five interviewees, it was deemed not serious enough to merit further changes. For some items, respondents were surprised that their responses changed and admitted that it was a mistake on their part. This means that the items were actually functioning properly. For item b6, there was no exact equivalent of the phrase "virtuoso performer" in Malay and thus, the translation of the item was not changed. However, the word "novis" in brackets, which is a borrowed word from English "novice", was added in brackets to item a42 to add more clarity to the item. With that, the TPI Malay version (see Appendix M) was deemed ready for the next stage of the study.

Stage II: Measurement Invariance Testing

Demographics. As shown in Table 11, the samples used in the psychometric analyses of the two versions of the TPI from the U.S. and Malaysia were quite closely matched according to demographics such as gender, percentage of time spent teaching, primary role, and the types of learners commonly taught. However, experience in years as teachers as well as practitioners in their profession differed slightly.

Table 11

Characteristics of the Faculties in the U. S. and Malaysian Samples

Characteristic	U.S. (n = 605) %	Malaysian (n = 561) %
Gender		
Male	39.3	36.0
Female	60.7	59.7
Missing	0	4.3
Percent Teaching		
90-100	16.0	9.6
60-80	40.0	53.5
30-50	43.9	31.7
10-20	0	5.1
Missing	0	0
Primary Role		
Teacher	66.3	60.7
Practitioner	7.8	5.2
Manager	4.6	1.6
Administrator	6.3	10.9
Researcher	6.0	18.4
Others	9.1	3.2
Missing	0	2
Usual Learners		
Undergrad	73.9	80.0
Post grad	26.1	18.0
Professional	0	1.8
Others	0	.2
Missing	0	0
Years Teaching		
26-50	8.4	6.2
16-25	15.7	11.6
6-15	33.0	31.5
1-5	36.3	18.0
Less than 1	5.3	31.4
Missing	1.2	0

Table 11 (Continued)

Years Practicing (e.g., A practicing lawyer besides teaching law)		
26-50	11.3	3.9
16-25	21.5	5.5
6-15	32.8	17.1
1-5	21.1	18.6
Less than 1	9.1	7.1
Missing	4.5	47.8

Note. Categories for the variable Percent Teaching matched those categories used by Pratt.

Descriptive statistics of the TPI. Descriptive statistics for each subscale of the TPI for the two countries are shown in Table 12. Each subscale has nine items. Many of the items' score distributions were negatively skewed for both countries but the Malaysian sample revealed more skewness than those of the U.S. sample in terms of number and size. There were nine items that showed large effect sizes (Cohen's *d*) to indicate that the means of each group were different from each other. The greatest differences were for items b3 and b5. The rest of the 36 TPI items appeared to be similar across the two groups especially item a39.

Table 12

Descriptive Statistics of the TPI for the U. S. (*n* = 605) *and the Malaysian* (*n* = 561) *Samples*

Items	U.S.			Malaysia			
	Mean (*SD*)	Skewness	Kurtosis	Mean (*SD*)	Skewness	Kurtosis	Effect Size
Transmission							
b1. Learning is enhanced by having predetermined objectives.	4.29 (0.76)	-1.16	1.81	4.55 (0.82)	-2.73	8.81	-0.33
b6. Teachers should be virtuoso performers of their subject matter.	3.27 (1.02)	-0.35	-0.63	3.52 (1.05)	-0.22	-1.06	-0.24
b11. Effective teachers must first be experts in their own subject.	3.75 (1.00)	-0.58	-0.45	3.70 (1.06)	-0.64	-0.32	0.05

Table 12 (Continued)

i16. My goal is to prepare people for content-related examinations.	2.80 (0.97)	0.14	-0.20	2.99 (1.06)	0.15	-0.74	-0.19
i21. I expect people will master a lot of information related to the subject.	3.50 (0.87)	0.05	-0.30	4.12 (0.73)	-0.75	0.84	-0.77
i26. I want people to score well on examinations as a result of my teaching.	3.77 (1.02)	-0.61	-0.17	4.02 (.97)	-0.83	-0.20	-0.25
a31. I cover the required content accurately and in the allotted time.	3.97 (.68)	-0.55	0.98	4.27 (0.80)	-1.29	2.51	-0.41
a36. My teaching is governed by the course objectives.	3.94 (0.79)	-0.52	0.47	4.31 (0.79)	-1.65	4.33	-0.47
a41. I make it very clear to people what they are to learn.	4.13 (0.80)	-0.69	0.21	4.28 (0.68)	-1.06	2.85	-0.20
Apprenticeship							
b2. To be a good teacher, one must be a good practitioner.	3.98 (0.91)	-0.93	0.63	4.20 (0.82)	-0.86	0.30	-0.25
b7. The best learning comes from working alongside good practitioners.	3.91 (0.77)	-0.70	0.76	4.00 (0.85)	-1.50	3.18	-0.11
b12. Knowledge and its application cannot be separated.	3.53 (1.09)	-0.33	-0.96	4.28 (0.80)	-1.85	5.35	-0.78
i17. My goal is to demonstrate how to perform or work in real situations.	4.23 (0.83)	-0.83	0.05	4.34 (0.68)	-0.85	0.75	-0.14
i22. I expect people to know how to apply the subject matter in real settings.	4.28 (0.73)	-0.83	0.63	4.30 (0.69)	-1.29	3.89	-0.03
i27. I want people to understand the realities of working in the real world.	4.37 (0.79)	-1.15	0.85	4.37 (0.78)	-1.03	0.37	0.00
a32. I link the subject matter with real settings of practice or application.	4.41 (0.67)	-0.94	0.91	4.33 (0.74)	-1.19	1.85	0.11

Table 12 (Continued)

a37. I model the skills and methods of good practice.	4.22 (0.68)	-0.66	0.99	4.10 (0.79)	-0.93	1.65	0.16
a42. I see to it that novices learn from more experienced people.	3.64 (0.89)	-0.31	-0.07	3.85 (0.85)	-0.57	0.21	-0.42
Developmental							
b3. Most of all, learning depends on what one already knows.	2.79 (1.04)	0.25	-0.77	3.94 (0.99)	-0.99	0.28	-1.13
b8. Teaching should focus on developing qualitative changes in thinking.	3.96 (0.81)	-0.78	0.88	4.01 (0.76)	-0.69	0.86	-0.06
b13. Teaching should build upon what people already know.	3.81 (0.95)	-0.65	-0.26	3.71 (0.96)	-1.03	0.80	0.11
i18. My goal is to help people develop more complex ways of reasoning.	4.34 (0.76)	-1.12	1.36	4.33 (0.75)	-0.81	-0.07	0.01
i23. I expect people to develop new ways of reasoning about the subject.	4.09 (0.77)	-0.51	0.01	3.22 (1.33)	-0.45	-0.96	0.81
i28. I want people to see how complex and inter-related things really are.	4.39 (0.74)	-1.21	1.57	4.17 (0.79)	-0.61	-0.17	0.29
a 33. I ask a lot of questions while teaching.	4.21 (0.78)	-0.61	-0.48	4.14 (0.79)	-0.36	-1.02	0.09
a38. I challenge familiar ways of understanding the subject matter.	3.85 (0.86)	-0.50	0.10	3.50 (1.15)	-0.69	-0.23	0.35
a43. I encourage people to challenge each other's thinking.	3.92 (0.93)	-0.60	-0.11	3.81 (1.01)	-0.64	-0.36	0.11
Nurturing							
b4. It's important that I acknowledge learners' emotional reactions.	4.01 (0.76)	-0.90	1.59	4.06 (0.70)	-0.81	1.78	-0.07
b9. In my teaching, building self-confidence in learners is a priority.	4.20 (0.78)	-0.95	1.08	3.99 (0.89)	-0.88	0.30	0.25

Table 12 (Continued)

b14. People's effort should be rewarded as much as achievement.	3.37 (1.03)	-0.44	-0.51	3.98 (0.99)	-1.25	1.22	-0.60
i19. My goal is to build people's self-confidence and self-esteem as learners.	4.18 (0.92)	-0.92	0.20	4.40 (0.74)	-0.84	-0.51	-0.26
i24. I expect that people will enhance their self-esteem through my teaching.	3.62 (1.04)	-0.35	-0.54	3.92 (1.00)	-0.59	-0.71	-0.29
i29. I want to provide a balance between caring and challenging as I teach.	4.41 (0.79)	-1.37	1.80	4.18 (0.85)	-0.94	0.89	0.28
a34. I find something to compliment in everyone's work or contribution.	3.93 (0.89)	-0.65	0.05	3.39 (1.27)	-0.67	-0.58	0.49
a39. I encourage expressions of feeling and emotion.	3.53 (1.09)	-0.27	-0.70	3.53 (1.23)	-0.73	-0.36	0.00
a44. I share my own feelings and expect my learners to do the same.	3.55 (1.05)	-0.39	-0.36	3.35 (1.14)	-0.10	-0.87	0.18
Social Reform							
b5. My teaching focuses on societal change, not the individual learner.	2.23 (0.88)	0.07	0.56	3.39 (0.96)	-0.50	-0.22	-1.26
b10. Individual learning without social change is not enough	3.09 (1.02)	-0.10	-0.63	3.61 (1.01)	-0.52	-0.32	-0.51
b15. For me, teaching is a moral act as much as an intellectual activity.	3.86 (0.96)	-0.65	-0.15	4.36 (0.81)	-2.18	6.88	-0.56
i20. My goal is to challenge people to seriously reconsider their values.	3.04 (1.09)	0.13	-0.49	3.94 (0.91)	-0.41	-0.33	-0.89
i25. I expect people to be committed to changing our society.	3.07 (1.08)	0.05	-0.49	3.75 (1.07)	-0.51	-0.61	-0.63

Table 12 (Continued)

i30. I want to make apparent what people take for granted about society.	3.31 (1.14)	-0.15	-0.70	3.84 (1.07)	-0.74	-0.28	-0.48
a35. I use the subject matter as a way to teach about higher ideals.	3.51 (0.99)	-0.25	-0.40	3.68 (0.98)	-0.57	-0.26	-0.17
a40. I emphasize values more than knowledge in my teaching.	2.92 (0.95)	0.26	-0.11	3.67 (1.03)	-0.26	-0.90	-0.76
a45. I link instructional goals to necessary changes in society.	3.02 (1.08)	0.01	-0.60	3.87 (1.01)	-0.81	-0.01	-0.81

Note. Effect size = (Mean for U.S. – Mean for Malaysia)/ Pooled *SD*.

Table 13 shows the descriptive statistics of the overall TPI model. Just like the individual items, the scores of the subscales were somewhat similar across the two groups. Both the U. S. and the Malaysian samples showed Apprenticeship as their dominant perspective. The Malaysian sample, however, was higher than the U.S. on the Transmission scale (effect size of Cohen's *d* of 0.58, moderate effect), the Apprenticeship scale (0.30, small) while the biggest difference was on the Social Reform scale (1.04, very large effect). The U.S., on the other hand, was slightly higher on the Developmental scale (0.12, small effect) and virtually the same on the Nurturing scale (0 effect). However, the interpretation of the resuts of these descriptives remained tentative in view of the invariance testing to be carried out.

Internal consistency reliability for the TPI with 45 items as well as the subscales with nine items each was tested for each group using Cronbach's alpha (Table 14). The internal consistency of the TPI measured with all 45 items for the U. S. group was .88 with an average inter-item correlation of .14. The item-to-total correlations ranged from .06 to .59. Meanwhile,

Table 13

Sum of Scores for the Five Perspectives for the U.S. (n=605) and Malaysian (n=561) Samples

Scale	Country	Min.	Max	*M*	*SD*	Skew	Kurtos	Effect Size
Transmission	U.S.	20.00	45.00	33.42	4.15	0.03	0.00	0.58
	Malay	21.00	45.00	35.75	3.87	-0.25	-0.11	
Apprenticeship	U.S.	20.00	45.00	36.57	4.13	-0.55	0.33	0.30
	Malay	24.00	45.00	37.78	3.92	-.030	-0.08	
Developmental	U.S.	13.00	44.00	35.37	4.00	-0.55	1.40	-0.12
	Malay	22.00	45.00	34.83	5.24	-0.56	0.07	
Nurturing	U.S.	12.00	45.00	34.81	5.44	-0.66	0.62	0.00
	Malay	20.00	45.00	34.79	5.58	-0.10	-0.93	
Social Reform	U.S.	11.00	45.00	28.06	6.02	-0.02	-0.20	1.04
	Malay	19.00	45.00	34.10	5.62	-0.10	-0.91	

Note. Effect size = (Mean for U.S. – Mean for Malaysia)/ Pooled *SD*. Potential range of scores was from 9 to 45.

the internal consistency of the TPI for the Malaysian group was higher at .93 which also had a larger range of item-to-total correlation starting from .03 reaching up to .68 with an average inter-item correlation that was also slightly higher at .23. As for the subscales, the U.S. sample had Cronbach's alphas that were slightly higher ranging from .67 for both Transmission and Developmental to .83 for Nurturing. The Malaysian sample's subscale alphas ranged from .59 for Transmission to .81 for Social Reform.

However, the average inter-item correlations of both the U. S. and the Malaysian groups were quite similar. The U.S. sample yielded average correlations ranging from the lowest for Transmission and Developmental with both at .19 to the highest for both Nurturing and Social Reform at .35. The average correlations for the Malaysian group ranged from .15 for Transmission to .32 for both Nurturing and Social Reform. The ranges of item-to-total

correlation for the subscales were similar for Apprenticeship and Nurturing for both groups but for the Transmission scale the range for the Malaysian sample was much larger when compared to the U.S. group which ranged from .06 to .50. Meanwhile the Transmission scale item-to-total correlations ranged from .26 to .45. Overall, the item-to-total correlations for both groups were acceptable but not overly high. The internal consistency for both groups was also on the lower side considering the moderate number of items (9) making up each scale of the TPI.

As mentioned earlier, many studies have been carried out with the TPI but at the time of this writing, only one study by Brown and Lake (2006) carried out a confirmatory factor analysis (CFA) of the TPI. However, they reported that the hierarchical model as proposed by Pratt and Collins (2001) was not a viable model. Therefore, they tested a modified correlated four-factor model based on only 11 of the original items under the subscales they labeled as Apprenticeship-

Table 14

Internal Consistency of TPI Subscales with Nine Items

Subscale	U.S. (n = 605)			Malaysia (n = 560)		
	Cronbach's α	Item-to-total correlation range	Average correlation	Cronbach's α	Item-to-total correlation range	Average correlation
Transmission	.67	.26 - .45	.19	.59	.06 - .50	.15
Apprenticeship	.72	.19 - .59	.24	.73	.14 - .50	.24
Developmental	.67	.04 - .50	.19	.78	.11 - .68	.29
Nurturing	.83	.38 - .67	.35	.80	.32 - .61	.32
Social Reform	.83	.33 - .70	.35	.81	.17 - .72	.32
All 45 items	.88	.06 - .59	.14	.93	.03 - .68	.23

Developmental, Nurturing, Social Reform, and Transmission. There were only two items under their Transmission scale of which one was a new item "My intent is to prepare people for examinations" and b11 (Effective teachers must first be experts in their own subject [areas]). Under the Apprenticeship-Developmental scale, the three items were b8 (Teaching should focus on developing qualitative changes in thinking), i18 (My goal is to help people develop more complex ways of reasoning) and a32 (I link the subject matter with real settings of practice or application). For the Nurturing scale the chosen items were b9 (In my teaching, building self-confidence in learners is a priority), i19 (My goal is to build people's self-confidence and self-esteem as learners), and a39 (I encourage expressions of feeling and emotion). As for their Social Reform scale, the items included b10 (Individual learning without social change is not enough), i25 (I expect people to be committed to changing our society), and a new item "I help people see the need for changes in society". The new model was found to have an acceptable fit, χ^2 (76, N = 1398) = 541.1, TLI = .88, CFI = .91, RMSEA = .07, among Queensland teachers (Brown & Lake, 2006). The correlations among the scales ranged from .30 to .70 and the factor loadings were from .30 to .85. In this current study however, a full five-factor model (see Figure 1) as originally proposed by Pratt and Collins (2001) was tested. The five-factor structure of the TPI includes 45 items grouped under five factors or perspectives of teaching called Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform where all the factors had nine items each. It was essential to ascertain via CFA if the five-factor model fit well for both samples before invariance testing could be carried out to determine if the measure was functioning equivalently across the two groups. The CFAs for both groups were conducted with

Mplus 7.0 (Muthén & Muthén, 2012) and the robust Maximum Likelihood (MLR) estimation method was used.

CFA for U. S. Faculty. When the correlated, five-factor model using the 45 items did not converge in the estimation process even after iterations were set at 25000, a process of step by step elimination of one item at a time to ascertain which item was causing the problem was undertaken. This process identified item b3 (Developmental: Most of all, learning depends on what one already knows) as the problem and therefore this item was dropped from the model. The 44 item model was successfully identified but the fit was found to be less than acceptable. Both the chi-square, χ^2 (892, N = 605) = 2539.71, $p < .001$ and the CFI (.75) indicated a lack of fit for the five-factor model underlying the TPI but the RMSEA (.06) met the criterion for acceptable fit. The standardized loadings ranged from .19 to .79. For the Transmission scale, standardized loadings ranged from .32 for item b6 to .58 for item a41, while the Apprenticeship scale had the lowest standardized loading for item b12 (.21) and the highest for item i17 with a loading of .76. After the exclusion of b3 from the Developmental scale, item b13 had the lowest standardized loading (.19) and item a38 had the largest standardized loading (.68). For the Nurturing scale, item b14 and item i19 had the lowest and highest standardized loadings of .42 and .76, respectively. Lastly, item b5 had the lowest standardized loading (.32) within the Social Reform scale and item a45 had the highest standardized loading (.79).

