From West to East: Adoption of Western Measurement Scales in Taiwan’s Organizational Research

Chu-Chen Rosa Yeh\textsuperscript{a,}\textsuperscript{∗}, Chu-Yen Lin\textsuperscript{b}, Shu-Yuan Chen\textsuperscript{c}

\textsuperscript{a}Graduate Institute of International Human Resource Development, National Taiwan Normal University, Taiwan
\textsuperscript{b}Department of Applied Psychology, Hsuan Chuang University, Taiwan
\textsuperscript{c}Department of Business Management, National United University, Taiwan

Abstract

Adopting existing measures in their research is common for organizational researchers. However, this practice is questionable when proper validation procedures are lacking. This study surveyed the extent of this practice in Taiwan’s academic society to determine the severity of the situation. The target of this investigation was those articles that include latent constructs in an organizational research and that were published in major Taiwanese academic journals between 2004 and 2008. This effort uncovered 146 organizational studies containing 791 research constructs. Of the 791 constructs, 263 were translated from an existing foreign scale, and 352 were measured using an adaptation of an existing foreign scale. Organizational commitment, job satisfaction, and performance were found to be the most frequently studied constructs in Taiwan during this period. Most studies applying existing Western scales to Taiwanese samples did not report proper scale validation procedures to ensure measurement equivalence. The implications of these research findings are discussed, and suggestions are provided to improve the rigor of organizational research.

Keywords: Organizational research, research instrument, scale validation, TSSCI, Taiwan

1. Introduction

Adopting existing measures is a common practice of organizational researchers. This practice is questionable when researchers fail to take necessary procedures to cross-examine the validity and the reliability of an existing measure in the new research context. This study surveyed the extent of this practice in Taiwan’s academic society with the purpose of understanding and improving the rigor of organizational research in Taiwan.

The field of organizational research deals with many constructs that are latent in nature. That is, these constructs are abstract, implicit, and cannot be directly measured. Examples of frequently researched latent constructs include organizational climate, leadership, organizational commitment, organizational learning, organizational citizenship behaviors, and employee satisfaction, among others. These latent constructs are important both academically and practically in understanding the inner workings of organizational management and organizational behaviors. However, the development of valid measures or measurement instruments for these latent constructs has always been a challenge for organizational management researchers. Aside from extra time and effort, it requires in-depth knowledge and experience in both the theory behind the construct and the most current methodological know-

\textsuperscript{∗}Corresponding author. Email: rosayeh@ntnu.edu.tw
how. Moreover, as social science methodology progresses, the bar is raised higher for the academic society to accept a newly developed measure. As a result, most researchers are discouraged to develop on their own and are more inclined to adopt measurement scales that have been published before when taking on a research.

This phenomenon is worsened when researchers work under the pressure of getting published. Many researchers would rush to undertake substantive research without any intent to devote time in crafting valid research instruments (Farh, Cannella and Lee, 2006). In Taiwan, the pressure is not only real, but it is also intensified because of an urgent need for social science researchers to publish in international journals, specifically journals indexed in the Social Science Citation Index (SSCI). To increase their chances of having their manuscripts accepted by international journals, organizational researchers in Taiwan often adopt the measurement scales that have appeared in internationally renowned journals, with most of these scales originally developed by scholars in the United States or other Western countries. This practice has become common for researchers who intend to 1) devote valuable research time to test substantive hypotheses, 2) avoid being challenged in their validation process of a new research instrument, or 3) compare results with other studies using the same research construct.

However, this practice has been under attack by a great number of research methodologists for many years because of the potential threats to measurement equivalence/invariance (ME/I) involved in this type of research (Cheung and Rensvold, 1999, 2002; Farh, Cannella and Lee, 2006; Little, 1997; Reise, Widaman and Pugh, 1993; Singh, 1995; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000). ME/I is required to ensure that the members of different groups have the same interpretation of the items used in survey-type instruments (Cheung and Rensvold, 2002). Researchers invoke assumptions about measurement equivalence when they adopt or translate an existing measurement scale in their substantive research.

There are many sources of threats to ME/I in this type of research, for instance, cultural difference, language/translation problem, organization influence, and response context (Robert, Lee and Chan, 2006). Cultural difference is a major source of threat to ME/I. It stems from the adoption of a measure for an emic construct without testing first whether the measure was developed under an etic or an emic assumption. The assumption of an emic construct is that it is culture specific, whereas an etic construct is universal or culturally invariant (Farh, Cannella and Lee, 2006). Theoretically, emic accounts describe thoughts and actions primarily in terms of the actor’s self-understanding of terms that are often historically or culturally bound (an insider’s view), whereas etic accounts (an outsider’s view) describe the phenomena in constructs that can be applied across cultures (Morris, Leung, Ames and Lickel, 1999). Although the distinction is described as a dichotomy, it is possible for an instrument to contain both aspects. When social science researchers adopt measures that were developed and validated in the United States or other Western cultures and use them in their local studies without first examining whether the construct under investigation is culture specific or that ME/I has been achieved, they do not know whether a measurement scale actually measures the same construct they want to measure. Therefore, the validity of the practice of adopting Western research instruments in local organizational studies in Taiwan is seriously undermined.

Korman (1974, p.194) argues, “the point is not that adequate measurement is ‘nice’. It is necessary, crucial, etc. Without it, we have nothing” Schoenfeldt (1984, p.78) echoes the same concern by offering this conclusion: “The construction of the measuring devices is perhaps the most important segment of any study. Many well-conceived research studies have never seen the light of day because of flawed measures”. According to Nunnally and Bernstein (1994, p.84), “validation is an unending process… Most psychological measures
need to be constantly evaluated and reevaluated to see if they are behaving as they should.” Many researchers agree that the issue of construct validation is by far the most important in the research planning process. However, as adopting existing measures has become a very common practice among researchers, tests of ME/I are now a necessary step in ensuring construct validity.

Almost all theories and data of contemporary management come from Western populations (e.g., Europeans, North Americans, Australians, etc.). However, 70% of people live in non-Western cultures (Triandis, 1995). If management is to become a universal discipline, it will need theories and data from the majority of the people. This assertion is especially true for the international business discipline, as it deals with cross-cultural issues and concerns emerging from the management of individuals with different cultural contexts. As Taiwan enters into the global village, more and more Taiwanese corporations are faced with decisions requiring in-depth understanding of the phenomenon involved. Many organizational research measures have been proved to be helpful in understanding organizational phenomenon. However, most measurement scales were developed and validated entirely in the Western context and possibly with student samples. When these measures are adopted in research involving Taiwanese business samples without ensuring first that the measurement equivalence has been achieved, inference from the study result may be invalid. In the past few decades, works by scholars from Taiwan, Hong Kong, and other places have discussed and debated the limitations of adopting Western concepts and methods to study non-Western populations (e.g., Yang, 1982, 2000; Yang and Bond, 1990; Yang and Chiu, 1987). Determining whether the ideas expressed in the literature have already made an impact on scale development practices in Taiwan’s organizational research would be informative.

This study proposes a large-scale study to survey the extent of adoption of Western research instruments in organizational research in Taiwan. The investigation focuses on the scale development approach in major Taiwanese academic journals, specifically those listed in the TSSCI index, targeting latent constructs in organizational research. The survey uses Farh, Cannella, and Lee’s 2006 framework of scale development approaches in Chinese management research, which categorizes the development of organizational research constructs into four approaches, namely, translation, de-contextualization, adaptation, and contextualization, according to two dimensions: 1) source of the scale and 2) expectations about cultural specificity. Research constructs that were measured with a direct translation of a Western-developed research instrument are tallied to uncover the most frequently adopted instruments by organizational researchers in Taiwan. This study hopes to achieve two objectives: to understand the practice in scale development in Taiwan’s organizational research society and to investigate the extent of practice in adopting Western measurement scales in Taiwan’s organizational research. After the current practice is revealed, implications and suggestions are discussed to improve the rigor of organizational research in Taiwan.

2. Research background

2.1 Bias and equivalence in cross-cultural assessment

Culture has a significant role in human behavior. The importance of culture has been recognized by a number of scholars for many years (Betancourt and Lopez, 1993). According to Triandis et al.’s (1980) definition, culture (or subjective culture) includes elements such as social norms, roles, beliefs, and values. These elements have a wide range of topics such as tradition, familial roles, affective styles, and values on personal control, individualism/collectivism, spirituality, and religiosity. As culture is composed of values and beliefs shared by the people of a society, it can produce significant psychological effects at
the individual level. The literature offers different views on how culture relates to psychological constructs (e.g., Oyserman and Lee, 2007; 2008; Oyserman and Sorensen, 2009). Take the commonly recognized cultural dimension of individualism versus collectivism (Hofstede, 1980) for example. This dimension depicts contrasting cultural values between Western and Eastern societies, with the Western considered more individualistic and the Eastern more collectivistic. Collectivists emphasize more on social contexts, situational constraints, and social roles, all of which shape the styles of thinking and reasoning different from those of individualists (Markus and Kitayama, 1991; Oyserman and Lee, 2008). Moreover, collectivists with interdependent selves are considered more attentive and sensitive to others than are individualists with independent selves (Oyserman and Lee, 2008; Triandis, 1989; 1995). Empirical studies show that the Chinese motive patterns directly reflect the collectivist or group-orientated tradition. The Chinese respondents revealed relatively high levels of need for socially orientated achievement (Yang and Lu, 2005), abasement (Bond, 1986), harmony (Westwood, 1997), affiliation (Hui and Villareal, 1989), nurturance, and order (Bond, 1986). These motives mostly originate from the desire to meet the expectations of significant others and groups, which are very different from the self-centered norm (Bontempo and Rivero, 1992) in the Western society.

Borrowing an instrument developed from a culture or society other than one’s own could lead to serious problems in research because of the potential cultural differences. When an instrument or scale is developed in one society and applied in a different cultural context, bias could occur. As the invariance of psychometric properties (i.e., reliability and validity) cannot be assumed, the score comparisons across cultures become invalid (Robert, et al., 2006; Vandenberg and Lance, 2000). Van de Vijver and Tanzer (1997) introduce three kinds of cross-culture assessment bias. The first one is construct bias, which occurs when the construct measured is not identical across cultural groups. When the concept of a construct is different in various cultures, the measure developed in or for one culture can be partial. The second is method bias, which can come from the incomparability of samples in aspects other than the target variable, problems from instrument characteristics (instrument bias), or problems in instrument administration (administration bias). The third is item bias or differential item functioning, which refers to distortions at the item level that causes different psychological meaning across cultures.