The correlations between the scales were positive and mostly low to medium in size with the lowest correlation between Social Reform and Transmission (r = .06) and the strongest between Social Reform and Nurturing (r = .66). Collins and Pratt (2011) reported the lowest correlation was r = .15 between Transmission and Nurturing and the highest was r = .58 between

Apprenticeship and Developmental which is higher than the U.S. sample in this study. However, another study carried out in China (Wang, 2012) found that the lowest interscale correlation was between Transmission and Nurturing (r = .04) and the highest was between Apprenticeship and Social Reform (r = .77). The rest of the interscale correlations for the U. S. sample are shown in Table 15.

Table 15

Correlations among the Five Scales for U. S. Sample (n = 605)

	Transmission	Apprenticeship	Developmental	Nurturing
Transmission				
Apprenticeship	.43*			
Developmental	.07	.43*		
Nurturing	.12	.50*	.55*	
Social Reform	.06	.43*	.65*	.66*

*p <.001

Analysis of the potential sources of misfit in the model began with an examination of the modification indices (MIs) reported in Mplus. Table 16 shows that there were 64 MIs greater than 10.83 (critical value for the χ^2 with 1 degree of freedom at the .001 level of significance). The four pairs of items that had high correlated errors had modification indices ranging from 55.23 for item i24 (Nurturing, I expect that people will enhance their self-esteem through my teaching) with item i19 (Nurturing, My goal is to build people's self-confidence and self-esteem as learners) to 83.54 for item i19 (Nurturing, My goal is to build people's self-confidence and self-esteem as learners) with item b9 (Nurturing, In my teaching, building self-confidence in

learners is a priority). Both i19 and b9 are from the Nurturing scale and the correlated error may be due to similar item wording and content.

Table 16

Correlated Errors as Indicated by the Modification Indices Reported for the Five-Factor Confirmatory Factor Analysis Model for the Teaching Perspective Inventory for the U. S. Sample (n = 605)

Modification Indices	Pairs of Items
10.00 -19.00	a44-b15, i28-i20, a44-a40, a37-i18, a37-i17, a40-i21, a37-a31, a39-a34, i23-i21, i30-b14, a39-i25, i23-i18, i21-b14, a42-b7, i29-b4, b14-b7, b8-b1, a42-b2, a45-a44, i19-i18, a32-b2, b10-b9, a42-a40, a44-b9, i27-i26, i21-i20, i26-b12, a44-i18, i30-i20, b10-b2, a37-a34, a38-a37, b12-b11, a33-a32, a35-b10, b7-b6, i25-i24, a45-i25, a43-a42, i30-i28, i26-b11, a36-b1, i23-i22, a34-i17, i24-b4, a40-i25, b10-b5, a34-a33, a40-a35
20.00 - 29.99	i21-i17, i24-b9, a39-i24, i29-i24, i22-i21, i27-i17, a39-i19, a37-a36, a44-i19, b6-b2, a39-b4, a40-a39, i28-i27, a39-b9
30.00 - 39.99	-
40.00 - 49.99	b11-b6, b7-b2
50.00 - 59.99	i24-i19
60.00 - 69.99	a44-a39
70.00 -79.99	i26-i16
80.00 - 89.99	i19-b9

CFA for Malaysian Faculty. The correlated, five-factor model with 44 items for the Malaysian sample produced even less acceptable fit compared to the U.S. faculty. The chi-square indicated a statistically significant lack of fit, χ^2 (892, N = 561) = 7783.63, $p < .001$ and both the

CFI (.43) and RMSEA (.12) revealed the same less than desirable fit. The standardized loadings ranged from .10 to .77. For the Transmission scale, standardized loadings ranged from .10 for item b6 to .63 for item i26, while the standardized loadings for the Apprenticeship scale ranged from .28 for item a42 to .65 for item i22. Meanwhile, item b13, a loading of .06, was the lowest for the Developmental scale and item i23 had the largest standardized loading (.72). As for the Nurturing scale, item b4 had the lowest and item i29 had the highest loadings of .22 and .73, respectively. Lastly, the Social Reform scale had the lowest loading for item b5 at .10 and the highest loading of .77 for item i25.

The correlations between the factors were generally higher than the U.S. sample. Higher than perfect correlations ($r = 1.0$) were obtained for two pairs of factors, Nurturing-Developmental and Social Reform-Nurturing (these Heywood cases, correlations greater than 1.0, may be due to random sampling error). Correlations for the other factors ranged from .46 for Social Reform-Transmission to .94 for Social Reform-Developmental. The rest of the correlations among the scales are shown in Table 17.

A major source of misfit in the model involved high correlated errors among many items (see Table 18) with modification indices ranging from 52.63 for the correlation between the errors for item a38 (I challenge familiar ways of understanding the subject matter) and item a35 (I use the subject matter as a way to teach about higher ideals) to 211.93 for item b15 (For me, teaching is a moral act as much as an intellectual activity) and item b12 (Knowledge and its application cannot be separated). These correlated errors may be due to method effects such as similarity in wording or meaning. There were a total of 389 MIs greater than 10.83 (critical value for the χ^2 for 1 degree of freedom at the .001 level of significance).

Table 17

Correlations among the Five Scales for the Malaysian Sample (n = 561)

	Transmission	Apprenticeship	Developmental	Nurturing
Transmission				
Apprenticeship	.68*			
Developmental	.66*	.86*		
Nurturing	.64*	.87*	1.00[a*]	
Social Reform	.46*	.89*	.94*	1.00[a*]

*$p < .001$. [a]Values exceeded 1.0 (i.e., 1.01) and were set to 1.0.

Confirmatory Factor Analyses of Each Scale from the TPI for Each Country

Since the overall fit for the five-factor model was poor for both groups, it was decided to analyze the fit of each of the five factors separately for each group to identify problems with the items.

Transmission. As shown in Table 19, the fit of the one-factor model of the nine item Transmission scale for the U. S. sample was less than adequate, χ^2 (27, N = 605) = 158.39, $p <$ .001, and the CFI (.75) and the RMSEA (.09) also showed less than desirable fit where the RMSEA should be less than .06. The lowest standardized loading was for item b6 at .03 and the highest was .58 for item a41. There were three modification indices higher than the critical value

Table 18

Correlated Errors as Indicated by the Modification Indices Reported for the Five-Factor Confirmatory Factor Analysis Model for the Teaching Perspective Inventory for Malaysian Sample (n = 561)

Modification Indices	Pairs of Items
50.00 - 59.99	b11-b2, a35-i24, b9-b1, a31-b11, a38-a35, a38-i27, a39-i24, a35-i25, i26-i24, b13-b12, a39-i27, a44-i24, a34-i30, i30-I28, a34-i24, i22-i21, i30-i26, a35-a34, a34-b10, b15-a13, a45-i26, b14-b1
60.00 - 69.99	i30-b14, a34-a33, b11-b7, a43-a35, i30-b2, b10-b1, b11-b10, b9 -b7, b10-b8, a45-b14, a38 -a37, b10-b7, a45-a43, b14-b11, b10-b5, b12-b5, i24-i23,
70.00 -79.99	a45-a34, a34-i26, a40-i24, i25-i23, b15-b11, i27-i24, a35-b9, i18-b14, i19-b9, b9-b8,
80.00 - 89.99	a34-i18, a36-a31, b12-b11, b12-b8, i25-i24, a38 -i24, b8-b7
90.00 -99.99	b10-b2, b8-b1, b15-b7
100.00 -199.99	b14-b2, b14-b12, b15-b10, b12-b2, a34-b14, b7-b1, a39-a38, b15-b8, b12-b10, b45-I30, b12-b7, b14-b10, b15-b14, b15 -b1, b12-b1
Above 200.00	b15-b12

of 10.83 which were, in descending order, for items i26-i16 (68.08), b11-b6 (45.88), a36-b1 (16.23), and i26-b11 (11.81).

The fit of the Transmission scale for the Malaysian sample was much worse, χ^2 (27, N = 561) = 300.83, $p < .00$, and the same lack of fit was also revealed by the CFI (.57) and the

RMSEA (.13) as shown in Table 19. The standardized loadings, on the other hand, showed that item a36 had the highest (.73) and item b1 had the lowest loading (.07). This lack of fit was further evidenced by the high number of modification indices above the critical value of 10.83. The modification indices for the Malaysian sample ranged from 11.64 for items a41 and a31 to 66.46 for items b11 and b1.

For the purpose of invariance testing a model with acceptable fit must be achieved. In order to do this, model respecification was required for both samples and this was done by adding more parameters into each model. In Mplus 7, this was carried out by adding correlated error terms to the model for the pairs of items that had the highest modification index. The fit indices were then inspected to see if the fit had improved to an acceptable level. If not, the next highest modification index was added along with the first one.

For the U. S. Transmission scale, two correlated error terms representing the highest modification indices for the pairs, i26-i16 and b11-b6, were added in the model to achieve the required fit. As shown in Table 20, the fit for the modified Transmission model for the U. S. sample was χ^2 (25, N = 605) = 47.92, p = .01, and the other fit indices also revealed that this was a much better fitting model (CFI = .96, RMSEA = .04) with a loss of 2 degrees of freedom. The standardized loadings changed from .29 to .23 for item b6 which was the lowest loading and from .58 to .63 for item a41 as the highest loading.

The same procedure was carried out to improve the fit for the Malaysian Transmission model. As shown in Table 19, the correlated errors of b11-b1, i16-b6, and a36-a31 were added to the model to achieve improved fit, χ^2 (24, N = 561) = 127.30, p = .01 (CFI = .84, RMSEA = .09). For the standardized loadings, the lowest loadings increased slightly to .11 for item b1 and the

highest loading was for item i26 (.70). More than three parameters were added to the model but no real improvements could be achieved. So, the most parsimonious model above was finally chosen; despite these modifications the fit for the Transmission model in the Malaysian sample was still below acceptable levels.

Table 19

Confirmatory Factor Analysis of the Five Scales of the TPI for the U.S. (n = 605) and Malaysian (n = 561) Samples

Scale		X^2	*df*	CFI	RMSEA	SRMR
Transmission	U. S.	158.39	27	.75	.09	.06
	Malay	300.83	27	.57	.13	.09
Apprenticeship	U. S.	115.17	27	.89	.07	.06
	Malay	498.08	27	.53	.18	.12
Developmental	U. S.	60.89	20*	.93	.06	.04
	Malay	233.84	20*	.80	.14	.09
Nurturing	U. S.	260.07	27	.83	.12	.07
	Malay	369.80	27	.73	.15	.09
Social Reform	U. S.	112.99	27	.93	.07	.04
	Malay	520.22	27	.67	.18	.11

Note. For Developmental Scale, Item B3 was omitted from the analyses.

Apprenticeship. For the U. S. sample, the fit of the one-factor model as shown in Table 18 was marginally acceptable, χ^2 (27, N = 605) = 115.17, p = .01 (CFI = .89, RMSEA = .07) so a correlated error term for the errors of items b2-b7 parameter was added to make it a better fitting model. As shown in Table 18, the fit for the modified Apprenticeship model for the U. S. sample was χ^2 (25, N = 605) = 73.00, p = .01, and the other fit indices also revealed that this was a much

better fitting model (CFI = .94, RMSEA = .06) with a loss of 1 degree of freedom. The standardized loadings changed for the lowest loading item of b2 from .29 (original model) to .27 (modified model with the added correlated error) but remained the same for the highest loading item of i17 at .80.

To improve the fit for the Malaysian Apprenticeship model, three new correlated error terms for the errors for b7-b12, b2-b12, and b2-b7 were added to the model; these modifications resulted in marginally acceptable fit, χ^2 (24, N = 561) = 144.05, p < .01 (CFI = .88, RMSEA = .09.) As for the standardized loadings, the lowest loading was originally for a42 (.22) but now in the modified model was for b7 (.12) and the highest loading increased for i17 from .69 to .73.

Developmental. Item b3 was dropped from the Developmental scale after it was discovered to be the source of the model's failure to converge, even after greatly increasing the number of iterations. As a result, the fit achieved was found to be reasonably adequate and no further modification was deemed necessary, χ^2 (20, N = 605) = 60.89, p < .01 (CFI = .93, RMSEA = .06) for the U. S. sample. The lowest standardized loading was for b13 (.16) and the highest standardized loading was for a38 (.69).

The Malaysian Developmental scale demonstrated marginally adequate fit, χ^2 (20, N = 561 = 233.84, p = .00 (CFI = .80, RMSEA = .14) and so the decision was made to further improve the model by adding two correlated error terms namely b13-i18 and i23-a33. The fit was further improved to χ^2 (18, N = 561) = 117. 10, p < .01 (CFI = .91, RMSEA = .10) was similar to those of the U. S. sample. The standardized loadings ranged from .29 (b13) to .78 (a38).

Nurturing. To improve the fit for the U. S. Nurturing scale, χ^2 (27, N = 605) = 260.07, p < .01 (CFI = .83, RMSEA = .12), two new parameters had to be added to the model. The

addition of two correlated error parameters for a39-a44 and b9-i19, resulted in more acceptable fit, χ^2 (25, N = 605) = 152.59, p < .01 (CFI = .91, RMSEA = .09). The standardized loadings ranged from .15 (b14) to .76 (i19).

The fit of the Malaysian Nurturing scale, χ^2 (27, N = 561) = 369.80, p < .01 (CFI = .73, RMSEA = .15) was much worse than the U. S. model. An additional four parameters representing correlated errors (b14-a34, i24-a44, i24-a34, and b9-i19) were added to the model but the fit was still not acceptable, χ^2 (23, N = 561 = 162.44, p < .01 (CFI = .89, RMSEA = .10). The standardized loadings ranged from .02 (b14) to .77 (i29).

Social Reform. The U. S. Social reform scale had good fit at χ^2 (27 N = 605) = 112.99, p < .01 (CFI = .93, RMSEA = .07) but it was also deemed necessary to improve the model by adding another correlated error parameter for the errors associated with items a40 and a35. The resulting fit was an improved model, χ^2 (26, N = 605) = 85.70, p < .01 (CFI = .95, RMSEA = .06). The standardized loadings ranged from .35 (b5) to .80 (a45).

The Malaysian Social Reform model had poor fit, χ^2 (27 N = 561) = 520.22, p < .01 (CFI = .67 , RMSEA = .18). Model fit improved to an acceptable level with the addition of four correlated error terms (a45-i30, b15-b5, b15-b10, and b10-b5). The fit of the revised model was χ^2 (23 N = 561) = 92.93, p < .01 (CFI = .95, RMSEA = .07). The standardized loadings ranged from .10 (b5) to .84 (i25).

Table 20

CFA of the Five Scales of the TPI for U. S. and Malaysian Samples with Correlated Errors

Scale	Corr. Errors	Country	X^2	*df*	CFI	RMSEA	SRMR
Transmission	i26-i16,b11-b6	U. S.	47.92	25	.96	.04	.03
	b11-b1, i16-b6, a36-a31	Malaysia	127.30	24	.84	.09	.06
Apprenticeship	b2-b7	U. S.	73.00	26	.94	.06	.05
	b7-b12, b2-b12, b2-b7	Malaysia	144.05	24	.88	.09	.06
Developmental	None	U. S.	49.40	20*	.95	.06	.05
	b13-i18, i23-a33	Malaysia	75.41	18*	.94	.08	.05
Nurturing	a39-a44, b9-i19	U. S.	146.74	25	.91	.09	.05
	b14-a34, i24-a44, i24-a34, b9-i19	Malaysia	199.90	23	.87	.12	.08
Social Reform	a40-a35	U. S.	85.70	26	.95	.06	.04
	a45-i30, b15-b5, b15-b10, b10-b5	Malaysia	92.93	23	.95	.07	.04

*For the Developmental Scale, Item b3 was omitted from the analyses.

Invariance Testing of Each of the Scales from the TPI

The overall inadequate fit of the five-factor model underlying the TPI precluded a full test of measurement invariance for the five-factor TPI. Based on the results from the previous sections, the decision was made to examine the measurement invariance of each of the TPI factors separately with the caveats that these individual factors do not represent Pratt's overall model and that these individual factor models had been modified to include one or more

correlated errors. As shown in Table 20, the Developmental and Social reform scales proved to be the best models for both countries. For the other three scales, the baseline models were problematic at best.

Configural invariance. The first step in carrying out a measurement invariance testing of the TPI was to ascertain whether the pattern or configuration of each of the five TPI scales was similar for the U. S. and Malaysian samples. This is the configural invariance testing level (Widaman & Reise, 1997) which is a prerequisite that must be fulfilled before considering the next level of invariance testing. To evaluate configural invariance, the correlated five factor TPI model was not carried out because the fit for this overall model was poor for both countries and the decision was to test each scale one at a time. Therefore, each scale with its nine indicators except for the Developmental scale, which had eight items after item b3 was dropped due to a non-convergence issue, was constrained to have the same pattern for both countries. Except for the U. S. Developmental scale, all the other scales were modified to include correlated errors to improve the fit in order to obtain acceptable baseline models (see Table 20) to facilitate comparisons between the two countries. This modified one-factor model for each country was run and the combined chi squares were used as the baseline model.

Metric invariance. Configural invariance is a necessary but not a sufficient condition for comparing the mean scores on the TPI across the two countries. A stronger test of invariance is required, which is the second level of Widaman and Reise's (1997) measurement invariance testing; this stronger form of invariance is called metric invariance. This invariance testing level presupposes that the item loadings (i.e., relations of the items to the factor) are equal across the two countries. To test for metric invariance, the loadings for the U.S. and Malaysian samples

were constrained to be equal and the change in chi square from the baseline model was used to evaluate if the assumption of equal loadings was tenable. Each time the overall null hypothesis of equal loadings was found to be not tenable and was rejected, follow up comparisons of individual items were conducted. To account for the eight multiple follow-up comparisons (i.e., each individual item), a Bonferroni correction of .05/8 was used for the significance level to reject the null hypothesis. The chi-square difference value along with the corresponding change in degrees of freedom was compared to the critical chi-square value at the .006 level of significance. This procedure was carried out for each of the five scales.