According to Van de Vijver and Tanzer (1997), three types of equivalence exist. The first type is construct equivalence, which is also called structural equivalence or functional equivalence. Based on the “universal” or “etic” position, construct equivalence indicates that all cultural groups measure the same construct, regardless of whether the measurement is based on identical instruments across all cultures. The second type is measurement unit equivalence, which is obtained when two metric measures have the same measurement unit but different origins (e.g., Kelvin and Celsius temperature scales). The third type is scalar equivalence (or full scale equivalence), which indicates that two metric measures have the same measurement unit and origin. The bias mentioned above can lower the level of equivalence. Construct bias is likely to lead to conceptual inequivalence, and method and item biases can easily compromise scalar equivalence. To obtain valid cross-cultural score comparisons, researchers must assume the responsibility to reduce these measurement biases and increase measurement equivalence.

2.2 Cross-cultural evidence of measurement bias

As previously discussed, several common measurement biases can occur when researchers conduct cross-cultural studies or use an instrument developed from another culture. First, because a construct may contain different concepts in distinct cultures, construct bias may likely occur if there is only partial overlap in the definitions of the
construct across cultures. For example, Yang and Bond (1990) argue that the salient Chinese personality traits (factors) are different from those of Americans. They used both Chinese and American descriptors for Chinese subjects and found that, of the five Chinese factors identified, only four corresponded to the American personality factors. Cheung et al. (2003) found that, based on the joint factor analysis of an indigenous Chinese personality assessment instrument (CPAI) and NEO-FFI (Costa and McCrae, 1992; Cheung et al., 2003), the interpersonal relatedness factor is not covered by NEO-FFI, and the openness domain of NEO-FFI is not covered by the original CPAI. An openness scale was added to the subsequent development of CPAI, i.e., CPAI-2 (Cheung, Cheung and Zhang, 2004).

Another common bias is frequently caused by inadequate item translation (Hambleton, 1994). Van de Vijver and Tanzer (1997) cited many relevant situations that could cause item bias during translation, such as ambiguities in the original item, low familiarity or low appropriateness of the item content in certain cultures, and influence of cultural specifics associated with the item wording. Oyserman and Lee (2008) note that, as language is closely related to culture, many culture-relevant values, norms, goals, beliefs, attitudes, and cognitive and motivational styles can be easily primed in one language but may not be so after being translated to another language. A good illustration is Yang’s (1982) study experience; he found that many Western psychological measurements use the method of semantic differential. This method typically uses two opposed descriptors (antonyms) at two opposite poles of an item (e.g., weak vs. strong), but finding antonyms in the Chinese language is difficult. Therefore, this method can induce more difficulties and bias in translation.

The response set or style, such as social desirability, acquiescence, and middle point (or extremity) scoring, can also induce method bias, which in turn jeopardizes the validity of cross-culture comparisons. Harzing (2006) conducted a survey of 26 countries and found that major differences exist in the response styles between countries. Country-level characteristics such as power distance, collectivism, uncertainty avoidance, and extraversion significantly influence response styles such as acquiescence and extreme response styles. Moreover, English-language questionnaires are shown to elicit a higher level of middle responses, whereas questionnaires in the respondent’s native language result in more extreme response styles.

Based on the previous discussion, the importation of a foreign instrument without any prior examination (e.g., examination of the construct content) or posteriori procedures (e.g., item analysis) clearly cannot establish measurement equivalence. Without measurement equivalence, it can be difficult to claim that the scores obtained from the same measurement can be compared in different cultures or societies. Therefore, some common practices can be used to enhance the validity of the instrument in different target populations, such as test adaptations or development of an indigenous instrument (Van de Vijver and Poortinga, 1997).

2.3 Scale development approaches in Chinese management research

Farh, Cannella, and Lee (2006) classified the survey scales or measures currently used in Chinese management research along two dimensions: source of the scale and expectations about its cultural specificity. Based on these two dimensions, they differentiated four approaches to scale development: translation, adaptation, de-contextualization, and contextualization, with each having different key assumptions, strengths, limitations, and roles in Chinese management research (Table 1). “Translation” involves using a direct translation of an existing Western scale to create a Chinese language version of the scale. “Adaptation” also involves translating an existing scale but with some modification to create a more meaningful Chinese version. “De-contextualization” involves assembling a scale from scratch in the Chinese context under the assumption that the target construct is etic (i.e., universal or culturally invariant). “Contextualization” involves developing a scale from
scratch in the Chinese context but under the assumption that the target construct is emic (i.e., relevant to China only).

Table 1. Four types of scale development approaches in Chinese management research

<table>
<thead>
<tr>
<th>Source of the scale</th>
<th>Expectations about cultural specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Etic orientation</td>
</tr>
<tr>
<td>Use or modify an existing scale</td>
<td>Translation</td>
</tr>
<tr>
<td>Develop a new scale</td>
<td>De-contextualization</td>
</tr>
</tbody>
</table>

Source: Farh, Cannella and Lee (2006)

The translation approach adopts two key assumptions: (1) the target construct is equivalent across cultures in terms of overall definition, content domain, and empirical representations of the content domain; and (2) a high-quality culturally unbiased Western scale for the target construct is available. Most researchers use the direct translation approach to develop their scales because this approach not only saves development time and cost in the validation procedure required by a brand new scale but also enables researchers to compare their findings directly with those from Western settings. However, with a direct translation approach, achieving semantic equivalence between the Chinese and Western scales is sometimes difficult. Culturally unbiased Western scales are difficult to come by.

The adaptation approach assumes that the target construct is equal between cultures in terms of overall definition and content domain and that high-quality Western scales for the target construct are available for adaptation. During the adaptation process, some items are modified and other items are used “as is.” The strengths of this approach include low-to-moderate developmental time and cost as well as ease of scholarly exchange of research findings with Western literature. Conversely, the major drawbacks of an adaptation approach are the difficulty in conducting cross-cultural research and the possibility of drastic adaptation creating a new scale that requires extensive validation in the Chinese context.

The key assumptions of the de-contextualization approach are as follows (1) target construct is etic or universal or culturally invariant, and (2) a high-quality scale for the target construct is unavailable in the literature. This approach provides the opportunity to develop a universal measure for target construct and ease of scholarly exchange of research findings with the Western literature. However, it takes a much longer developmental time and a high developmental cost. Moreover, items developed under this approach tend to be phrased at the more abstract level, which may limit its informational and practical value.

Contrary to the de-contextualization approach, the contextualization approach assumes the target construct to be emic or culture specific and that high-quality emic scales for the target construct are unavailable in the literature. Contextualization provides the opportunity to develop scales that are highly relevant in the Chinese context and can contribute context-specific knowledge to Chinese management. Similar to the de-contextualization approach, it takes longer developmental time and higher developmental costs. New scales developed under this approach have limited generalizability. Moreover, using this approach makes communicating research findings with the Western literature difficult.

As majority of management researchers in Taiwan are Chinese and identify with the Chinese culture, Farh et al.’s classification provides a very good framework by which to observe and categorize scale development approaches in Taiwan’s organizational research.

2.4 Guidelines of scale development and validation in the cross-cultural context
Much has been said in the literature about the challenges of and methods for cross-cultural measurement development. Guidelines have been offered by professional organizations (e.g., AERA, APA, NCME, 1999; ITC, 2010; SIOP, 2003), and important articles and books have been published (e.g., Hambleton, Merenda and Spielberger, 2005; Harkness, Villar and Edwards, 2010; Van de Vijver and Leung 1997).

Scale development should always begin by an evaluation of the extent to which the constructs a scale is designed to measure are relevant and important to the intended use of the scale. Therefore, procedures should be taken to ensure construct validity in a newly developed scale. Evidence of construct validity reveals the extent to which a scale measures what it is purported to measure (Hinkin, 1998). The Standards for Educational and Psychological Testing, collaborated by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), has proposed many criteria to assess the psychometric soundness of a measurement scale. Appropriate construct validation should include a demonstration of content validity, criterion-related validity, and internal consistency (AERA, APA and NCME, 1999).

Hinkin (1998) proposes a series of assessments to ensure construct validity of a measurement scale. To assess content validity, he presents three procedures that involve matching and classifying measurement items to their definitions. Once the content validity has been ensured, exploratory factor analysis (EFA), internal consistency reliability test, and confirmatory factor analysis (CFA) should be performed to further refine the new scales, to ensure that the sampling domain has been captured adequately, and to quantitatively assess the quality of the factor structure. Obtaining more evidence of construct validity can be accomplished by examining convergent validity, discriminant validity, and criterion-related validity. Hinkin (1998) defines convergent validity as the extent to which the scales correlate with other measures of similar constructs, discriminant validity as the extent to which the scales do not correlate with dissimilar measures, and criterion-related validity as the relationships with other variables with which the measures are expected to correlate.

Many well-established measurement scales are used by other studies. According to Farh et al. (2006), there are two types of use in the cross-cultural context: translation and adaptation. When the construct being measured is etic in nature, direct translation is performed. The quality of the translation is often assessed by back-translation despite many criticisms on its use (e.g., Harkness, 2003; Harkness, Pennell and Schoua-Glusberg, 2004). Adaptation refers to the deliberate modification of a measure to create a new one. To better fit the needs of a new population, language, or condition, or any combination of these, the scales often go through various degrees of adaptation before they are administered. Adaptation can be made to the content, response options, or the format of any part of an instrument (Harkness, et al., 2010).

The International Test Commission (ITC) established a detailed set of guidelines for adapting psychological and educational tests for use in various different linguistic and cultural contexts (Van de Vijver and Hambleton, 1996), as the same tests are used in more and more countries and as tests developed in one country are frequently translated or adapted for use in another (ITC, 2012). Although the guidelines do not offer a step-by-step procedure for adapting the tests, they do call for the need to repeat the construct validation procedure as outlined in Hinkin (1998). For example, guideline C2 specifies that “the amount of overlap in the construct measured by the test or instrument in the populations of interest should be assessed” (ITC, 2010, p. 2). Guideline D4 states that “test developers/publishers should provide evidence that item content and stimulus materials are familiar to all intended populations” (ITC, 2010, p. 2). To comply with these guidelines, content validity procedures should be performed on the new test sample. Moreover, guideline D8, which states that “test
developers/publishers should provide information on the evaluation of validity in all target populations for whom the adapted versions are intended” (ITC, 2010, p. 2) implies the need to report evidence of validity check when the tests are used in a different population.

Recognizing the influence of cultural context within which a test is to be used and its serious threat to measurement equivalence, ITC places special emphasis on ensuring and reporting measurement equivalence in several of its guidelines. For instance, guideline D5 specifies that “test developers/publishers should implement systematic judgmental evidence, both linguistic and psychological, to improve the accuracy of the adaptation process and compile evidence on the equivalence of all language versions” (ITC, 2010, p. 2). The same view has been advocated by many scholars when discussing cross-cultural comparative research (e.g., Cheung and Rensvold, 1999, 2002; Robert, et al., 2006; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000).

In sum, establishing the validity and the reliability of measures used in any scientific research is very important. When these measures are used in a different research context, especially when administered to a culturally different sample, the evidence of validity and reliability will need to be examined again. If a cross-cultural comparison is intended in a study, aside from the reporting of cross-validation result, evidence of measurement equivalence should be presented before proceeding to any substantive analysis.