If the loadings for the nine items that make up the Transmission scale are the same across the U.S. and Malaysian groups, then measurement invariance is supported. As shown in Table 21, The Transmission metric invariance model (i.e., all the factor loadings constrained to be equal) was compared to the baseline configural model. The difference χ^2 test between the baseline and the metric invariance model based on the Satorra-Bentler scaled chi square (Satorra & Bentler, 2010) was significant at χ^2 (8, N = 1166) = 28.41, p < .001 and the CFI (.86) and RMSEA (.07) revealed marginal fit. To detect the source of the significant difference, an item by item analysis was carried out for the scale. The assumption of equal loadings for items i16, i26, and a31 were found not to be tenable, that is, the null hypothesis of equal loadings for these three items was rejected. It was decided that the scale be tested in the next level of invariance testing which was, the scalar invariance testing. The baseline model that was used included equality constraints on loading for both countries for all the items except items i16, i26 and a31, which were set free to vary across groups.

Table 21

Transmission Metric Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjust-ed with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Configural (b1= reference)	173.99			49		.89		.07		.05	
Metric Invariance	223.18	28.41*	**Yes**	57	8	.86	.03	.07	0	.07	.02
Metric_b6	174.87	0.51	No	50	1	.89	0	.07	0	.00	0
Metric_b11	181.12	7.14	No	50	1	.89	0	.07	0	.05	0
Metric_i16	182.94	9.13*	**Yes**	50	1	.89	0	.07	0	.05	0
Metric_i21	181.11	5.70	No	50	1	.89	0	.07	0	.05	0
Metric_i26	193.27	16.30*	**Yes**	50	1	.88	.01	.07	0	.06	.01
Metric_a31	184.71	8.43*	**Yes**	50	1	.88	.01	.07	0	.06	.01
Metric_a36	180.88	5.85	No	50	1	.89	0	.07	0	.05	0
Metric_a41	172.14	0.29	No	50	1	.90	.01	.07	0	.05	0
Everything equal except i16, i26 & a31	201.53	26.53*	**Yes**	54	5	.87	.02	.07	0	.07	.02

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

Table 22 summarizes the tests of metric invariance for the Apprenticeship scale. The overall test of equal loadings showed a significant difference between the baseline model (configural) and the metric invariance model, $\Delta \chi 2$ (8, N = 1166) = 34.36, p < .001. When individual items were tested using the Satorra-Bentler Scaled Chi Square difference test, none of the items revealed statistically significant difference, that is, the assumption of equal loadings for all the items was revealed to be tenable. The scale was therefore deemed suitable for the next level of invariance testing and the baseline model to be used was the metric invariance model.

Table 22

Apprenticeship Metric Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjust-ed with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Configural (b5 = reference)	214.21			50		.91		.08		.05	
Metric Invariance	248.58	34.36*	**Yes**	58	8	.89	.02	.08	0	.08	.03
Metric_b7	219.04	4.83	No	51	1	.91	0	.08	0	.06	.01
Metric_b12	214.75	0.44	No	51	1	.91	0	.07	.01	.05	0
Metric_i17	214.11	0.34	No	51	1	.91	0	.07	.01	.05	0
Metric_i22	213.84	0.38	No	51	1	.91	0	.07	.01	.05	0
Metric_i27	214.41	0.01	No	51	1	.91	0	.07	.01	.05	0
Metric_a32	214.41	0.98	No	51	1	.91	0	.07	.01	.06	.01
Metric_a37	216.08	2.35	No	51	1	.91	0	.08	0	.06	.01
Metric_a42	216.61	1.67	No	51	1	.91	0	.08	0	.06	.01

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

The overall test of equal loadings for the Developmental scale as shown in Table 23 showed a significant difference between the baseline model and the metric invariance model, $\Delta \chi^2$ (7, N = 1166) = 37.19, p < .001. However, both the CFI (.93) and the RMSEA (.07) showed an acceptable fit. When individual items were tested using the Satorra-Bentler Scaled Chi Square difference test, items i23 and a38 revealed statistically significant differences. The assumption of equal loadings for the rest of the five items was found to be tenable and the null hypothesis for the assumption of equal loadings was not rejected for these five items. The scale was also brought forward for the next level of invariance testing with the baseline model including equal loadings for all items except items i23 and a38, which were set to be freely estimated for the U.S. and Malaysian groups.

Table 23

Developmental Metric Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.007 (adjusted with .05/7)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Configural (b8_ref)[a]	124.66			38		.94		.06		.04	
Metric Invariance	160.10	37.19*	**Yes**	45	7	.93	.01	.07	.01	.07	.03
Metric_b13	125.91	0.94	No	39	1	.94	0	.06	0	.04	0
Metric_i18	125.85	0.01	No	39	1	.94	0	.06	0	.04	0
Metric_i23	133.57	15.18*	**Yes**	39	1	.94	0	.06	0	.05	.01
Metric_i28	126.56	1.40	No	39	1	.94	0	.06	0	.04	0
Metric_a33	126.46	1.89	No	39	1	.94	0	.06	0	.05	.01
Metric_a38	131.72	8.40*	**Yes**	39	1	.94	0	.06	0	.05	.01
Metric_a43	127.31	2.38	No	39	1	.94	0	.06	0	.05	.01
Everything but i23 and a38	136.83	11.81	No	43	5	.94	0	.06	0	.06	.02

*$p < .001$. Yes = assumption of equality rejected. No = assumption of equality is tenable. [a]B3 is dropped from the list.

In Table 24, the metric invariance testing of the Nurturing scale produced an outcome that also revealed a significant difference between the baseline model and the metric invariance model, $\Delta \chi^2$ (8, N = 1166) = 105.66, $p < .001$. Both the CFI (.86) and the RMSEA (.11) confirmed the poor fit of the model as well. When individual items were tested using the Satorra-Bentler Scaled Chi Square difference test, all items showed a statistically significant difference and the assumption of equal loadings for these items was found to be not tenable and the null hypothesis was rejected. The scale was considered to be not credible for the next level of invariance testing.

Table 24

Nurturing Metric Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjust-ed with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Configural (b4_ref)	316.29			48		.90		.10		.06	
Metric Invariance	420.65	105.66*	**Yes**	56	**8**	.86	.03	.11	.01	.11	.05
Metric_b9	364.92	23.22*	**Yes**	49	**1**	.89	0	.11	.01	.07	.01
Metric_b14	356.75	26.87*	**Yes**	49	**1**	.89	0	.10	0	.06	0
Metric_i19	354.15	18.50*	**Yes**	49	**1**	.89	0	.10	0	.07	0
Metric_i24	354.04	21.00*	**Yes**	49	**1**	.89	0	.10	0	.06	0
Metric_i29	392.27	38.44*	**Yes**	49	**1**	.88	0	.11	.01	.09	.01
Metric_a34	357.60	25.55*	**Yes**	49	**1**	.89	0	.10	0	.07	0
Metric_a39	376.90	32.54*	**Yes**	49	**1**	.88	.01	.11	.01	.08	.02
Metric_a44	366.45	29.69*	**Yes**	49	**1**	.89	.01	.11	.01	.07	.01

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

Metric invariance testing of the Social Reform scale (see Table 25) revealed a significant difference between the baseline model and the metric invariance model, $\Delta \chi^2$ (8, N = 1166) = 139.17, p < .001 (CFI = .91, RMSEA = .09). Follow-up comparisons of the loadings for items b10, i20, i25, i30, a35, and a40 were significantly different using the Satorra-Bentler Scaled Chi Square test. The assumption of equal loadings across the two groups was found to be tenable only for items b5 and a45. The decision was to continue to the next invariance testing for this scale by constraining the loadings for items b5 and a45 to be equal while letting the remaining six items identified above as significantly different to vary across groups.

Table 25

Social Reform Metric Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjust-ed with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Configural (i15_ref)	178.16			49		.95		.07		.04	
Metric Invariance	306.75	139.17*	**Yes**	57	8	.91	.04	.09	.02	.12	.08
Metric_b5	179.13	1.00	No	50	1	.95	0	.07	0	.04	0
Metric_b10	199.56	31.50*	**Yes**	50	1	.95	0	.07	0	.06	.02
Metric_i20	212.90	53.74*	**Yes**	50	1	.94	.01	.07	0	.07	.03
Metric_i25	213.12	46.25*	**Yes**	50	1	.94	.01	.08	.01	.0	.03
Metric_i30	198.50	25.47*	**Yes**	50	1	.95	0	.07	0	.06	.02
Metric_a35	206.53	40.69*	**Yes**	50	1	.94	.01	.07	0	.07	.03
Metric_a40	235.46	97.85*	**Yes**	50	1	.93	.02	.08	.01	.08	.04
Metric_a45	184.91	7.08	No	50	1	.95	0	.07	0	.05	.01
Everything except b10, i20, i25. i30, a35 & a 40	186.26	8.24	No	52	2	.95	0	.07	0	.05	.01

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

Scalar invariance. The next step in the invariance testing process was to assess scalar invariance or the equality of the item intercepts (Widaman & Reis, 1997). This level of invariance concerns the equality of item intercepts in the regression equations that connect the observed variables to their latent construct. Both the factor loadings and the intercepts for the items are constrained to be equal for both groups in order to establish evidence for scalar or strong factorial invariance.

Table 26 summarizes the scalar invariance tests for the Transmission scale. When the loadings and the intercepts for all the items were constrained to be equal across the two groups, the chi square difference based on the Satorra-Bentler adjustment scale revealed a significant

difference as compared to that of the baseline model. Scalar invariance tests on the five individual items identified as metric invariant were also carried out to ascertain the source of the difference and item b11 and a41 were found to be significantly different. The rest of the items showed no significant difference from the baseline model and the assumptions of equal loadings and equal intercepts across groups were found to be tenable for these items.

Table 26

Transmission Scalar Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjusted with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Everything but i16, 26, a31	201.53			54		.87		.07		.07	
Scalar Invariance	279.28	81.33*	**Yes**	59	5	.81	.06	.08	.01	.07	0
Scal_b6	202.19	0.04	No	55	1	.87	0	.07	0	.07	0
Scal_b11	227.54	71.37*	**Yes**	55	1	.85	.02	.07	0	.07	0
Scal_i21	203.15	1.87	No	55	1	.87	0	.07	0	.07	0
Scal_a36	202.98	1.20	No	55	1	.87	0	.07	0	.07	0
Scal_a41	211.02	9.46*	**Yes**	55	1	.87	0	.07	0	.07	0

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

Table 27 shows the summary for the Apprenticeship scale's scalar invariance tests. The chi square difference based on the Satorra-Bentler adjustment scale revealed a significant difference between the scalar invariant model as compared to that of the metric model. Scalar invariance tests on the individual items to ascertain the source of the difference revealed six items (b12, i17, i22, i27, a32 and a37) as significantly different. Meanwhile items b7 and a42 showed no significant difference and the assumptions of equal loadings and equal intercepts across groups were found to be tenable for these two items.

Table 27

Apprenticeship Scalar Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjust-ed with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline	248.58			58		.89		.08		.08	
Scalar Invariance	441.96	204.04*	**Yes**	66	8	.79	.10	.10	.02	.12	.04
Scalar_b7	251.95	3.62	No	59	1	.89	0	.08	0	.08	0
Scalar_b12	264.60	17.83*	**Yes**	59	1	.88	.01	.08	0	.09	.01
Scalar_i17	260.79	10.08*	**Yes**	59	1	.89	0	.08	0	.09	.01
Scalar_i22	265.09	13.56*	**Yes**	59	1	.88	.01	.08	0	.09	.01
Scalar_i27	265.52	13.96*	**Yes**	59	1	.88	.01	.08	0	.09	.01
Scalar_a32	272.65	18.57*	**Yes**	59	1	.88	.01	.08	0	.09	.01
Scalar_a37	277.93	29.97*	**Yes**	59	1	.88	.01	.08	0	.09	.01
Scalar_a42	250.53	2.48	No	59	1	.89	0	.08	0	.08	0

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

Table 28 summarizes the scalar invariance testing for the Developmental scale which revealed the chi square difference based on the Satorra-Bentler adjustment scale between the scalar invariant model and that of the metric model was significantly different. Scalar invariance testing on the individual items to ascertain the source of the difference showed three items as significantly different (items b13, i28, and a38). The three items of i18, a33, and a43 showed no significant intercept differences and the assumptions of equal loadings and equal intercepts across groups were found to be tenable for these items.

Table 29 shows the summary for Social Reform's scalar invariance tests. The chi square difference based on the Satorra-Bentler adjustment scale revealed a significant difference between the scalar invariant model as compared to that of the metric model. Scalar invariance testing on the remaining two individual items to ascertain the source of the difference revealed item b5 was significantly different. Only item a45 showed no significant difference from the baseline model and the assumption of equal loadings and equal intercepts across groups was found to be tenable for the item.

Table 28

Developmental Scalar Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.007 (adjusted with .05/7)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Everything but i23	200.12			44		.91		.08		.07	
Scalar Invariance	485.82	324.29*	**Yes**	51	7	.74	.17	.12	.04	.10	.03
Scalar_b13	266.45	84.13*	**Yes**	45	1	.87	.07	.09	.01	.08	.01
Scalar_i18	201.94	2.41	No	45	1	. 91	0	.08	0	.07	0
Scalar_i28	218.72	15.22*	**Yes**	45	1	.90	.01	.08	0	.07	0
Scalar_a33	204.25	4.23	No	45	1	.91	0	.08	0	.07	0
Scalar_a38	220.64	17.32*	**Yes**	45	1	.90	.01	.08	0	.07	0
Scalar_a43	153.10	5.42	No	45	1	.91	0	.08	0	.07	0

*p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable. [a]B3 is dropped.

Table 29

Social Reform Scalar Invariance

Model	X^2	*Satorra-Bentler Scaled* ΔX^2	*P=.006 (adjusted with .05/8)*	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Baseline Everything except b10, i20, i25. i30, a35 & a 40	186.26			51		.95		.07		.05	
Scalar Invariance	551.18	337.77*	**Yes**	59	**8**	.83	.12	.12	.05	.16	.11
Scalar_b5	238.68	65.46*	**Yes**	52	**1**	.93	.02	.08	.01	.07	.02
Scalar_a45	188.47	1.81	No	52	**1**	.95	0	.07	0	.05	0

* p<.001. Yes = assumption of equality rejected. No = assumption of equality is tenable.

Table 30 summarizes the invariance testing outcome. In order to make valid comparisons of the mean scores between the two countries, the items on the instrument have to exhibit metric and scalar invariance. In other words, the item loadings and the item intercepts need to be equal for both samples. In Table 30, none of the scales were successful in meeting the invariance criteria. At the subscale level, only three items of the Transmission scale had equal loadings and intercepts, while Apprenticeship showed two items with equal loadings and intercepts, and

Developmental had three that met that requirement. Meanwhile, there was only one item that had equal loadings and intercepts in the Social Reform scale while Nurturing had none. The unexpected result considering the outcome of the translation and the adaptation process was item b6. It was one of the two problematic items that was found to have no linguistics equivalence in Malay but it was one of the transmission items that showed invariance in the final analysis.

Table 30

Summary of Invariance Testing

Perspective	Metric (Item Loadings)		Scalar (Item Intercepts)	
	Not Significantly Different	Significantly Different	Not Significantly Different	Significantly Different
Transmission (B1 = reference)	B6 B11 I21 A36 A41	I26 A31	B6 I21 A36	B11 A41
Apprenticeship (B5 = reference)	B7 B12 I17 I22 I27 A32 A37 A42		B7 A42	B12 I17 I22 I27 A32 A37
Developmental (B8 = reference)	B13 I18 I28 A33 A43		I18 A33 A43	B13 I28 A38

Table 30 (Continued)

Nurturing (B4 = reference)		B9 B14 I19 I24 I29 A34 A39 A44		
Social Reform (I15 = reference)	B5 A45	B10 I20 I25 I30 A35 A40	A45	B5

Summary

In summary, the psychometric analyses of the adapted TPI and the original TPI did not support measurement invariance as proposed by Meredith (1997) who stated that measures across groups are considered to be on the same scale if relationships between the indicators or items used to measure the latent trait are the same across groups, which require the equality of item factor loadings and item intercepts. Both metric invariance (equality of factor loadings) and scalar invariance (equality of intercepts) were found to be not tenable for the subscales as well as for many of the items in each of the subscales. There was no need for stronger invariance testing to assess for equality of item residual variances as the condition for the weak invariance was not even fulfilled. In other words, the measurement equivalence of TPI adapted into Bahasa Malaysia and that of the original English TPI has not been established even after a rigorous process of translation and adaptation.

Chapter Five:

Discussion

The major purposes of the study were to translate and adapt the Teaching Perspectives Inventory (Pratt, 1992, 1990) from English to Bahasa Malaysia and compare the psychometric properties of the two versions. These two purposes were realized by means of two major stages involving forward translation of the TPI, backward translation of the TPI, a panel review, a pilot study, and cognitive interviews for the first stage, and a psychometric evaluation (reliability, factor structure, invariance testing) of the original and the adapted instruments for the second stage. Discussion of the findings of this study will consist of five sections. A review of teaching beliefs and its significance to teaching in higher education as well as the construct of teaching perspectives as proposed by Pratt (1992) will be covered in the first section. The second section will explicate the findings from the translation and adaptation stage while the third section will discuss the psychometric evaluation stage. Findings from the two phases are synthesized in the fourth section to form the basis for discussion as to the significance of the current study in contributing to the area of research across linguistic and cultural boundaries. The final section will explore implications of the current findings on teacher beliefs in higher education, limitations, and suggestions for future research.

Teaching Beliefs in Higher Education

The renewed attention towards effective teaching and learning in higher education has not only benefited many nations in the western hemisphere but has also shown to be reaching other countries as well. The experiences of many faculties of the European and American universities in their struggles to improve teaching and learning would be of a great help for their colleagues in less developed nations like Malaysia. This can only happen if the experiences and lessons learned are proven to be similar and can be assimilated into localized settings. Therefore, there is a need to assess the cross-cultural generality of our theories and assumptions about teaching and learning before any comparisons can be made about them across different groups of people.