3. Methodology

This study selected the seven most important and representative management journals in Taiwan (including six TSSCI journals), namely, Asia Pacific Management Review, NTU Management Review, Sun Yat-Sen Management Review, Journal of Management, Chiao Da Management Review, Management Review, and Journal of Human Resource Management. The purpose is to examine the characteristics of the organizational management measurements used in the empirical studies published in these journals, specifically targeting the fields of organization behavior and human resource (OB/HR). We examined the articles published in these seven journals from 2004 to 2008. The criteria used to select the papers are the following: (1) empirical studies, (2) organizational management studies, and (3) studies that involve the use of a measurement scale to measure the research construct. All three criteria have to be met to be included in this review.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>2005</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>2006</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>2007</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>15</td>
<td>22</td>
<td>15</td>
<td>7</td>
<td>9</td>
<td>67</td>
</tr>
</tbody>
</table>

As shown in Table 2, this study included 146 papers, at a range of 16 to 39 papers each year. In these journals, the Journal of Human Resource Management, with its specific focus on human resource management field, has the most number of papers at 67 and is selected
more times than any other journal in this study. Chiao Da Management Review has the least number of papers that match the criteria of this study at 7.

After the sample was determined, a team of three graduate students went through a comprehensive review of each of these 146 articles using a coding table to record important information related to the purpose of this study. The coding table requires specific information such as basic reference data, field of study, study population and sample, instruments used, whether the instrument is a measurement scale, whether the measurement scale is developed for the study or adopted from existing scale, source of the measurement scale if adopted, name of the measurement scale, constructs being measured in the scale, development/adoption process, validation process, and evidence of validity and reliability of each scale.

Once the comprehensive review was completed, the resultant 146 review tables and their respective articles were sent to two doctoral students for cross-checking. Information that was coded incorrectly, incomplete, or in doubt was discussed and reconciled in the review meetings. To uncover the nature of the measurement scales used by these 146 studies, the research team went through another coding effort targeting each measurement scale used in these studies. Measurement scales were identified and attributes of each measurement scale were coded in an Excel spreadsheet using a pre-determined coding scheme. The coding task was divided among the two doctoral students, who also cross-checked the other person’s coding for accuracy. Again, review meetings were used to reconcile any differences in judgment.

The attributes of measurement scales that use a numerical coding scheme are research field, study sample, scale development approach, the use of back-translation, cited reason for scale modification, evidence of validity, and evidence of reliability. The research fields include the fields of OB, HR, and others. The study sample includes Taiwan, Taiwan plus others, and others. The scale development approach combined Farh et al.’s (2006) framework of translation, adaptation, de-contextualization, and contextualization, and three other categories uncovered in the course of this study: the adoption of a Taiwanese scale, the adaptation of a Taiwanese scale, and not specified. The use of backward translation was coded as yes, no, and not applicable. The cited reasons for scale modification include cultural differences, background/context differences, population differences, not stated, and not applicable. This study followed Hinkin’s (1998) suggestions on the indicators of validity and reliability. Evidence of validity includes seven categories, namely, EFA, CFA, content validity, criterion-referenced validity, convergent validity, discriminant validity, and others, with each dummy being coded as yes or no. Evidence of reliability includes Cronbach’s alpha and others, with each dummy being coded as yes or no.

4. Findings and discussion

This section is divided into three parts. The first part reports the results of our analysis on the most frequently studied constructs and the measurement scales used to measure these constructs in organizational management studies by Taiwanese researchers. The second part reveals the research orientation of organizational management scholars in Taiwan. The third part provides a detailed analysis of the attributes of measurement scales used in Taiwan’s organizational management studies.

4.1 Popular research constructs and their measures

We tabulated several of the most frequently used measures, and the result indicates that the measure of organizational commitment is the most frequently used in OB/HR studies. A total of 42 studies involved organizational commitment, 12 of these studies adopted Porter, Steers, Mowday, and Boulian’s (1974) organizational commitment questionnaires in their
studies, and 10 studies adopted Mowday, Steers, and Porter’s (1979) version. Other organizational commitment questionnaires such as those of Allen and Meyer (1990) (5 out of 42 studies) and Meyer, Allen, and Smith (1993) (6 out of 42 studies) are also commonly used in OB/HR studies. From the citation sources, we found several studies adopting the Mandarin version of organizational commitment scales translated by Taiwanese scholars from Porter et al. (1974).

The measures of job satisfaction and performance are frequently used in OB/HR studies as well. For job satisfaction measures, the result indicates that the Minnesota Satisfaction Questionnaire (MSQ) developed by Weiss, Dawis, England, and Lofquist (1967) is the most frequently used (7 out of 36 studies). Another equally frequent measure used in job satisfaction studies is the job satisfaction subscale of the Michigan Organizational Assessment Questionnaire (7 out of 36 studies) developed by Cammann, Fichman, Jenkins, and Klesh (1979). We also found some studies adopting job satisfaction scales developed by Taiwanese scholars.

A total of 43 studies involved the measures of job performance. These studies adopted scales that measure performance at different levels, such as team, individual, and organizational performance. Most studies (32 studies) focused on individual job performance. The scales of Kostova (1999) and Zahra and George (2002) are the most frequently used. Some studies claimed that they adopted questionnaires developed (or translated) from studies by Taiwanese scholars.

Organizational commitment, job satisfaction, and job performance seem to be the most studied constructs in Taiwan from 2004 to 2008. Although these are important issues, the fact that each of them occupies more than one quarter of the organizational management literature in Taiwan is an indication that other equally important issues have received disproportionate attention from the research community. The tabulation also reveals several popular measures such as those of Porter et al. (1974) and Mowday et al. (1979), initially confirming our assumption that the adoption of Western measures is a widespread phenomenon in Taiwan.

4.2 Research orientation of Taiwan organizational management scholars

After reviewing all the articles, we classified the results in a framework according to two dimensions: 1) research field (OB vs. HR) and 2) level of analysis (individual vs. team and organizations). Based on the two dimensions, we developed a framework to place the constructs reviewed in a typology.

Figure 1. Typology of the research orientation
1) Individual HR orientation: These constructs focus on employee perceptions toward HR regular functions (e.g., selection and recruitment, training, and retention). For example, organizational attractiveness and training effectiveness are individual HR-oriented constructs. A total of 88 (11.1%) constructs reviewed in this study can be classified in this category.

2) Team and organizational HR orientation: The constructs with team and organizational HR orientation focus on human resource management practices and organizational-level performance. HRM practices, HRM practice system, and organizational performance are constructs classified as having team and organizational HR orientation. A total of 83 (10.5%) constructs reviewed in this study can be classified in this category.

3) Individual OB orientation: The individual OB oriented constructs focus on individual attitudes, behaviors and outcomes. For example, job satisfaction, organizational commitment, individual performance, and OCB are individual OB-oriented constructs. Majority of the constructs reviewed in this study (i.e., 539 or 68.1%) belong to this category.

4) Team and organizational OB orientation: These constructs focus on team- or organizational-level contextual conditions. For example, organizational climate, culture, and perception of justice are classified as team and organizational OB-oriented constructs. A total of 30 (3.8%) constructs reviewed can be classified in this quadrant.

The result of this analysis has two important implications for Taiwan's organizational research. First, the fact that most of the constructs in our investigation fall into one category (i.e., individual OB orientation, 68.1%) indicates that other areas of organizational research are understudied. Speculating the reasons for the lower percentage of team/organizational-level research is not difficult because of the challenge to elicit institutional participation. This assertion is especially true in Taiwan, where corporations are more reluctant to disclose information. However, it is puzzling to observe that individual HR-oriented constructs only receive 11.1% of the attention from the research community, an indication of the lack of interest in studying HR policies and processes from the employee perspective.

Second, most of the constructs studied in the individual OB orientation category require attitudinal and behavioral responses and are deeply influenced by culture (Oyserman and Lee, 2008). Therefore, the application of an existing measure from another culture may lack validity in Taiwanese samples. Let us take job satisfaction as an example. An employee in a relation-orientated or collectivistic culture such as Taiwan may put more weight on the importance of social satisfaction than an employee in an individualistic culture. Therefore, a job satisfaction measure lacking social satisfaction items may create some bias on evaluating employee job satisfaction. This assumption raises the importance of paying attention to the issue of ME/I in organizational management research.

4.3 Attributes of measurement scales

As shown in Table 3, our study uncovered 791 research constructs in the selected sample, of which 547 (69.2%) can be categorized as OB constructs and 179 (22.6%) as HR constructs. A total of 65 (8.2%) constructs are multi-disciplinary, combining OB/HR constructs with constructs from other research fields such as marketing and operation, among others. Clearly, OB remains the largest and most preferred research field among Taiwan’s organizational management scholars.
Table 3. Attributes of the measurement scales in Taiwan’s organizational research

<table>
<thead>
<tr>
<th>Field of Constructs Studied</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>OB</td>
<td>547</td>
<td>69.2</td>
</tr>
<tr>
<td>HR</td>
<td>179</td>
<td>22.6</td>
</tr>
<tr>
<td>Others</td>
<td>65</td>
<td>8.2</td>
</tr>
<tr>
<td>Total</td>
<td>791</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sources of Samples Studied by Taiwanese Organizational Management Researchers</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan only</td>
<td>725</td>
<td>91.7</td>
</tr>
<tr>
<td>Taiwan + others</td>
<td>14</td>
<td>1.8</td>
</tr>
<tr>
<td>Other countries</td>
<td>52</td>
<td>6.6</td>
</tr>
<tr>
<td>Total</td>
<td>791</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale Development Approaches</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>263</td>
<td>33.2</td>
</tr>
<tr>
<td>Adaptation</td>
<td>352</td>
<td>44.5</td>
</tr>
<tr>
<td>De-contextualization</td>
<td>59</td>
<td>7.5</td>
</tr>
<tr>
<td>Contextualization</td>
<td>36</td>
<td>4.6</td>
</tr>
<tr>
<td>Using local scale</td>
<td>46</td>
<td>5.8</td>
</tr>
<tr>
<td>Adapting local scale</td>
<td>31</td>
<td>3.9</td>
</tr>
<tr>
<td>Unspecified</td>
<td>4</td>
<td>.5</td>
</tr>
<tr>
<td>Total</td>
<td>791</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use of Backward Translation</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>41</td>
<td>5.2</td>
</tr>
<tr>
<td>no</td>
<td>567</td>
<td>71.7</td>
</tr>
<tr>
<td>N/A(no need)</td>
<td>183</td>
<td>23.1</td>
</tr>
<tr>
<td>Total</td>
<td>791</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reported Measurement Scale Revision Reasons</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural difference</td>
<td>8</td>
<td>1.0</td>
</tr>
<tr>
<td>Research context difference</td>
<td>61</td>
<td>7.7</td>
</tr>
<tr>
<td>Research sample difference</td>
<td>96</td>
<td>12.1</td>
</tr>
<tr>
<td>Unspecified</td>
<td>392</td>
<td>49.6</td>
</tr>
<tr>
<td>No revision needed</td>
<td>234</td>
<td>29.6</td>
</tr>
<tr>
<td>Total</td>
<td>791</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The research samples in the studies we selected include both Taiwanese and subjects from other regions. As indicated by the analysis results in Table 3, most constructs in the studies used samples from Taiwan only (i.e., 725 or 91.7%), and 14 (1.8%) constructs used samples from both Taiwan and other regions. A small number of constructs (i.e., 52 or 6.6%) collected data from samples living in regions outside of Taiwan such as Hong Kong and Mainland China. Therefore, we can conclude that the OB/HR researchers in Taiwan conduct their research mainly on Taiwanese samples.