In a globalized world where physical and geo-political boundaries are fast disappearing, such knowledge can be shared among people all over the world quickly and effectively if socio-cultural and linguistic boundaries can be overcome as well. This can be done if issues like teachers' conceptions about teaching and learning can be shown to be equivalent across cultures. Due to the strong relationship between beliefs and behaviors (Menges, 1990) and the way that educators perceive teaching as somehow having an impact on student learning (Borrich, 1999; Clark, 1986; Kagan, 1992; Pajares, 1992; Thompson, 1992), discoveries of cross-cultural studies regarding these phenomena can be extremely beneficial for all. The two types of teaching conceptions of Gow and Kember (1993) have already seen some impact in the ways teaching and learning are viewed in higher education across the globe where more deep learning is called for as opposed to surface learning only (Biggs & Tang, 2011; Dolmans & Gijbels, 2013; Hartley, 2008; Ke & Chavez, 2013). Deep learning and active learning (Bonwell & Eison, 1991) are two

recent trends that many Malaysian educators are trying to promote as students in the country are known to be very passive learners (Nik, Nazli, & Maliah, 2013).

The Malaysian Education Act (1996) specifically pointed out the need to exert some quality control on the teaching and learning in the country's public universities and one of the ways is to look into the teaching methods of faculty. Just like the notion of deep learning and active learning, Pratt's (1992) view is that there is no one best way of teaching, thus proposing the five conceptions of teaching, which are then manifested in the teachers' beliefs, intentions, and actions in the classroom. Pratt's conceptualization of teaching culminated in the design of the TPI, which has made its way across national boundaries (Brown & Lake, 2006; Wang, 2013). The dissemination of this conceptualization has been even more prolific with translations of the TPI instrument having been carried out in some of the world's prominent languages like Spanish, German, French, Indonesian, Japanese and a major dialect of Chinese. Two studies have been carried out to assess the conceptual invariance of the TPI comparing the original version with Chinese versions (Lu, 2006; Ruan, 2004), as reported by Collins and Pratt (2012), but these studies did not provide any psychometric information about the translations (no formal invariance testing was reported).

The TPI has been used as an instrument to assist faculty development (Ratcliff & Rocco, 2003), which is consistent with Collins and Pratt's suggestion that it be used as "a discussion tool" to help educationists learn more about teaching (Collins & Pratt, 2011, p. 373). There may be a possibility that TPI be used as an evaluation instrument even though Collins and Pratt (2011) advised against the idea. The probability of the TPI being used in public universities of Malaysia as one of the instruments for faculty evaluation and development is also to be expected

as the government hastens its move to make the country the center of educational excellence in the region. So far, no study has shown that the teaching conceptions as measured by the TPI are equivalent across languages and cultures. According to Johnson (1998), the equivalence of survey items is more important than reliability and validity. Since cross-cultural research in Malaysia is scarce (Fontaine & Richardson, 2003) and this is especially so in higher education, methodological insights into the process and procedures of obtaining an invariant research intrument are still an unfulfilled requirement. The need for an invariant instrument such as the TPI is crucial because experts agree that borrowing instruments for research without checking their relevance and equivalence in other countries and contexts is seriously defective (Chapman & Carter, 1979; Douglas & Nijssen, 2003). Even though Collins and Pratt (2011) have reported that the TPI scales were unbiased across criterion groups, there is no study except for the current one that reports the measurement invariance property of the instrument.

Challenges in Translation and Adaptation of Research Instruments

According to Acquadro, Conway, Hareendran, and Aaronson (2008) there are many challenges to producing translated instruments of high quality. They argue that the development of good translated instruments requires extensive work, and because guidelines to improve the effectiveness of translated instruments are lacking, they advise using a variety of methods to overcome these challenges. Meanwhile, Maneesriwongul and Dixon (2004) strongly suggest that even though all instruments that are used across cultures need to employ many techniques and approaches to translation and adaptation, there is yet no agreement as to which ones to use. However, relying solely on back translation alone is not an option that is recommended. In Pratt's recommended 24 steps to translate the TPI (see Appendix B), the first step in the

translation process is to look for questions or items that have an approximate equivalent in the target language. However, there was no mention of steps to deal with items such as those that do not have equivalent forms in the target language (e.g., Bahasa Malaysia language for items such as b6 and a35 found in this study).

For this particular study, the translation of the instrument employed some of the steps as proposed by McGorry (2000) who also advocated a combination of survey translation methods. In this case, forward and back translations were carried out with the recommended two or more translators for each step. The wisdom of employing more than one translator came to bear when one of the initial translators in this study acknowledged that she was not proficient enough in English to confidently find equivalent forms in Malay for a number of items. Besides, compared to the other two initial translators, she had to spend six hours on separate occasions to complete the translation. As for the back translation, one of the translators left many items partly completed as he could not think of the English equivalents to some of the Malay words used. Some extra form of quality control in the selection of translators has to be put in place to avoid similar circumstances in future research. In this case, translators were chosen based on recommendations by the translators' department head as well as based on academic qualifications as shown in Table 2. A much more effective approach is needed to choose better translators. The shortcomings of the translation stages were made up for by the use of an expert review panel which was not part of McGorry's (2000) recommendation. The six panel members, which included the researcher, came up with a pre-final version of the Malay TPI after deliberating for two hours on two separate occasions. The usefulness of an expert panel review after the back translation process cannot be over-emphasized here and many researchers have

reported using it (Daouk-Oyry & McDowal, 2012; Hyrkas, Appelqvist-Schmidlechner, & Oksa, 2003).

It was also discovered that even after the review by the expert panel, there were challenges with the translation process because a number of items were found to be unacceptable translations. The researcher decided to leave the items unchanged due to difficulties in translating certain items as reported by all translators as well as expert panel members. The difficulties in finding truly equivalent forms of the original items highlighted the need to use one of the recommended steps made by McGorry (2000), which was not used in this study, called decentering. Decentering involves making changes to both the original and the translated version. The original instrument can only be finalized once the translation process is completed (Werner & Campbell as cited in McGorry, 2000). Brislin, Lonner and Thorndike (1970) have pointed out that with decentering a one-to-one correspondence between the original and the translated form is not a necessity as long as equivalence in meaning can be maintained. However, as Chapman and Carter (1979) have pointed out, decentering is only possible if both versions of the instrument can be revised, and this was the reason why this step was omitted in this study. The original TPI version was not open for revision. One of the advantages of decentering, on the other hand, merits its consideration especially in cases where there is no exact equivalent form available in the target language. This was mentioned by translators and expert panel members for items b6 and a35. Panel members objected to the translators' choice of words in translating the phrase "virtuoso performer" of item b6, which is translated in Malay as "pengamal yang luarbiasa", meaning "extraordinary practitioners", which to the panel members is not exactly the meaning being portrayed by the original version. The same goes for

the expression “higher ideals” of item a35, which was translated as “kesempurnaan yang lebih tinggi”, which according to the panel failed to capture the essence of the original meaning of the English version. In the back translation process, the words used were “higher perfection” instead.

Albeit the issues mentioned above, the translated Malay instrument was produced from the expert panel review and this was used to carry out a pilot study to obtain preliminary basic statistical data as well as the test-retest reliability of the scores of the instrument. Pilot testing was one of the recommendations made by McGorry (2000). The pilot study had to be carried out with two separate groups because nine of the 25 original respondents did not respond after quite some time. This delay could have affected the outcome of the pilot study and so a new group of nine respondents was identified and test-retest analyses were conducted. Data analyses revealed nine items (b1, b6, b7, i17, b11, i19, i26, i29, and a42) to be inconsistent. One way to overcome the low response rate that was evident in the current pilot study is to offer some form of reward or incentive. With a small number of participants involved in a pilot study, this approach would not be too prohibitive from a cost standpoint.

The pilot study proved to be a beneficial move after the translation process had been carried out as this paved the way for cognitive interviews to be carried out. The use of cognitive interviews as part of the tools to develop cross-cultural research instruments has been reported in many studies (Enache, Gonzalez, Castillo, & Gonzalez, 2012; Fujishiro et al., 2010; Goerman & Caspar, 2010; Willis et al., 2008). One study by Daouk-Oyry and McDowal (2012) employed cognitive interviewing as a way to enhance the semantic equivalence of English personality inventories that were translated into three languages: Arabic, Mandarin, and Spanish. The authors reported that 67 out of 136 items were amended based on the findings that were gleaned

from the cognitive interviews and recommended a second cognitive interview may be required after field testing is done.

In this study, only one round of cognitive interviews was carried out. If Daouk-Oyry and Macdowal's suggestion to go for another round of cognitive interviews was to be implemented in this study, it would have to be after the measurement invariance testing had been carried out. As shown in the stage two findings of the current study, many items were found to be not invariant and these results could be used as a basis for selection of problematic items. For the current study, the cognitive interviews revealed that besides the nine items identified by the test-retest analysis, there were other items that had differences in responses. As for the nine items, the cognitive interviewees provided three main reasons for the differences. The first of these reasons was the presence of confusing words such as items b6 and b11 with words like "virtuoso performers" and "expert", respectively. Another reason brought up was the ambiguity of the situation or frame of reference that the item was referring to. According to one of the respondents, items i17 and i19 were dependent upon the type of students or subject to be taught. The last reason for the differences was about the scope of the domain being measured. Many items like i26, i29, and a42 refer to 'people' as the object of the sentence. The interviewees expressed confusion as to whom the sentences with 'people' in them were directed at. They reported that if they thought the people were 'students', their responses were to disagree but when they thought that the sentences referred to 'people in general', they chose to agree instead. Based on the feedback from the cognitive interviews, for item a42, the word 'novis' a direct translation of the word 'novice' was added in brackets after the expression "mereka yang lebih berpengalaman" (those with more experience) to clarify the meaning. Since the expression

"virtuoso performer" was not found in the Malay corpus, the current translation was deemed the most equivalent and no change was made. To deal with the other sources of confusion mentioned above, additional reminders were added in the instruction in the translated version as advised by Pratt and Collins (personal communication, August 10, 2013) to stress that the context and people referred to by the items must specifically be for the most recent course that the faculty have taught. The latest TPI website was also updated to carry additional instructions regarding the specificity of the focus of the items (Pratt & Collins, 2013).

Another method proposed by Pratt in his recommendation to translate the TPI involves using bilingual versions of the TPI (see Appendix B). This is done after back translations have been carried out, and following a pilot study and discussion among native speakers of the target language to obtain a pre-final version. The pre-final version is then administered to two groups of bilingual speakers of the same size. One of the bilingual groups responds to the translated version while the other group responds to the original version. Equivalence of the measures is then ascertained by comparing the means of the responses for each of the 45 items between the original and the translated versions as well as the means of composite scores obtained for the 15 subscale levels, the five perspectives levels, and the total score. The present study did not use this recommended step as it was deemed not practical to obtain a sample of 20-30 bilingual teachers to conduct these analyses, and it was reasoned that the expert panel, pilot study, and cognitive interviews would provide the necessary information about the conceptual similarity and differences in the English and Malaysian translated forms. Besides, according to Sperber, Devellis, and Boehlecke (1994), the assumption that a bilingual person's response patterns to translated items can be generalized to a monolingual person's responses may be false.

Measurement Properties of the Teaching Perspectives Inventory

Standard 9 from the *Standards for Psychological and Educational Measurement* (AERA, APA, & NCME, 1999) states that instrument developers must provide evidence for measurement invariance across language variations and must take the initiative to explore the possibilities that the instrument may not function equivalently across different groups. Extending this idea to an instrument like the TPI, which has been translated into eight languages and has been used in a number of cross-cultural research studies (Lu, 2006; Ruan, 2004), the measurement properties of the original TPI have to conform to the required measurement standards and when translated and adapted, the new version must conform to these same standards.

Looking at the basic psychometric properties of the original, English-version of the TPI, the reliabilities of the five subscales were mostly moderate. Pratt and Collins (2010) reported test-retest correlations that ranged from .48 (Apprenticeship) to .81 (Nurturing) and internal consistency reliabilities that ranged from .70 to .83. The results from this study showed that the reliabilities of the English version of the TPI with the data obtained from the TPI's database (n = 605) were similar to those reported by Pratt and Collins with the lowest internal consistency reliability equal to .67 for Transmission and Developmental and the highest for Nurturing and Social Reform (α=83). These coefficients, while acceptable, were not as high as expected for a widely used instrument like the TPI. A recent study by Wang (2012) in China revealed lower reliabilities as well. This may partially explain why the factor structure was less than ideal.

Although the TPI has been used by many researchers (Chan, 1994; Deggs, Machtmes, & Johnson, 2008; Wang, 2012) few confirmatory factor analyses (CFA) have been carried out to assess the adequacy of the fit of the five-factor model. Confirmatory factor analysis is a critical

source of validity evidence as it can show how well the data fit the model, and unlike Cronbach's alpha, CFA can reveal additional sources of model misfit, such as items that have secondary loadings on other factors and correlated item errors. Results of the CFA conducted in the present study on the correlated five-factor structure of the English version of the TPI found that the fit of the model was marginal. In addition, an estimation problem resulting in a lack of convergence was evident and was only resolved when item b3 from the Developmental subscale was removed from the model. These results are consistent with the problems with model fit reported by Brown and Lake (2006) with the English-version of the TPI, which led them to reduce the TPI to 11 items and four factors in order to achieve acceptable model fit. In this study, both the original English TPI and the translated Malay versions were found to have less than acceptable fit with the Malay version displaying much worse fit. This may be partly explained by the low average correlations between the 45 TPI items, which tend to have an impact on some of the fit measures, such as the comparative fit index (Kenny & McCoach, 2003).

Invariance testing is normally conducted on the full measurement model (i.e., five-factor TPI). Because the fit of the factor structure for the five-factor model was inadequate it was necessary to evaluate the structure one factor at a time. Subscale CFAs for the English TPI revealed that the Developmental subscale, which had one less item after the exclusion of b3, had the best model fit and required no further modification. The remaining four subscales were modified by adding one or more parameters into the model in the form of highly correlated item errors. For the Malaysian TPI, all the subscales showed less than acceptable fit and required three or more additional parameters to improve model fit. The Malaysian Transmission subscale revealed low standardized loadings for items b1 (.07) and b6 (.11), which support the findings of

the pilot study that indicated problems in the translation process involving complexities of the words.

Even after the individual models were made to fit adequately within each country, when invariance testing was carried out across countries, the results did not support the existence of measurement invariance for the English and the Malay versions. Measurement invariance, according to Meredith (1997) is present when the indicators or items used to measure the latent trait are the same across groups, which requires the equality of item factor loadings, item intercepts, and item residual variances. Both metric invariance (equality of factor loadings) and scalar invariance (equality of intercepts) were found to be not tenable for the subscales as well as for many of the items in each of the subscales. None of the five subscales was successful in showing metric and scalar invariance. Because there was no evidence of metric and scalar invariance, there was no further need to conduct stronger invariance testing to assess for equality of item residual variances as the condition for weak invariance was not fulfilled. Overall, despite the rigorous process of translation and adaptation used in the present study the results of confirmatory factor analyses do not support the measurement equivalence of the TPI adapted into Bahasa Malaysia and that of the original English TPI. The limitations of the present study along with directions for future research will provide suggested next steps for enhancing the translation and adaptation process with the goal of moving closer to achieving an equivalent Malaysian version of the TPI.

Limitations

There are three main limitations to this study. The first limitation is the lack of a screening process to ensure that all translators had adequate translation ability and proficiency

required of the task. Relying on just the recommendation of the translators' superior and lists of academic and professional qualifications were not sufficient to identify qualified translators, and two out of the five translators self-reported that they had difficulties in translating many of the items.

The second limitation of the study, related to the first, was that it was necessary to use translators who came forward as volunteers to carry out the forward- and back-translations. The translators were mostly language experts and linguist who had some background in doing translations from English into Malay. Even though all the recruited translators and reviewers had taught a minimum of six years in university, they were not experts in teaching at the higher education level. If funds were available it would be possible to attract highly qualified and certified translators. Increasing the number of forward- and back-translators plus having a larger expert panel would also have strengthened the translation and adaptation process.

Lastly, the low rate of return in stage two of the psychometric analysis of the TPI was a major drawback in this study. The final sample does not accurately reflect the distribution pattern of all the faculties in the 20 government-funded universities in the country. Although the sample size for the Malaysian group was large (n = 561), it still represented only a small percentage (about 2%) of the population of faculty in the 20 government-funded universities in the country which, according to the most recent government statistics, was 24, 571 strong (Ministry of Higher Education, 2010). Also, it was not possible to randomly select faculty to be part of the study and so the generalizability of the results is limited. Furthermore, only a single ethnic group was chosen for the Malaysian sample (i.e., the Malays) to compare with the U. S. sample, which consists of more than one ethnic group. Another limitation related to the psychometric stage of

the study was that matching of the samples was only done based the samples' profiles on selected variables related to the faculty members' teaching experience (e.g., years teaching, types of students taught). Therefore, any generalizations from the findings of this study can only be made to people who are similar to the participants of this study.

Suggestions for Future Research

A similar study such as the present study is recommended, taking into consideration the limitations that have pointed out. As for the translation process, selection of a greater number of qualified translators and reviewers would be beneficial. With adequate funding, a well-trained team of translators and reviewers could be hired to enhance the translation process for the TPI. These experts would need to be familiar with the complexities of teaching in higher education and be well-versed with the philosophy that undergirds the TPI. It would also be advantageous to have some form of a screening process, which could be a translation proficiency test such as those used by many translator and interpreter agencies. The Interpreter Language and Interpreting Skills Assessment Tool (ILSAT) (Center for Education and Training, n.d.) and the ATA Certification Exam (American Translators Association, 2013) are good models to be used as a screening test to check the overall ability of a translator.