We classified the study measures into seven categories: translation, adaptation, de-contextualization, contextualization, adoption of a Taiwanese scale, adaptation of a Taiwanese
scale, and not specified. The frequency analysis results in Table 3 show that, of the 791 constructs, 263 (33.2%) constructs were translated directly from an existing foreign scale, 352 (44.5%) constructs were measured using an adaptation of an existing foreign scale, 59 (7.5%) constructs used a measurement scale developed for the study in a de-contextualization manner, 36 (4.6%) constructs used a measurement scale developed specifically for the study in a contextualization manner, 46 constructs (5.8%) were measured using an existing Taiwanese scale, 31 constructs (3.9%) were measured using an adaptation of an existing Taiwanese scale, and the source of measurement scale of 4 (.5%) constructs were left unspecified. This analysis shows that majority of the measurement scales (i.e., 615 or 77.7%) used by Taiwanese organizational management researchers in our sample are in the categories of translation and adaptation, both of which involve the use of an existing foreign scale.

Among the constructs that were a direct translation or an adaptation of an existing foreign measure, only 41 constructs (5.2%) were reported using the backward translation procedure, and 567 constructs (71.7%) were not reported using this procedure to ensure that the translation of a foreign scale to Chinese was accurate. Seemingly, most Taiwanese researchers still do not take the backward translation procedure as a necessary step to ensure the validity of their translated measures.

Among the 557 measures that were modified from an existing scale, 8 measures (1.0%) were modified because of cultural differences, 61 measures (7.7%) because of the differences in the study context, and 96 measures (12.1%) because of the differences in population characteristics; no reason for the modification was specified for the 392 (49.6%) measures. Most Taiwanese researchers left the reasons for the modification of an existing measure unstated in their papers.

Table 4 shows the current practice in validity and reliability reported in Taiwan’s organizational management research. As regards to the indicators of validity, of the 791 constructs, 330 reported using EFA as the validity indicator, 231 reported using CFA to ensure the validity, 71 reported content validity analysis, 11 reported criterion-referenced validity, 140 reported convergent validity, 144 reported discriminant validity, and 39 reported other validity indicators. Most of the constructs that reported validity used multiple validity indicators; for example, convergent validity usually goes with discriminant validity. However, 193 constructs were studied without reporting any evidence of validity.

As for the reliability indicators of measures, 699 out of the 791 constructs reported their internal consistency using Cronbach’s alpha as their reliability indicator, and 19 constructs
used other reliability indicators. A total of 79 constructs did not report any reliability analysis. Cronbach’s alpha is the most frequently reported reliability indicator.

Some problems underlying the measurement validation process in recent OB/HR studies in Taiwan are revealed in our analysis. First, OB/HR researchers in Taiwan have the tendency to adapt mostly the existing measurement scales developed in the West without presenting cross-validation evidence. Although they adapted scales that were well validated in previous studies, attempts to validate in the current research context are lacking. Clearly, the potential threats to validity and ME/I are not given much attention. Second, most OB/HR studies do not indicate clearly how they adapted the existing Western-developed scales. As the analysis indicated, only 263 out of 615 researched constructs reported using scale translation. Moreover, only 41 constructs clearly explained the translation process (e.g., back translation process); the others did not report how the scales were translated, an indication that the researchers might have performed the scale translation without checking the accuracy of the translated scales. Third, most researchers did some minor revisions on the adapted scales for their studies without explaining how or why they conducted the revision process. Based on our analysis, at least 50% of the studies did not explain the revision process, although they did report that they revised the scales as required by their study.

5. Conclusion and suggestions

This study attempts to understand the practice of scale development in Taiwan’s organizational research community and to examine the common practice of adopting Western measurement scales in Taiwan’s organizational research. Conclusions are drawn after reviewing 146 OB/HR studies published in Taiwan’s major journals from 2004 to 2008, and suggestions are offered to enhance the validity of Taiwan’s organizational research.

5.1 Common practices in adopting Western scales in Taiwan

We found that most of the published OB/HR studies in Taiwan adapted or modified the existing scales. According to our analysis, 615 out of 791 constructs studied in the 146 research adapted the existing scales, with 263 constructs using an existing translated version of a Western scale in the studies. Only 141 out of 791 constructs reported efforts of developing a new scale for use in their studies. The result confirms our previous assumption that scale development is not a popular approach in Taiwan’s OB/HR studies. Most researchers still prefer to adapt existing measures. This condition is consistent with that in the current academic society in Taiwan where researchers try to increase their chances of getting manuscripts accepted by international journals by using an existing scale to avoid being challenged in their validation process of a new research instrument. This phenomenon results in the lack of well-validated indigenous measurement scales in Taiwan.