Another method worth considering in the translation and adaptation process is to get permission from the TPI developers to allow the original instrument be opened for modification so that the decentering approach as recommended by McGorry (2000) could be carried out. Items like b6 with the idiomatic expression "virtuoso performer" is a unique English expression which has no equivalent form in the Malay language and even translators of high caliber with the desired qualifications may not be able to satisfactorily translate the items to the original intent of

the English version. The decentering process could be of assistance to find a common ground bearing in mind the claims of the Sapir-Whorf Hypothesis that it is language that determines the way that individuals organize their thoughts about the world and their experiences in it and because they vary to a certain extent from culture to culture (Whorf, 1956). Similar experiences may be perceived and articulated differently by different languages. Some form of compromise must be achieved to come to terms with these cross-cultural differences that exist between the English version and the target language during the decentering process.

For the psychometric phase of the research, a larger and more representative sample of faculty is recommended to determine if the results of the present study are replicated with this new sample. For the low rate of return shown by the Malaysian sample, a more persistent approach with multiple follow-up requests for participation in the research is needed to obtain more respondents for both the pilot study as well as the main data collection stage for the Malaysian sample. Some form of reward or other incentives may be required to boost participation.

A closely-matched pair of samples based on a single ethnic group from each sample would be advantageous. However, each sample should accurately represent the population they are supposed to come from. It would be misguided if the U. S. samples were obtained to look exactly like the Malaysian sample because some demographic variables like teaching experience of faculty have different distribution patterns. There are more faculty members in the U. S. who have longer years practicing than those in Malaysia and there are more faculty members teaching postgraduate students as compared to their Malaysian counterparts.

Future validation studies of the TPI would gain considerably by providing space on the electronic instrument forms for feedback from respondents about the items. This can be in the form of closed or open-ended questions to ascertain how the items are functioning. One issue of interest that can be investigated is the social desirability in the responses among the Malaysian participants. Collins (email to author, November 11, 2011) stated that Malaysian respondents from the south of Peninsular Malaysia were found to have endorsed high positive agreements to almost all of the items for all the perspectives when taking the original English TPI from the TPI website. This is deemed to be implausible because some of the perspectives present views of teaching that are in contrast to the others. The explanation given by Collins was that the Malaysian respondents seemed to think that the TPI is like a test and they wanted to score high on every item on the scales. Future Malay TPI survey would benefit from an incorporation of items specifically measuring social desirability items as proposed by Crowne and Marlowe (1960) to ascertain to what extent socially desirable responding (SDR) is present among Malaysian respondents when taking the survey. Social desirable responding as defined by van de Mortel (2008) is the tendency of individuals to respond to items in a manner that would make them appear favorable and in the meta-analytic study that was carried out by van de Mortel revealed that almost half of the 14275 studies identified were found to have been influenced by SDR. From the description provided by Collins above, Malaysians responding to the TPI items may manifest a certain amount of this bias. If this is the case, then high positive agreements to the TPI items should correlate highly with the SDR scale.

As for the psychometric properties of the TPI, the CFA results obtained from the original version of the TPI suggest a lack of adequate fit. Future research should continue to evaluate,

using CFA, the five-factor model of the TPI and particular items such as b3 (Most of all, learning depends on what one already knows) and b6 (Teachers should be virtuoso performers of their subject matter) of the Developmental scale. Item b3 was found to be problematic in the analyses of the original TPI. This item requires close attention in future research that endeavors to employ the TPI to study teacher beliefs. The item did not pose any difficulties for the translation process and it did not appear to be a problem in the translated Bahasa Malaysia version. However, the item caused a statistical problem in the original TPI model specification for this particular study and it was subsequently dropped from the final model to be used in the invariance testing. Future research needs to consider that an item may be acceptable from the viewpoint of the translators and expert panel but can still pose statistical problems in subsequent analyses. Likewise, an item may be viewed as problematic by the translators and expert panel but can still not pose statistical complications. The large number of correlated errors for items as revealed by the CFA of both the original and the translated TPI also suggests the need for some items to be reviewed in terms of wording and content. Items b2 and b7 showed high correlated errors in the Apprenticeship scale for the English version and both items contain "good practitioners" in them and they are both under the same subscale. The same problem was found for items a35 and a40 of the Social Reform scale. It would be advisable to reword the items to avoid these similarities. Items that are able to discriminate between dissimilar constructs would help improve the fit of the TPI and this will help in making translations much easier for future undertakings.

For assessing measurement equivalence, at the time of this writing, there is yet to be universal agreement on how to do invariance testing. Some experts argue that when metric invariance is not obtained, the invariance testing process should be terminated. In other words,

when the fit gets significantly worse as compared to configural invariance, the two models are not equal and therefore no further testing is required. In this study, however, the decision was made to explore specifically where the source of misfit was even when the two versions of the TPI were not found to be equal in the metric invariance. The purpose was to gain more insight on which items were functioning better than others similar to doing item analyses based on other less complex procedures such as factor analyses. Lastly, using multiple psychometric methods such as CFA and item response theory (IRT) is recommended to effectively test for measurement invariance in cross-cultural research. This is to shed light on issues such as the inconsistencies in the results of the translation and adaptation process with the psychometric analyses such as the case of item b6. The item was revealed to be problematic during the translation and adaptation process due to the fact that there was no equivalent form in the Malay language. However, the outcome of the invariance testing showed that it was one of the three invariant items of the Transmission perspective. Further investigation is needed to explain the discrepancy. As noted by Van de Vijver (2003), "statistical sophistication in data analysis cannot compensate for poor quality of study design nor for lack of cultural sophistication…Only through a combination of cultural awareness and statistical sophistication can we arrive at high quality survey research" (p. 233).

References

Acuna, E., & Rodriguez, C. (2004). The treatment of missing values and its effect in the classifier accuracy. In D. Banks, L. House, F. R. McMorris, P. Arabie, & W. Gaul (Eds.), *Classification, clustering and data mining applications.* (pp. 639-648). Heidelberg, Germany: Springer-Verlag.

Acquadro, C., Conway, K., Hareendran, A., & Aaronson, N. (2008). Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value in Health, 11*(3), 509-521.

Adler, N. J. (2001). *International dimensions of organizational behavior* (4th ed.). Cincinnati, OH: South-Western College Publishing.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D. C.: American Educational Research Association.

American Translators Association. (2013). *ATA certification Program*. Retrieved from http://www.atanet.org/certification/aboutpractice_test.php

Ariely, G., & Davidov, E., (2009, July). *Measurement equivalence in cross-national surveys: The neglected invariance analyses*. Paper presented at the annual meeting of the ISPP 32nd Annual Scientific Meeting, Trinity College, Dublin, Ireland.

Bar, R. B., & Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change: The Magazine of Higher Learning, 17*(6), 12-26.

Beeman, R. R. (1998). Teaching and Learning in the 21st Century. *Almanac, 44 (31).* Retrieved from http://www.upenn.edu/almanac/v44/n31/tatbeeman.html

Biggs, J. (2003). *Teaching for quality learning at university* (2nd ed.). Maidenhead, UK: Society for Research into Higher Education/Open University.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university*. (4th ed.). Maidenhead, UK: McGraw-Hill International.

Bonwell, C., & Eison, J. (1991). *Active learning: Creating excitement in the classroom AEHE-ERIC Higher Education Report No. 1*. Washington, D.C.: Jossey-Bass.

Borich, G. D. (1999). Dimensions of self that influence effective teaching. In R. P. Lipka & T. M. Brinthaupt (Eds.), *The role of self in teacher development*, (pp. 92-117). Albany, NY: State University of New York Press.

Bowden, J., & Marton, F. (1998). *The university of learning*. London, UK: Kogan Page.

Brancato, V. C. (2003). Professional development in higher education. *New Directions for Adult and Continuing Education, 2003*(98), 59–66.

Brislin, R. (1970) Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216.

Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research method*. New York, NY: Wiley.

Brown, G. T. L., & Lake, R. (2006, November). *Queensland teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers.* Paper presented at the Annual Conference of the Australian Association for Research in Education (AARE), Adelaide, Australia.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bruner, J. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

Budgell, G. L., Raju, N. S., & Quartetti, D. A. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement, 19(4*), 309-321.

Burns, D. J., & Brady, J. (1992). A cross-cultural comparison of the need for uniqueness in Malaysia and the United States. *The Journal of Social Psychology, 132*(4), 487-495.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.

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*, 456-466.

Caffarella, R. S., & Zinn, L. F. (1999). Professional development for faculty: A conceptual framework of barriers and supports. *Innovative Higher Education, 23*(4), 241-254.

Campbell, H. L., Barry, C. L., Jilliam N. J., & Finney, S. J. (2008). Configural, metric, and scalar invariance of the Modified Achievement Goal Questionnaire across African American and White university students. *Educational and Psychological Measurement, 68*(6), 988-1007.

Cha, E., Kim, K. K., & Erlen, J. A. (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing, 58*(4), 386-395.

Chan, C. H. (1994). *Operationalization and Prediction of Conceptions of Teaching in Adult Education*. Unpublished doctoral dissertation, The University of British Columbia, Vancouver, British Columbia.

Chapman, D. W., & Carter, J. F. (1979). Translation procedures for the cross cultural use of measurement instruments. *Educational Evaluation and Policy Analysis, 1*(3), 71-76.

Center for Education and Training. (n. d.). Translation & Interpreter services: Interpreter language proficiency testing. Retrieved from http://www.tcet.com/interpreter/proficiency-test.php

Cheung, G. W., & Rensvold, R. B. (1998). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science, 6,* 93-110.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1-27.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Clark, C. M., & Yinger, R. J. (1979). Teachers' thinking. In P. L. Peterson, & H. J. Walberg, (Eds.), *Research on teaching: Concepts, findings, and implications*. (pp. 231-263). Berkeley, CA: McCutchan.

Clark, C. M. (1986). Ten years of conceptual development in research on teacher thinking. In M. Ben-Peretz, R, Bromine, & Halkes, R (Eds.). *Advances of research on teacher thinking.* Lisse: Swets & Zeitlinger.

Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research, 20*, 389-417.

Collins, J. B., & Pratt, D. D. (2011).The teaching perspectives inventory at 10 years and 100,000 respondents: Reliability and validity of a teacher self-report inventory. *Adult Education Quarterly 61*(4), 358-375.

Collins, J. B., Selinger, S. J., & Pratt. D. D. (2001). *How do perspectives on teaching vary across disciplinary majors for students enrolled in teacher preparation?* Retrieved from http://www.teachingperspectives.com/PDF/howdoteachers.pdf

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Cronbach, L. J., & Meehl, P. C. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Cross, P. K. (1998). *What do we know about students' learning and how do we know it.* AAHE's National Conference on Higher Education. Retrieved from http://cshe.berkeley.edu/publications/docs/ROP.Cross.7.05.pdf

Crowne, D. P., & Marlowe, D. (1960) A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349-354.

Daouk-Oyry, L., & McDowal, A. (2012). Using cognitive interviewing for the semantic enhancement of multilingual versions of personality questionnaires. *Journal of Personality Assessment, 95*(4), 407-416.

Deggs, D. M., Machtmes, K. L., & Johnson, E. (2008). The significance of teaching perspectives among academic disciplines. *College Teaching Methods & Styles Journal, 4*(8), 1-7.

Dolmans, D., & Gijbels, D. (2013). Research on problem-based learning: Future challenges. *Medical Education*, *47*(2), 214-218.

Douglas, S. P., & Nijssen, E. J. (2003). On the use of "borrowed" scales in cross-national research: A cautionary note. *International Marketing Review, 20,* 621-642.

Doyle, M. (1997). Beyond life history as a student: Preservice teachers' beliefs about teaching and learning. *College Student Journal, 31*, 519-531.

Enache, C. M., González, I. G., Merino, D. C., & González, O. L. (2012). Protean and boundaryless career attitudes scales: Spanish translation and validation. *Intangible Capital, 8*(1), 1-16.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*(5), 713-740.

Eison, J. (2002). Teaching strategies for the twenty-first century. In R. Diamond (Ed.), *Field guide to academic leadership*. (pp. 157-174). San Francisco, CA: Jossey-Bass.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38(1)*, 47-65.

Fenstermacher, G. D. (1986). Philosophy of research on teaching: Three aspects. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.

Fink, L. D. (2003). *Creating significant learning experiences*. San Francisco, CA: Jossey Bass.

Flaherty, J. A., Gavira, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman J. A., … Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease, 176,* 257-263.

Fontaine, R., & Richardson, S. (2003). Cross-cultural research in Malaysia. *Cross-Cultural Management: An International Journal, 10*(2), 75-89.

Fujishiro, K., Gong, F., Baron, S., Jacobson, C. J., Jr., DeLaney, S., Flynn, M., & Eggerth, D. E. (2010). Translating questionnaire items for a multilingual worker population: The iterative process of translation and cognitive interviews with English-, Spanish-, and Chinese-speaking workers. *American Journal of Industrial Medicine*, *53*(2), 194-203.

Geisinger, K. F. (1994). Cross-Cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*(4)*,* 304-312.

Goerman, P. L., & Caspar, R. A. (2010). A preferred approach for the cognitive testing of translated materials: Testing the source version as a basis for comparison. *International Journal of Social Research Methodology*, *13*(4), 303–316.

Gow, L., & Kember, D. (1993). Conceptions of teaching and their relationship to student learning. *British Journal of Educational Psychology, 63*(1), 20-33.

Grant, J. S., & Davis, L. T. (1997). Selection and use of content experts in instrument development. *Research in Nursing & Health, 20*(3), 269-274.

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*(12), 1417-1432.

Gulliksen, H., & Wilks, S. S. (1950). Regression tests for several samples. *Psychometrika*, *15*(2), 91-114.

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*(3), 164-172.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications.

Hambleton, R. K., & Rodgers, J. (1995). Item bias review. *Practical Assessment, Research & Evaluation, 4(6).* Retrieved from http://PAREonline.net/getvn.asp?v=4&n=6

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Applied Testing Technology, 1*(1), 1-16.

Hargreaves, A., & Fullan, M. G. (1992). *Understanding teacher development*. London, UK: Cassell.

Harkness, J. A, Pennell, B. E., & Schoua-Glusberg, A. (2004). Survey Questionnaire Translation and Assessment. In S. Presser, J. M. Rothgeb, M. P. Couper, J. L. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*. (pp. 453-473). Hoboken, NJ: Wiley.

Harkness, J. A. (2007). Improving the comparability of translations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.). *Measuring attitudes cross-nationally: Lessons from the European social survey.* London: Sage Publications.

Hartley, J. (2008). *Learning and studying: A research perspective*. London, UK: Routledge.

Hayduk, L. A. (1989*). Structural equation modeling: Essentials and advances.* Baltimore, MD: The John Hopkins University Press.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82*(6), 903–918.

Horn, J. L. (1991). Comments on issues in factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change*. (pp. 114-125). Washington D.C.: American Psychological Association.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117-144.

Howell, D. C. (2012) *Treatment of missing data: Part I.* Retrieved from http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html

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

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53-69.

Hyrkas, K., Appelqvist-Schmidlechnerb, K., & Oksac, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies, 40*(6), 619-625.

IBM Corp. Released 2012. *IBM SPSS Statistics for Windows, Version 21.0*. Armonk, NY: IBM Corp.

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *Zuma Nachrichten Spezial, 3*, 1-40.

Johnstone, A. H., & Percival, F. (1976). Attention breaks in lectures. *Education in Chemistry, 13*, 49-50.

Kagan, D. M. *(*1990*)*. Ways of evaluating teacher cognition: Inferences concerning the Goldilocks principle. *Review of Educational Research, 60*(3), 419-469.

Kagan, D. M. (1992). Implication of research on teacher belief. *Educational Psychologist. 27* (10), 65-70.

Ke, F., & Chávez, A. F. (2013).*Web-based teaching and learning across culture and age.* New York, NY: Springer

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 10*(3), 333-351.

Kember, D., & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *Journal of Higher Education, 65*(1), 58-74.

Kember, D. (1997). A reconceptualisation of research into university academics/conceptions of teaching. *Learning and Instruction. 7*(3), 255-275.

Kember, D. (2000). Misconceptions about the learning approaches, motivation and study practices of Asian students. *Higher Education, 40*(1), 99-121.

Kember, D. (2009). Promoting student-centered forms of learning across an entire university. *Higher Education, 58*(1), 1-13.

Khader, F. R. (2012). The Malaysian experience in developing national identity, multicultural tolerance and understanding through teaching curricula: Lessons learned and possible applications in the Jordanian context. International Journal of Humanities and Social Science, 2(1), 270-288.

Kember, D., & Kwan, K. (2000). Lecturers' approaches to teaching and their relationship to conceptions of good teaching. *Instructional Science, 28*(5), 469-490.

Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies,* (2nd ed.). London, UK: RoutledgeFalmer.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53-76.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Lu, X. (2006). *Conceptual equivalence of teaching perspectives in Chinese and English* (Unpublished master's thesis). University of British Columbia, Vancouver, British Columbia, Canada.

Lueddeke, G. R. (2003). Professionalising teaching practice in higher education: A study of disciplinary variation and 'teaching-scholarship'. *Studies in Higher Education, 28* (2). Retrieved from http://pdfserve.informaworld.com/224337_731200553_713696325.pdf

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35*(2), 382-386.

Ljubojevic, D., & Laurillard, D. (2010). *A theoretical approach to distillation of pedagogical patterns from practice to enable transfer and reuse of good teaching.* Proceedings of the 2010 European LAMS & Learning Design Conference http://lams2010.lamsfoundation.org/papers.htm

Marsh, H. W. (1994) Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling 1*(1), 5-34.

McCredie, J. (2003). Does IT matter to higher education. *Educause Review*, *38*(6), 15-22.

McGorry, S. Y. (2000) Measurement in a cross-cultural environment: Survey translation issues. *Qualitative Market Research: An International Journal, 3*(2*),* 74-81.

McKeachie, W. J., & Svinicki, M. (2005). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*. (12th Ed.). New York, NY: Houghton Mifflin Company.

Menges, R. J. (1990). Teaching: Beliefs and behaviors. *Teaching Excellence, 2*(6*),* 1-2.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*(4), 525-543.

Merrill, M. D. (2013). *First principles of instruction: Identifying and designing effective, efficient, and engaging instruction*. San Francisco, CA: John Wiley and Sons, Inc.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education, 30*(4), 159-167.

Ministry of Higher Education (2010). *Chapter 2: Public higher education institution (HEI).* Retrieved from http://www.mohe.gov.my/web_statistik/statistik2010/BAB2_IPTA.pdf

Mohammad Aziz Shah, M. A., Fauziah, M. S., Abdul Malek, A. R., Syed Sofian, S. S., & Mohammad Nasir, B. (2013). Translation, validity and reliability of multidimensional self-concept scale (MSCS) questionnaire among Malaysian teenagers. *Procedia Social and Behavioral Sciences, 84,* 1455-1463.

Musa, R., Fadzil, M. A., & Zain, Z. (2007). Translation, validation and psychometric properties of Bahasa Malaysia version of the depression anxiety and stress scales (DASS). *ASEAN Journal of Psychiatry, 8*(2), 82-89.

Middendorf, J., & Kalish, A, (1996) The 'change up' in lectures. *National Teaching & Learning Forum, 5*(2), 1-5.

Muthén, B. O., Jo, B., & Brown, H. (2003). Comment on the Barnard, Frangakis, Hill& Rubin article, Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association, 98*, 311-314.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide.* (6th Ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide.* (7th Ed.). Los Angeles, CA: Muthén & Muthén

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies, 19*(4), 311-328.

Nik Nazli, N. A., & Maliah, S. (2013). Case-studies in a passive learning environment: Some Malaysian evidence. *Accounting Research Journal*, *26*(3), 2-2.

Northcote, M. (2003). Online assessment in higher education: The influence of pedagogy on the construction of students' epistemologies. *Issues In Educational Research, 13* (1), 66-84.

Nur Fazidah, E. (2012). Using a deductive approach in validating a measurement model. *Australian Journal of Basic and Applied Sciences, 6*(8), 329-336.

Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332.

Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly, 42*(4), 203-220.

Pratt, D. D. (1998). The research lens: A general model of teaching. In D. D. Pratt & Associates (Eds.). *Five perspectives on teaching in adult and higher education,* (pp. 3 -14). Malabar, FL: Kreiger Publishing Company.

Pratt, D. D., & Collins, J. B. (2001). *Teaching perspectives inventory*. Retrieved from http://www.teachingperspectives.com

Pratt, D. D., Collins, J. B., & Selinger, S. J. (2001). *Development and use of the Teaching Perspectives Inventory (TPI).* Paper presented at the 2001 American Educational Research Association Annual Meeting, Seattle, WA. Retrieved October 20, 2009, from http://www.teachingperspectives.com/PDF/development1.pdf

Pratt, D. D., & Collins, J. B. (2013). *Teaching perspectives inventory*. Retrieved from http://www.teachingperspectives.com/drupal

Pratt, D. D. (2002). Good teaching: One size fits all? In J. Ross-Gordon (Ed.). *An up-date on teaching theory*. San Francisco, CA: Jossey-Bass Publishers.

Pratt, D. D. (2005). Personal philosophies of teaching: A false promise? *ACADEME, 91*(1), 32-35.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education. 93*(3), 223-231.

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education.* Buckingham, UK: SRHE and Open University Press.

Quek, K. F., Chua, C. B., Low, W. Y., Razack, A. H., & Loh, C. S. (2002). Reliability and validity of the Malay version of the Health-Related Quality of Life instrument in a Malaysian population. *BJU International, 90*(1), 37-40.

Rahimah, A. (1998). Educational development and reformation in Malaysia: Past, present and future. *Journal of Educational Administration, 36*(5), 462-475.

Ratcliff, S., & Rocco, T. (2003). *The teaching perspectives inventory and teaching log as tools to inform faculty development.* Unpublished manuscript, Florida International University.

Rensvold, R. B., & Cheung, G. W. (1998). Testing measurement models for factorial invariance: A systematic approach. *Educational and Psychological Measurement, 58*(6), 1017-1034.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analyses and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552-566.

Reynolds, J. (2000). Learning-centered learning: Theory into practice. *Inquiry, 5*(2), 1-9.

Riordan, C. R., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management*, *20*(3), 643-671.

Robinson, V. M. J. (1993). *Problem-based methodology: Research for the improvement of practice.* Oxford, UK: Pergamon Press.

Rokeach, M. (1968). *Beliefs, attitudes and values: A theory of organization and change*. San Francisco: Jossey-Bass.

Ruan, X. (2004). *Conceptual equivalency in translating teaching perspectives inventory (TPI): A Chinese perspective* (Unpublished master's thesis). University of British Columbia, Vancouver, British Columbia, Canada.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-92.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work. *Social Work Research, 27*(2), 94–104.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*(2), 243-248.

Schaubroeck, J., & Green, S. G. (1989). Confirmatory factor analytic procedures for assessing change during organizational entry. *Journal of Applied Psychology*, *74*(6), 892-900.

Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, *17*(3), 343-358.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London, UK: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*(1), 1-10.

Schriersheim, C., & Neider, L. L. (Eds.). (2001). *Equivalence in measurement.* Greenwich, CT: Information Age Publishing.

Seldin, P., & Associates. (1995). *Improving college teaching.* Bolton, MA: Anker.

Sharom, A. (1980). Nation building and the university in developing countries: The Case of Malaysia. *Higher Education, 9*(6), 721-741.

Shavelson, R. J., & Stern, P. (1981). Research of teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*(4), 455-498.

Sireci, S. G., Patsula, L., & Hambleton, R. K. (2005). In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), Adapting educational and psychological tests for cross-cultural assessment, (pp. 93-116). Mahwah, NJ.: Lawrence Erlbaum Associates, Inc., Publishers.

Smith, R., & Geis, G. (1996), Professors as clients for instructional development. In L. Richlin (Ed.), *To Improve the Academy, 15*. Stillwater, OK: New Forums Press.

Solarno-Flores, G., Backhoff, E., & Contreras-Nino, L. A. (2009). Theory of test translation error. *International Journal of Testing, 9*(2), 78-91.

Sperber, A. D., Devellis, R. F., & Boehlecke, B. (1994). Cross-cultural translation: Methodology and validation. *Journal of Cross-cultural Psychology, 25*(4), 501-524.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-107.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2007). *Testing measurement invariance using mulitigroup CFA: Differences between educational groups in human values measurement*. Retrieved from http://www.springerlink.com/content/41p1302164175954/fulltext.pdf

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Suan-Phaik Khoo, B. D. S., Yap, M. A. U. J., Head, B. U., Bulgiba, A. M., & Khoo, S. P. (2008). Translating the research diagnostic criteria for temporomandibular disorders into Malay: Evaluation of content and process. *Journal of Orofacial Pain, 22*(2), 131-138.

Swaine-Verdier, A., Doward, L. C., Hagell, P., Thorsen, H., & McKenna, S. P. (2004). Adapting life quality instruments. *Value in Health, 7*(1), 27-30.

Swami, V. (2012). Further examination of the psychometric properties of a Malay version of the Rosenberg self-esteem scale. In S. De Wals, & K. Meszaros (Eds.), Handbook on psychology of self-esteem (pp. 371-380). Hauppauge, NY: Nova Science Publishers.

Svinicki, M. (1998). Divining the future for faculty development: Five hopeful signs and one caveat. In M. Kaplan, (ed.). *To Improve the Academy, 17*. Stillwater, OK: New Forums Press.

Tan, P. L. (2005). *Approaches to learning and learning values: An investigation of adult learners in Malaysia.* (Unpublished doctoral dissertation). Queensland University of Technology, Brisbane, Australia.

Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. Grouws, (Ed.). *Handbook of research in mathematics teaching and learning*. (pp. 127-146). New York, NY: MacMillan.

Tight, M. P. (2003) *Researching higher education.* Maidenhead, UK: Open University Press.

Trigwell, K., & Prosser, M. (1996a). Changing approaches to teaching: A relational perspective. *Studies in Higher Education, 21*(3), 275-284.

Trigwell, K., & Prosser, M. (1996b). Congruence between intention and strategy in university science teachers' approaches to teaching. *Higher Education, 32*(1), 77-87.

Trigwell, K., & Prosser, M. (December 2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review, 16*(4), 409-424.

U. S. Department of State. (2010). Accessed at http://www.state.gov/r/pa/ei/bgn/2777.htm

van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*(2), 89-99.

van Widenfelt, B. M., Treffers, P. D. A., de Beurs, E., Siebelink, B. M. & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychological Review, 8*(2), 135-147.

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology, 78*(4), 557-568.

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*, *2*, 4-69.

Von Hippel, P. T. (2004). Biases in SPSS 12.0 missing value analysis. *The American Statistician, 58*(2), 160-164.

Wang, L. (2012). Teaching perspectives of pre-service physical education teachers in Shanghai: A pilot study. *Asian Social Science*, *8*(1), 209-216.

Watkins, D., & Ismail, M. (1994). Is the Asian learner a rote learner? A Malaysian perspective. *Contemporary Educational Psychology, 19*(4), 483-488.

Whorf, B. (1956). *Language, thought & reality*. Cambridge, MA: MIT Press.

Widaman, K. F., & Reise, S. P., (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance abuse domain. In: K. J. Bryant, & M. Windle, (Eds.). *The science of prevention: Methodological advance from alcohol and substance abuse research.* (pp. 281–324).Washington, D.C.: American Psychological Association.

Willis, G. B. (1999). *Cognitive interviewing: A "how to" guide.* Research Triangle Institute. 1999 Meeting of the American Statistical Association. Research Triangle Park, NC: Research Triangle Institute.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Willis, G. B., Kudela, M. S., Levin, K., Norberg, A., Stark, D. S., Forsyth, B. H., … Hartman, A. M. (2010). Evaluation of a multistep translation process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B. E. Pennell & T. W. Smith (Eds*.), Survey methods in multicultural, multinational, and multiregional contexts* (pp. 141-156). Hoboken, NJ: John Wiley & Sons.

Willis, G., Lawrence, D., Hartman, A., Kudela, M. S., Levin, K., & Forsyth, B. (2008). Translation of a tobacco survey into Spanish and Asian languages: The tobacco use supplement to the Current Population survey. *Nicotine & Tobacco Research*, *10*(6), 1075–1084.

Wood, E. J. (2004). Problem-based learning: Exploiting knowledge of how people learn to promote effective learning. *Bioscience Education, 3.* Retrieved from http://journals.heacademy.ac.uk/doi/pdf/10.3108/beej.2004.03000006

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation. 12*(3), 1-26.

Yoo, B. (2002). Cross-group comparisons: A cautionary note. *Psychology and Marketing, 19,* 357–368.

Zaidatul Akmaliah, L. P., Jamaliah, A. H., & Rahil, M. (2007). Quality teaching in learning business studies at instituions of higher education. *Pertanika Journal of Social Sciences & Humanities, 15*(2), 113-125.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (Ordinal) item scores.* Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Appendices

Appendix A: Teaching Perspectives Inventory Items – English Version Sorted by Perspective

Transmission

b1. Learning is enhanced by having predetermined objectives.
b6. Teachers should be virtuoso performers of their subject matter.
b11. Effective teachers must first be experts in their own subject.

i16. My goal is to prepare people for content-related examinations.
i21. I expect people will master a lot of information related to the subject.
i26. I want people to score well on examinations as a result of my teaching.

a31. I cover the required content accurately and in the allotted time.
a36. My teaching is governed by the course objectives.
a41. I make it1 very clear to people what they are to learn.

Apprenticeship

b2. To be a good teacher, one must be a good practitioner.
b7. The best learning comes from working alongside good practitioners.
b12. Knowledge and its application cannot be separated.

i17. My goal is to demonstrate how to perform or work in real situations.
i22. I expect people to know how to apply the subject matter in real settings.
i27. I want people to understand the realities of working in the real world.

a32. I link the subject matter with real settings of practice or application.
a37. I model the skills and methods of good practice.
a42. I see to it that novices learn from more experienced people.

Developmental

b3. Most of all, learning depends on what one already knows.
b8. Teaching should focus on developing qualitative changes in thinking.
b13. Teaching should build upon what people already know.

i18. My goal is to help people develop more complex ways of reasoning.

i23. I expect people to develop new ways of reasoning about the subject.
i28. I want people to see how complex and inter-related things really are.

a33. I ask a lot of questions while teaching.
a38. I challenge familiar ways of understanding the subject matter.
a43. I encourage people to challenge each other's thinking.

Nurturing

b4. It's important that I acknowledge learners' emotional reactions.
b9. In my teaching, building self-confidence in learners is a priority.
b14. People's effort should be rewarded as much as achievement.

i19. My goal is to build people's self-confidence and self-esteem as learners.
i24. I expect that people will enhance their self-esteem through my teaching.
i29. I want to provide a balance between caring and challenging as I teach.

a34. I find something to compliment in everyone's work or contribution.
a39. I encourage expressions of feeling and emotion.
a44. I share my own feelings and expect my learners to do the same.

Social Reform

b5. My teaching focuses on societal change, not the individual learner.
B10. Individual learning without social change is not enough.
b15. For me, teaching is a moral act as much as an intellectual activity.

i20. My goal is to challenge people to seriously reconsider their values.
i25. I expect people to be committed to changing our society.
i30. I want to make apparent what people take for granted about society.

a35. I use the subject matter as a way to teach about higher ideals.
a40. I emphasize values more than knowledge in my teaching.
a45. I link instructional goals to necessary changes in society.

Appendix B: Pratt's Recommended Steps to Translate the TPI into Other Languages

1. Discussions to determine if each question/item has an approximate equivalent in the new language

2. A first trial translation to begin a first-trial document.

3. Team discussions to determine if the new translation captures the sense of the items in the original language.

4. Team discussions among speakers of different forms of the new language (Portugal/Brazil, Canada/France) to decide how questions should be phrased appropriate to all dialects.

5. Revisions to the new translation...as necessary.

6. A BACK-translation, back into the original language performed by a bilingual expert who was NOT involved in the initial translation steps. (Blind back-translations).

7. Review of the word- and conceptual equivalencies of the back-translation to the original language form.

8. Revisions to the new language translation to bring phrases and concepts into conformance with the original intent.

9. Trial completions of the new translations by 15-20 native speakers of the new language version.

10. Scoring the results of the native speakers according to the established scoring protocol.

11. Discussion among these native speakers about whether each question/item makes sense in the new language.

12. Comparisons of native speaker response profiles to known norms.

13. Revisions to the new language translation on the advice of native-speaker responses.

14. Administration of the revised translation to bilingual speakers; half responding to the new language version first and half to the first language version first.

15. Comparisons of average level of endorsement (item-by-item for all 45 items) between the two language versions.

16. Comparisons of average levels of endorsement for 15 subscales, 5 Perspective scales, 3 Biases scales, and 1 overall total between the two versions.

17. Plotting endorsement means against new language/first language correlations for all 45 + 15 + 5 + 3 + 1 scales.

18. Discussion about the implications of discrepancies in (1) endorsement or (2) correlation.

19. Semi-Final revisions to the new language translation.

20. Administration to a large cohort of new language speakers.

21. Scoring cohort responses according to the established protocol for deriving scales and scale scores.

22. Comparisons of new language cohort responses to national/international norms.

23. Final revisions to wording of the new language translation.

24. Installing the new language version as an online alternative to the first-language version.

Appendix C: Letter for Translators/Expert Panel Members

Jecky Misieng Center for language Studies
UNIMAS
Sarawak Malaysia.

August 30, 2011

Lecturers
Center for Language Studies
UNIMAS
Sarawak
Malaysia.

INVITATION TO BE A TRANSLATOR/EXPERT PANEL MEMBER FOR A RESEARCH STUDY THAT WILL INVOLVE TRANSLATING AN INSTRUMENT FROM ENGLISH TO BAHASA MALAYSIA

1. Respectfully, the above topic is referred to here.
2. I am pleased to inform you that I am now in my final year of my doctoral studies at the University of South Florida majoring in Curriculum and Instruction with specialization in Educational measurement and research.
3. For my dissertation, I will be conducting a study on teacher beliefs about teaching and learning in Higher Education based on an instrument called the Teaching Perspectives Inventory that will be translated from English to Bahasa Malaysia before proceeding to investigate the instrument's psychometric properties pertaining to measurement equivalence.
4. In order to ensure that the translation process is done correctly and effectively, guidelines as proposed by Hambleton and Kanjee (1995) will be used.
5. We are very pleased to say that you have all the requirements needed to be one of our translators or an expert panel member.
6. If you agree to be one of our translators or an expert panel member please respond via email @ jmisieng@usf.edu and I will arrange a meeting with you to discuss the details of the research and translation procedures. If you have any questions, you can do so via the same email address.
7. Lastly, I hope to hear a favorable reply from you as soon as possible.

Thank you.

Sincerely,
(Jecky Misieng)
630101-13-5677

Appendix D: Letter to Participants of Pilot Study

Dear Prof./Dr./Mr./Ms. «Name»,

I am a lecturer at Universiti Malaysia, Sarawak and currently a doctoral candidate at the Department of Measurement and Research at the College of Education, University of South Florida, Tampa. I am now working on my dissertation and the focus of my research is comparing teaching beliefs across cultures specifically between faculty in higher education of the United States and Malaysia. I will be using a very widely used survey instrument called the Teaching Perspectives Inventory (TPI) which was developed by Pratt to measure perceptions of faculty about teaching and learning. The TPI has been translated into many languages but there has not been a study done to find out if the TPI works the same way across cultures. As fellow educators, I would like to ask if you would be willing to volunteer to participate in a web-based survey which is approximately 50 minutes long. You will be participating in the phase of the study that is designed to examine the test-retest reliability of the TPI. In order to do the test-retest, I will have to put a study number on your electronic form so that I can link the first and second response later. Your second response would be within 2 weeks of the first one. Once your data are linked together the code will be removed and the data will be made confidential with no names or identifiers revealed. Although participation in this study may not benefit you directly, your responses will provide valuable insight on how beliefs about teaching influence the aims and goals of instruction and ultimately determine what actually goes in the classroom.