In terms of reporting validity and reliability, there appears to be room for improvement for the organizational research community in Taiwan. Studies often jump to the findings and conclusions without presenting a sufficient discussion of the scale validation process. Most Taiwanese management researchers usually forgo describing the scale translation process and how translation quality was ensured, such as through backward translation. When scales were modified, the reasons and implications for the modification were usually not discussed. Although most studies (75.6%) reported the scale validity, less than half (41.7%) reported EFA analysis and about 30% reported CFA analysis. Moreover, the result indicates that most of the OB/HR studies did not perform more than one kind of validity test when they adapted Western scales for their studies. An alarming over-reliance on using Cronbach’s alpha (89% of studies) was apparent, which is the only evidence of reliability in the Taiwanese literature.

Another major finding of this study is the fact that the most frequently studied constructs in Taiwan’s organizational research are individual’s attitudinal and behavioral outcomes, for
instance, organizational commitment and job satisfaction. Moreover, the samples in most of these studies consisted of Taiwanese only (91.7%). This practice of applying measures developed from an individualistic culture such as the United States in a more collectivistic society such as Taiwan poses a serious threat to ME/I.

5.2 Suggestions for improving validity in Taiwan’s organizational research

We suggest future management studies on Taiwan to examine the validity of existing scales more fully before administering to local samples. More effort should be devoted to cross-validation to ensure measurement equivalence when a Western scale is first introduced to the Taiwan context. When such cross-validation has been done, later research can more confidently adopt the same measure citing the cross-validation result. Guidelines provided by professional organizations (e.g., AERA, et al., 1999; ITC, 2010) and organizational research scholars (e.g., Harkness, et al., 2010; Hinkin, 1998) should be followed when conducting studies using imported measures. Moreover, the use of an appropriate translation quality assessment is suggested to reduce threat to the validity and the measurement equivalence of the scale. Reports on the validity indicators will help readers evaluate the quality of a study.

Another suggestion can be made in the reporting of reliability. Reliability is the consistency of a measure. Note that different reliability indices are associated with different types of measurement errors. In certain type of studies, some sources of measurement errors are more likely to occur and cause more harmful effects on measurement outcomes than others. Depending on the measurement process and the intended uses of the measurement, some examples of the methods used to estimate reliability include test-retest reliability, internal consistency reliability, and parallel-test reliability (Schwab, 2005). Each method deals with the problem of finding the source of error in the test differently. Cronbach’s alpha is the most common estimate for internal consistency reliability, but it does not guarantee the other types of reliability. Therefore, the evaluation and reporting of reliability should be much more than a mindless check for Cronbach’s alpha and should include a discussion on the potential sources of measurement errors and their respective methods to estimate reliability in the study.

We also consider that an investigation of ME/I of the most frequently adopted Western measures is necessary to understand the extent of validity of past research results and to ensure the validity of future applications. As people live in the web woven with reference to their own cultural traditions (Geertz, 1973), their attitudinal and behavioral outcomes in organizations depend largely on the person’s cultural context (Oyserman and Lee, 2008). Taiwan is a collectivist culture with high power distance and uncertainty avoidance (Hofstede, 1980); therefore, the practice of applying attitudinal and behavioral measures developed in an individualistic culture such as the United States raises the issue of ME/I. Vandenberg and Lance (2000) suggest the sequence for conducting ME/I tests using structural equation modeling techniques that can be applied after data collection. Multi-group comparison procedures in structural equation modeling tools such as LISREL and AMOS can be used to test the ME/I of these instruments between Taiwanese samples and samples from the country of the scale’s origin such as the United States.

This study provides concrete evidence to the academic society in Taiwan that the current OB/HR studies do not pay sufficient attention to the potential threats to ME/I. It also shows that OB/HR scholars in Taiwan lack the awareness of a solid scale validation process or the reporting of it when they adopt existing Western scales. In sum, we would like to call for more attention on the issue of validity and measurement equivalence in management research. We strongly recommend journal editors to request authors of studies that applied existing measures from a foreign culture to report appropriate validation procedures to enhance the quality of management research around the globe.
5.3 Research contributions and limitations

In this paper, an ambitious effort to review and critique scale development practices in organizational research in Taiwan has been demonstrated. Based on extensive data collection and analysis, striking findings are reported and important questions are raised regarding the rigor of Taiwan’s organizational research. The outlining of the methodology and the results is a contribution of its own in highlighting the problems. Another contribution is the discussion and suggestions that are grounded on a comprehensive review of the relevant literature and up-to-date knowledge about methodological issues. This paper serves as a phenomenon-delineating piece with some normative prescriptions that can potentially contribute to the improvement of organizational research in Taiwan, East Asia, or other places.

This research has some limitations. First, to make the research scope manageable, we selected only a small portion of the vast number of publications available. Although a census approach was taken to identify the 146 OB/HR studies published in the seven selected journals from 2004 to 2008, a greater number of articles were produced by Taiwanese organizational researchers during this period, such as those published in SSCI-indexed journals and other indigenous publication not included in the TSSCI index. Second, majority of the articles reviewed were published in the Journal of Human Resource Management (46%). As the preferences of journal editors and their review panels could influence the type of research topics accepted and the methodologies used, some of the findings in this research could manifest a bias toward the journal with greater representation. Therefore, caution is advised in generating conclusions or judgments on the overall quality of Taiwan’s organizational research from this paper.

Acknowledgement

This research was funded by Taiwan’s National Science Council (NSC 97-2410-H-003-123-). We appreciate the support from the funding agency and the constructive inputs from the anonymous reviewers.

References