If you are not the right participant for this study, please kindly reply to this e-mail to notify me so that you will not receive any reminder e-mails.

If you are willing to participate in this voluntary study, you will be asked to complete a simple online survey. Needless to say, I understand that you may not wish to participate in this purely voluntary study. This unfunded research is considered to be a minimal risk investigation and compensation is unfortunately not available to pay you for your participation. This research will be confidential in nature, and the survey results will be reported in an aggregate manner. About 1500 individuals will be asked to participate in the study.

If you have any questions, concerns or complaints about this study please contact Jecky Misieng via e-mail at jmisieng@usf.edu. Additionally, if you have questions about your rights as a participant in this study, or have any complaints, concerns or issues you want to discuss with someone outside the research, call the Division of Research Integrity and Compliance of the University of South Florida at (813) 974-9343 (IRB Study Pro00001701).

I appreciate your time and would like to thank you in advance for your consideration participating in this study.By clicking the link below to go directly to the survey, you are hereby granting your informed consent to take part in this research.
http://teslmalaysia.com/tpi/index.php?sid=86443&lang=ms

Appendix E: Letter to Cognitive Interviewees

Dear Prof./Dr./Mr./Ms.,

I am a lecturer at University Malaysia, Sarawak and currently a doctoral candidate at the Department of Measurement and Research at the College of Education, University of South Florida, Tampa. I am now working on my dissertation and the focus of my research is comparing teaching beliefs across cultures specifically between faculty in higher education of the United States and Malaysia. I will be using a very widely used survey instrument called the Teaching Perspectives Inventory (TPI) which was developed by Pratt to measure perceptions of faculty about teaching and learning. The TPI has been translated into many languages but there has not been a study done to find out if the TPI works the same way across cultures. For my study, the TPI will be translated into Malay and will evaluate whether it functions the same way as the original instrument. As fellow educators, I would like to ask if you would be willing to volunteer to participate in a web-based survey which is approximately 50 minutes long.

Although participation in this study may not benefit you directly, your responses will provide valuable insight on how beliefs about teaching influence the aims and goals of instruction and ultimately determine what actually goes in the classroom.

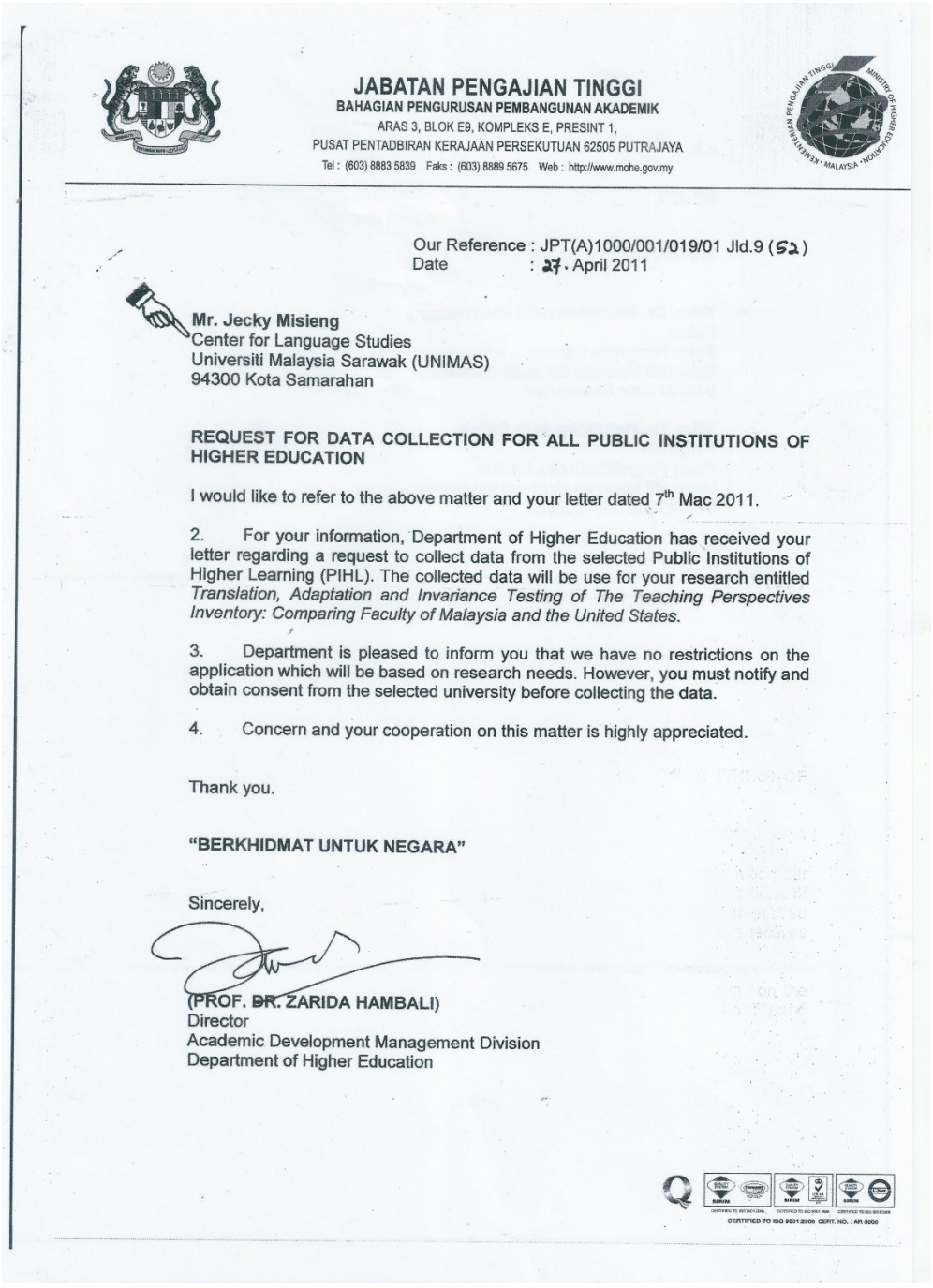
I have chosen you to be among five people out of the twenty-five who were involved in the pilot study of the instrument to be interviewed about the questionnaire items. If you are willing to participate in this cognitive interview, you will be asked probing questions, e.g. what do you believe the question is asking?; what do specific words and phrases in the item mean to you?; what information do you need to recall in order to answer the question?; do you devote enough mental effort to answer it accurately? and can you match your internally generated response to one of the response options given easily?

Needless to say, I understand that you may not wish to participate in this purely voluntary study. This unfunded research is considered to be a minimal risk investigation and compensation is unfortunately not available to pay you for your participation. All of your responses will be confidential in nature and your identity or name will not be identified.

If you have any questions, concerns or complaints about this study please contact Jecky Misieng via e-mail at jmisieng@usf.edu. Additionally, if you have questions about your rights as a participant in this study, or have any complaints, concerns or issues you want to discuss with someone outside the research, call the Department of Health and Human Servicers along with the Division of Research Integrity and Compliance of the University of South Florida at (813) 974-5638 (IRB Study Pro00001701).

I appreciate your time and would like to thank you in advance for consideration participating in this study.

Appendix F: Letter of Approval from the Malaysian Ministry of Higher Education



JABATAN PENGAJIAN TINGGI
BAHAGIAN PENGURUSAN PEMBANGUNAN AKADEMIK
ARAS 3, BLOK E9, KOMPLEKS E, PRESINT 1,
PUSAT PENTADBIRAN KERAJAAN PERSEKUTUAN 62505 PUTRAJAYA
Tel : (603) 8883 5839 Faks : (603) 8889 5675 Web : http://www.mohe.gov.my

Our Reference : JPT(A)1000/001/019/01 Jld.9 (52)
Date : 27. April 2011

Mr. Jecky Misieng
Center for Language Studies
Universiti Malaysia Sarawak (UNIMAS)
94300 Kota Samarahan

REQUEST FOR DATA COLLECTION FOR ALL PUBLIC INSTITUTIONS OF HIGHER EDUCATION

I would like to refer to the above matter and your letter dated 7th Mac 2011.

2. For your information, Department of Higher Education has received your letter regarding a request to collect data from the selected Public Institutions of Higher Learning (PIHL). The collected data will be use for your research entitled *Translation, Adaptation and Invariance Testing of The Teaching Perspectives Inventory: Comparing Faculty of Malaysia and the United States*.

3. Department is pleased to inform you that we have no restrictions on the application which will be based on research needs. However, you must notify and obtain consent from the selected university before collecting the data.

4. Concern and your cooperation on this matter is highly appreciated.

Thank you.

"BERKHIDMAT UNTUK NEGARA"

Sincerely,

(PROF. DR. ZARIDA HAMBALI)
Director
Academic Development Management Division
Department of Higher Education

CERTIFIED TO ISO 9001:2008 CERT. NO. : AR 5006

Appendix G: A Letter to all Malaysian Public Universities

Jecky Misieng
Center for Language Studies,
Universiti Malaysia Sarawak (UNIMAS)
94300 Kota Samarahan
Sarawak

28 August, 2012

Through and cc to:
The Dean,
Center for Language Studies,
Universiti Malaysia Sarawak (UNIMAS)
94300 Kota Samarahan
Sarawak

Vice- Chancellor

Prof./Dr./Sir/Ms,

REQUEST FOR DATA COLLECTION FROM PUBLIC HIGHER EDUCATION INSTITUTIONS MALAYSIA

1. I, Jecky Misieng, a doctoral candidate at the University of South Florida, Tampa (USF) and also a lecturer at UNIMAS, would like your permission to carry out a study on teaching beliefs of faculty in your institution.
2. My study entitled "Translation, Adaptation and Invariance Testing of the Teaching Perspectives Inventory: Comparing Faculty of Malaysia and the United States" will require a total of 1500 lecturers of all public universities in Malaysia whose mother tongue is the Malay language.
3. To protect the privacy of all the respondents, I am not allowed to have direct contact with any of them. With your permission, I would like to contact representatives of your institution to assist me to reach all the Malay lecturers for my research. I would be very grateful if their contact information is made available so that I can email them my survey invitation and they can help forward my email to everyone concerned.
4. It is my hope that all the responses can be collected by the end of this September.
5. Once this instrument is deemed ready for use, it will be uploaded online to be accessed by all Malaysians.
6. For your information, I have obtained permission from the Department of Higher Education to collect these data for my study.
7. The institutional review board (IRB) of the University of South Florida has also approved the study (IRB# Pro00001701).

Thank you.

Sincerely,

(Jecky Misieng)
Lecturer
Center for Language Studies
UNIMAS

c.c.
Director
Department of Higher Education
MOHE
Malaysia

Appendix H: E-mail to Survey Participants

Dear Prof./Dr./Mr./Ms.,

I am a lecturer at Universiti Malaysia, Sarawak and currently a doctoral candidate at the Department of Measurement and Research at the College of Education, University of South Florida, Tampa. I am now working on my dissertation and the focus of my research is comparing teaching beliefs across cultures specifically between faculty in higher education of the United States and Malaysia. I will be using a very widely used survey instrument called the Teaching Perspectives Inventory (TPI) which was developed by Pratt to measure perceptions of faculty about teaching and learning. The TPI has been translated into many languages but there has not been a study done to find out if the TPI works the same way across cultures. As fellow educators, I would like to ask if you would be willing to volunteer to participate in a web-based survey which is approximately 50 minutes long. The questionnaire will be anonymous so we will not ask for your name.

Although participation in this study may not benefit you directly, your responses will provide valuable insight on how beliefs about teaching influence the aims and goals of instruction and ultimately determine what actually goes in the classroom.

If you are not the right participant for this study, please kindly reply to this e-mail to notify me so that you will not receive any reminder e-mails.

If you are willing to participate in this voluntary study, you will be asked to complete a simple online survey. Needless to say, I understand that you may not wish to participate in this purely voluntary study. This unfunded research is considered to be a minimal risk investigation and compensation is unfortunately not available to pay you for your participation. This research will be confidential in nature, and the survey results will be reported in an aggregate manner. About 1500 individuals will be asked to participate in the study.

If you have any questions, concerns or complaints about this study please contact Jecky Misieng via e-mail at jmisieng@usf.edu. Additionally, if you have questions about your rights as a participant in this study, or have any complaints, concerns or issues you want to discuss with someone outside the research, call the Department of Health and Human Servicers along with the Division of Research Integrity and Compliance of the University of South Florida at (813) 974-5638 (IRB Study Pro00001701).

I appreciate your time and would like to thank you in advance for your consideration participating in this study. There will be one follow-up email thanking those who have already agreed to participate and a reminder that the site is still open for those who would like to participate.By clicking the link below to go directly to the survey, you are hereby granting your informed consent to take part in this research.
http://teslmalaysia.com/tpi/index.php?sid=86443&lang=ms

Appendix I: Reminder E-mail to Survey Participants

Dear Prof./Dr./Mr./Ms.,

A couple of weeks ago I emailed you to ask for your voluntary participation in a research study comparing teaching beliefs across cultures specifically between faculty in higher education of the United States and Malaysia. I will be using a very widely used survey instrument called the Teaching Perspectives Inventory (TPI) which was developed by Pratt to measure perceptions of faculty about teaching and learning. The TPI has been translated into many languages but there has not been a study done to find out if the TPI works the same way across cultures. For my study, the TPI will be translated into Malay and will evaluate whether it functions the same way as the original instrument. As fellow educators, I would like to ask if you would be willing to volunteer to participate in a web-based survey which is approximately 50 minutes long. The questionnaire will be anonymous so we will not ask for your name.

If you have completed the questionnaire, please accept my sincere thanks. If you have not completed the questionnaire but would still like to, the site is still open and you can click on the following link:
http://teslmalaysia.com/tpi/index.php?sid=86443&lang=ms

If you have any questions, concerns or complaints about this study please contact Jecky Misieng via e-mail at jmisieng@usf.edu. Additionally, if you have questions about your rights as a participant in this study, or have any complaints, concerns or issues you want to discuss with someone outside the research, call the Department of Health and Human Servicers along with the Division of Research Integrity and Compliance of the University of South Florida at (813) 974-5638 (IRB Study Pro00001701).

I appreciate your time and would like to thank you in advance for your consideration participating in this study.

Appendix J: Letter of Approval from Pratt and Collins

Department of Educational Studies
Mailing address:
2125 Main Mall
Vancouver, B.C. Canada V6T 1Z4

Tel: 604-822-5374
Fax: 604-822-4244
http://www.edst.educ.ubc.ca

Mr. Jecky Misieng
Department of Educational Measurement
and Research, COEDU
4202 E. Fowler Ave. EDU105
University of South Florida
Tampa, FL 33620

October 5, 2011

Dear Mr. Misieng,

Thank you for inquiring about the Teaching Perspectives Inventory (TPI) and outlining your research agenda to translate it into the Malaysian language, then to compare your results with a comparable group of recent respondents from our files.

This letter provides you with official permission to translate the TPI and to use it in your graduate research. We trust that it will satisfy any permission requirements from your thesis committee as well as from the Institutional Review Board at the University of South Florida. As well, should any Malaysian Authorities question your use of the TPI for translation and research purposes, this letter will also confirm our permission to proceed.

We look forward to your progress and results and assure you our willingness to help in any way possible. We are pleased to note the enthusiasm for the TPI in many countries, including several throughout Asia.

Please notify us how we can best support your work, and let us know if you encounter any questions regarding official permission to proceed.

Sincerely,

John B. Collins

John B. Collins, Adjunct Professor
(for) Daniel D. Pratt, Professor

Appendix K: Intraclass Correlations for the Malaysian sample

SUMMARY OF DATA

Number of clusters	20
Average cluster size	28.050

Estimated Intraclass Correlations for the Y Variables

Variable	Correlation	Variable	Correlation	Variable	Correlation
B1	0.053	B2	0.054	B4	0.057
B5	0.034	B6	0.053	B7	0.045
B8	0.052	B9	0.066	B10	0.027
B11	0.041	B12	0.052	B13	0.062
B14	0.043	B15	0.052	I16	0.039
I17	0.031	I18	0.040	I19	0.037
I20	0.017	I21	0.035	I22	0.028
I23	0.055	I24	0.033	I25	0.033
I26	0.054	I27	0.018	I28	0.046
I29	0.026	I30	0.035	A31	0.030
A32	0.026	A33	0.036	A34	0.041
A35	0.041	A36	0.043	A37	0.033
A38	0.021	A39	0.021	A40	0.037
A41	0.058	A42	0.038	A43	0.063
A44	0.026	A45	0.058		

Appendix L: Results of Test-Retest Correlations

Item	Administration	Mean	SD	Correlations
b1	initial test	4.50	0.58	0.37
	retest	4.27	0.82	
b2	initial test	4.27	0.72	0.65*
	retest	4.27	0.78	
b3	initial test	4.15	1.08	0.67*
	retest	4.00	0.94	
b4	initial test	4.42	0.58	0.53*
	retest	4.38	0.75	
b5	initial test	3.62	1.06	0.55*
	retest	3.46	1.21	
b6	initial test	3.46	0.95	0.81*
	retest	3.73	1.00	
b7	initial test	4.04	0.60	0.77*
	retest	4.00	0.69	
b8	initial test	4.15	0.97	0.86*
	retest	4.04	0.96	
b9	initial test	4.31	0.68	0.57*
	retest	4.00	0.94	
b10	initial test	4.27	0.72	0.79*
	retest	4.19	0.75	
b11	initial test	4.31	0.55	0.12
	retest	4.38	0.57	
b12	initial test	4.15	0.97	0.70*
	retest	4.00	1.06	
b13	initial test	3.81	0.85	0.51*
	retest	3.58	1.10	
b14	initial test	4.27	0.60	0.89*
	retest	4.27	0.60	
b15	initial test	4.58	0.64	0.70*
	retest	4.54	0.71	
i16	initial test	3.19	0.90	0.76*
	retest	3.27	0.92	
i17	initial test	4.69	0.47	0.38
	retest	4.38	0.70	
i18	initial test	4.46	0.76	0.66*
	retest	4.19	0.94	
i19	initial test	4.54	0.58	0.65*
	retest	4.27	0.87	
i20	initial test	4.15	0.78	0.59*
	retest	3.92	0.98	

i21	initial test	4.15	0.88	0.58*
	retest	3.96	1.04	
i22	initial test	4.27	0.87	0.79*
	retest	4.15	0.93	
i23	initial test	3.88	0.86	0.78*
	retest	3.85	0.93	
i24	initial test	4.19	0.80	0.84*
	retest	4.04	0.82	
i25	initial test	4.00	1.13	0.86*
	retest	3.85	1.16	
i26	initial test	4.23	1.03	0.20
	retest	4.15	1.01	
i27	initial test	3.92	1.06	0.70*
	retest	3.88	1.07	
i28	initial test	3.88	1.07	0.85*
	retest	3.77	1.11	
i29	initial test	4.15	0.73	0.49
	retest	3.77	0.99	
i30	initial test	3.73	1.04	0.83*
	retest	3.58	1.07	
a31	initial test	3.77	0.95	0.98*
	retest	3.73	0.96	
a32	initial test	4.19	0.75	0.97*
	retest	4.23	0.71	
a33	initial test	4.27	0.67	0.92*
	retest	4.19	0.90	
a34	initial test	3.19	1.13	0.88*
	retest	3.00	1.20	
a35	initial test	3.58	0.76	0.81*
	retest	3.54	0.91	
a36	initial test	4.04	0.87	0.85*
	retest	3.96	0.87	
a37	initial test	4.00	0.80	0.97*
	retest	3.96	0.77	
a38	initial test	3.58	1.07	0.75*
	retest	3.42	1.14	
a39	initial test	3.46	1.03	0.98*
	retest	3.50	0.99	
a40	initial test	3.81	0.85	0.55*
	retest	3.73	0.92	
a41	initial test	4.08	0.63	0.91*
	retest	4.08	0.69	
a42	initial test	3.92	0.89	0.32
	retest	3.73	1.04	

a43	initial test	3.46	1.14	0.57*
	retest	3.08	1.23	
a44	initial test	3.77	0.95	0.86*
	retest	3.58	1.10	
a45	initial test	3.65	1.09	0.62*
	retest	3.50	1.03	

Note. n=26, *p=.00

Appendix M: Teaching Perspectives Inventory Items – Bahasa Malaysia Version Sorted by Perspective

Transmission

b1. Pembelajaran diperkukuh apabila mempunyai objektif yang telah ditentukan terlebih dahulu.
b6. Para pengajar sepatutnya menjadi pengamal yang luarbiasa terhadap subjek yang diajar.*
b11. Pengajar yang berkesan mesti terlebih dahulu pakar dalam bidangnya.

i16. Hasrat saya adalah untuk mempersiap-kan individu untuk peperiksaan.
i21. Saya meng-kehendaki individu untuk menguasai banyak maklumat berkaitan subjek.
i26. Saya mahu individu memperoleh keputusan cemerlang dalam peperiksaan hasil daripada pengajaran saya.

a31. Saya menyelesaikan keperluan kandungan kursus dengan tepat dan dalam masa yang diperuntukkan.
a36. Pengajaran saya berpandukan objektif kursus.
a41. Saya menerangkan dengan jelas kepada individu tentang perkara yang akan mereka pelajari.

Apprenticeship

b2. Untuk menjadi seorang pengajar yang berkesan, seseorang itu mesti juga pengamal yang berkesan.
b7. Pembelajaran terbaik wujud daripada kerjasama dengan pengamal-pengamal yang baik.
b12. Ilmu pengetahuan dan aplikasinya tidak dapat dipisahkan.

i17. Hasrat saya adalah untuk menunjuk ajar cara melakukan sesuatu atau bekerja dalam situasi sebenar.
i22. Saya meng-kehendaki individu untuk mengetahui cara mengaplikasi kandungan pelajaran dalam situasi sebenar.
i27. Saya mahu individu memahami realiti bekerja dalam dunia sebenar.

a32. Saya meng-hubungkaitkan kandungan pelajaran secara praktis dengan dunia sebenar atau aplikasi.
a37. Saya mencontohi kemahiran dan kaedah pengajaran yang baik.
a42. Saya memastikan individu yang kurang berpengalaman belajar daripada mereka yang lebih berpengalaman.

Developmental

b3. Yang paling penting ialah proses pembelajaran bergantung kepada asas pengetahuan sedia ada pada seseorang.
b8. Pengajaran harus berfokus kepada membina perubahan kualitatif dalam pemikiran.
b13. Proses pengajaran seharusnya berasaskan pengetahuan sedia ada seseorang.

i18. Hasrat saya adalah untuk membantu individu mengembang-kan penaakulan yang lebih kompleks.
i23. Saya mengkehendaki individu untuk membangunkan kaedah baru dalam mempertimbangkan hal-hal berkaitan kandungan pelajaran.
i28. Saya mahu individu melihat betapa kompleks dan saling bergantungnya sesuatu perkara itu.

a33. Saya bertanya banyak soalan semasa mengajar.
a38. Saya mencabar kaedah-kaedah lazim yang digunakan untuk memahami kandungan pelajaran.
a43. Saya menggalakkan individu mencabar pemikiran antara satu sama lain.

Nurturing

a4. Penting untuk saya mengambil kira reaksi emosi pelajar.
a9. Dalam pengajaran saya, membina keyakinan diri dalam diri pelajar menjadi keutamaan.
a14. Dalam pembelajaran, usaha individu perlu diberi ganjaran setimpal dengan pencapaian-nya.

i19. Hasrat saya adalah untuk membina keyakinan dan harga diri individu sebagai pelajar.
i24. Saya berharap individu dapat meningkatkan harga diri mereka melalui pengajaran saya.
i29. Saya mahu menyediakan keseimbangan antara mengambil berat dan mencabar kemampuan pelajar semasa saya mengajar.

a34. Saya mencari sesuatu untuk dipuji dalam setiap sumbangan seseorang.
a39. Saya menggalakkan ekspresi perasaan dan emosi.
a44. Saya berkongsi perasaan saya dan meng-kehendaki pelajar saya juga berbuat demikian.

Social Reform

b5. Pengajaran saya berfokus kepada perubahan masyarakat, tidak pada seseorang pelajar.
b10. Pembelajaran individu tanpa perubahan sosial adalah tidak memadai.
b15. Pada saya, mengajar ialah satu tindakan moral yang juga aktiviti intelektual.

i20. Hasrat saya adalah untuk mencabar individu mempertimbangkan semula nilai diri secara serius.
i25. Saya meng-kehendaki individu untuk komited melakukan perubahan kepada masyarakat.

i30. Saya mahu mendedahkan perihal masyarakat yang diambil mudah oleh individu.
a35. Saya menggunakanbahan pengajaran sebagai cara untuk mengajar mencapai kesempurnaan yang lebih tinggi*
a40. Saya lebih memberi penekanan kepada nilai-nilai murni dalam pengajaran saya berbanding ilmu pengetahuan.
a45. Saya menghubung-kaitkan matlamat pengajaran dengan perubahan yang diperlukan dalam masyarakat.

Appendix N: 20 Malaysian Public Universities' Academic Staff by Position and by Gender in 2010

Institution	Professors		Associate Professors		Lecturers		Language Teachers		Tutors		Grand Total		
	M	F	M	F	M	F	M	F	M	F	M	F	Total
UM	233	116	225	189	553	579	47	45	95	125	1153	1054	2207
USM	156	27	262	125	583	521	52	127	0	0	1053	800	1853
UKM	190	84	220	225	565	733	28	63	57	101	1060	1206	2266
UPM	127	46	202	133	382	497	12	39	0	0	723	715	1438
UTM	137	29	296	89	683	513	0	0	130	167	1246	798	2044
UUM	23	7	74	41	434	485	11	23	42	127	584	683	1267
UIAM	106	12	152	56	448	460	202	356	12	59	920	943	1863
UNIMAS	41	1	52	23	215	253	6	11	43	80	357	368	725
UMS	31	6	82	30	305	289	7	8	35	46	460	379	839
UPSI	35	4	35	8	246	208	11	25	55	121	382	366	748
UiTM	91	37	556	529	2490	4528	0	0	119	45	3256	5139	8395
UniSZA	12	2	11	8	152	154	4	12	55	103	234	279	513
UMT	14	2	34	12	130	156	2	9	25	71	205	250	455
USIM	24	5	6	10	143	166	12	41	37	74	222	296	518
UTHM	23	5	35	4	314	223	5	15	161	158	538	405	943
UTeM	18	0	25	3	314	208	3	9	108	55	468	275	743
UMP	20	4	21	8	241	184	3	17	21	15	306	228	534
UniMAP	21	2	40	4	235	173	6	14	95	63	397	256	653
UMK	16	0	8	2	50	51	5	9	15	32	94	94	188
UPNM	8	2	13	2	76	64	4	8	9	27	110	103	213
Total	1326	391	2349	1501	8559	10445	420	831	1114	1469	13768	14637	28405

Appendix O: Sum of Scores for the 5 Perspectives for the Test-Retest Study Sample (n=25)

Scale	Administration	Min.	Max	*M*	*SD*	Skewness	Kurtosis	Effect Size
Transmission	Test	30.00	43.00	35.60	3.03	0.11	0.43	0.07
	Retest	31.00	43.00	35.40	2.94	0.33	0.37	
Apprenticeship	Test	33.00	45.00	37.60	3.24	0.27	-0.66	0.26
	Retest	30.00	45.00	36.72	3.52	0.47	0.03	
Developmental	Test	28.00	43.00	35.84	4.54	-0.23	-1.10	0.37
	Retest	28.00	43.00	34.24	4.04	.49	-0.63	
Nurturing	Test	31.00	43.00	36.40	3.38	0.17	-0.64	0.47
	Retest	29.00	41.00	34.84	3.20	0.4.0942	-0.03	
Social Reform	Test	28.00	43.00	35.52	4.09	0.21	-0.12	0.30
	Retest	26.00	43.00	34.40	3.42	-0.30	2.19	

Effect size = (Mean for Test – Mean for Retest)/ Pooled *SD*.

Appendix P: Summary of TPI Alternative Models

Model	Country	χ^2	df	CFI	RMSEA	SRMR
Five-factor model	U.S.	2539.71	892	0.75	0.06	0.06
	Malaysia	7783.63	892	0.43	0.12	0.14
Three-factor model*	U.S.	-	-	-	-	-
	Malaysia	6314.12	899	0.55	0.10	0.12
Higher Order model	U.S.	2340.85	920	0.79	0.05	0.06
	Malaysia	7345.66	920	0.48	0.11	0.14

**Model 2 Not feasible for U.S. due to non-convergence with and without b3.*

Auto Invariance for Five-factor Model of Teaching Perspectives

Invariance	χ^2	Satorra-Bentler Scaled Δ X^2	*P*=.001	*df*	Δ *df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Configural	10691.43	-	-	1784		0.54	-	0.09	-	0.11	-
Metric	10952.47	254.92	No	1823	39	0.52	0.02	0.09	0	0.12	0.01
Scalar	12072.26	1310.01	No	1862	78	0.47	0.07	0.10	0.01	0.13	0.02

Appendix Q: U.S. TPI Subscores (n=605)

Scale	Subscale	Min.	Max	*M*	*SD*	Skewness	Kurtosis
Transmission	T_B	6.00	15.00	11.31	1.90	-0.21	-0.53
	T_I	3.00	15.00	10.07	2.09	-0.17	0.09
	T_A	7.00	15.00	12.04	1.66	-0.28	-0.19
Apprenticeship	A_B	5.00	15.00	11.41	1.88	-0.45	0.29
	A_I	5.00	15.00	12.88	1.91	-0.89	0.63
	A_A	7.00	15.00	12.28	1.56	-0.37	0.05
Developmental	D_B	3.00	15.00	10.56	1.84	-0.07	0.21
	D_I	3.00	15.00	12.82	1.72	-0.87	1.7
	D_A	5.00	15.00	11.99	1.96	-0.56	0.10
Nurturing	N_B	3.00	15.00	11.59	1.85	-0.58	0.84
	N_I	3.00	15.00	12.21	2.13	-0.74	0.46
	N_A	3.00	15.00	11.01	2.40	-0.47	0.05
Social Reform	S_B	4.00	15.00	9.18	2.08	-0.12	0.12
	S_I	3.00	15.00	9.42	2.57	-0.04	-0.22
	S_A	3.00	15.00	9.46	2.44	-0.16	-0.26
Belief		39.00	72.00	54.06	5.84	-0.06	-0.02
Intention		37.00	74.00	57.41	6.44	-0.11	0.06
Action		33.00	75.00	56.76	6.90	-0.13	-0.01

Appendix R: Malaysia TPI Subscores (n=561)

Scale	Subscale	Min.	Max	*M*	*SD*	Skewness	Kurtosis
Transmission	T_B	4.00	15.00	11.77	2.00	-1.33	3.58
	T_I	6.00	15.00	11.13	2.00	-0.26	-0.38
	T_A	5.00	15.00	12.85	1.69	-1.17	2.75
Apprenticeship	A_B	4.00	15.00	12.48	2.00	-1.50	4.16
	A_I	7.00	15.00	13.01	1.71	-0.86	0.81
	A_A	7.00	15.00	12.27	1.65	-0.34	0.34
Developmental	D_B	4.00	15.00	11.66	1.95	-0.76	1.04
	D_I	6.00	15.00	11.72	2.25	-0.46	-0.37
	D_A	5.00	15.00	11.45	2.29	-0.39	-0.56
Nurturing	N_B	3.00	15.00	12.03	1.86	-0.54	0.80
	N_I	7.00	15.00	12.49	2.06	-0.36	-0.94
	N_A	3.00	15.00	10.27	2.61	0	-0.73
Social Reform	S_B	3.00	15.00	11.35	2.17	-1.22	3.54
	S_I	4.00	15.00	11.53	2.53	-0.37	-0.58
	S_A	4.00	15.00	11.22	2.35	-0.21	-0.63
Belief		21.00	75.00	59.29	8.21	-1.69	6.24
Intention		34.00	75.00	59.89	8.66	-0.30	-0.52
Action		28.00	75.00	58.05	8.49	-0.22	0.04

Appendix S: Summary of Subscores by Country (n=561)

		Malaysia (*n*=561)		U. S. (*n*=605)		
Scale	Subscale	*M*	*SD*	*M*	*SD*	Effect Size
Transmission	T_B	11.77	2.00	11.31	1.90	.24
	T_I	11.13	2.00	10.07	2.09	.52
	T_A	12.85	1.69	12.04	1.66	.48
Apprenticeship	A_B	12.48	2.00	11.41	1.88	.55
	A_I	13.01	1.71	12.88	1.91	.07
	A_A	12.27	1.65	12.28	1.56	-.01
Developmental	D_B	11.66	1.95	10.56	1.84	.58
	D_I	11.72	2.25	12.82	1.72	-.55
	D_A	11.45	2.29	11.99	1.96	-.25
Nurturing	N_B	12.03	1.86	11.59	1.85	.24
	N_I	12.49	2.06	12.21	2.13	.13
	N_A	10.27	2.61	11.01	2.40	-.30
Social Reform	S_B	11.35	2.17	9.18	2.08	1.02
	S_I	11.53	2.53	9.42	2.57	.83
	S_A	11.22	2.35	9.46	2.44	.73
Belief		59.29	8.21	54.06	5.84	.73
Intention		59.89	8.66	57.41	6.44	.32
Action		58.05	8.49	56.76	6.90	.17

Appendix T: Expedited Approval for Initial Review IRB#: Pro00001701

USF
UNIVERSITY OF
SOUTH FLORIDA

DIVISION OF RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-5618

January 13, 2012

Jecky Misieng
Edu Measurement & Research

RE: Expedited Approval for Initial Review
IRB#: Pro00001701
Title: Translation, Adaptation and Invariance Testing of the Teaching Perspectives Inventory (TPI): Comparing Faculty of Malaysia and the United States

Dear Jecky Misieng:

On 1/13/2012 the Institutional Review Board (IRB) reviewed and APPROVED the above referenced protocol. Please note that your approval for this study will expire on 1-13-2013.

Approved Items:
Protocol Document(s):

Proposal for Translation and Adaptation of the TPI from English to Bahasa Malaysia

Consent/Assent Documents:

Name
Online Consent form with a Waiver of Informed Consent Documentation
(Consent forms with Waivers are not stamped)

It was the determination of the IRB that your study qualified for expedited review which includes activities that (1) present no more than minimal risk to human subjects, and (2) involve only procedures listed in one or more of the categories outlined below. The IRB may review research through the expedited review procedure authorized by 45CFR46.110 and 21 CFR 56.110. The research proposed in this study is categorized under the following expedited review category:

(7) Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Your study qualifies for a waiver of the requirements for the documentation of informed consent as outlined in the federal regulations at 45CFR46.117 (c) which states: an IRB may waive the requirement for the investigator to obtain a signed consent form for some or all subjects if it finds either:(1) That the only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality. Each subject will be asked whether the subject wants documentation linking the subject with the research, and the subject's wishes will govern; or (2) that the research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context.

As the principal investigator of this study, it is your responsibility to conduct this study in accordance with IRB policies and procedures and as approved by the IRB. Any changes to the approved research must be submitted to the IRB for review and approval by an amendment.

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

Sincerely,

John Schinka, PhD, Chairperson
USF Institutional Review Board

Cc: Various Menzel, CCRP
USF IRB Professional Staff