# Fear of Offending Others in Japan and Canada:
# A Mixed Methods Evaluation of the Taijin Kyofu Sho Questionnaire

Momoka Watanabe

Division of Social and Transcultural Psychiatry

Department of Psychiatry, McGill University

845 Sherbrooke Street West, Montreal, Quebec, Canada

December 2015

A thesis submitted to the Department of Psychiatry of McGill University in partial fulfilment of the requirements for the degree of Master of Science in Psychiatry

## TABLE OF CONTENTS

ABSTRACT

*Taijin Kyofusho* (TKS) has long been considered a Japanese culture-bound social anxiety disorder in the Western literature. One of the defining features of TKS is the fear of offending *others*, whereas the existing understanding of the "Western" form of social anxiety disorder (SAD) is the fear of embarrassing the *self*. TKS, however, does not appear to be as culturally specific as previously thought, and has been reported across different cultural contexts. The Taijin Kyofu Sho Questionnaire (TKSQ: Choy et al., 2008) was developed to assess symptoms of TKS in American and Korean patients with SAD. The scale, however, has not yet been validated in Japan nor in other Western cultural contexts. Cognitive processes and contextual factors shaping the experience of completing the TKSQ, as well as the item functioning and cross-cultural equivalence of the TKSQ, need to be examined in a comprehensive research design. This study examined the validity and psychometric properties of a recently developed measure of TKS symptoms, the 30-item Taijin Kyofu Sho Questionnaire (TKSQ: Choy et al., 2008), using a mixed-methods design. The specific objectives were: (1) to evaluate translation errors, questionnaire designs, and unique socio-cultural factors influencing the participants' responses on the TKSQ to better understand TKSQ across cultural contexts; and (2) to evaluate the cross-cultural equivalence of the TKSQ. Cognitive Interviewing (CI), a well-established qualitative method for evaluating survey questions was conducted with Japanese ($n$ = 12) and Euro-Canadian ($n$ = 9) university students with high levels of TKS symptoms. In the quantitative component of the study, two datasets were used to examine the validity of the scale. First, the TKSQ was administered to students in Japan ($n$ = 190) and Canada ($n$ = 78). Differential item functioning (DIF) analyses were conducted to detect items with item bias in TSKQ (Dataset 1). The external validation study was conducted in a separate dataset to examine correlation analysis among TKSQ, TK-S, SIAS, and SCS, also in Japan ($n$ = 124) and Canada ($n$ = 114). Analysis of CI results showed that the major source of error was related to clarity of the scale for the Euro-Canadian sample. Different interpretations of certain items by Japanese participants appeared to be due to translation errors. DIF analysis identified three items demonstrating DIF in Dataset 1, although the TKSQ was positively correlated with TK-S, SIAS, and Interdependent Self-Construal for both groups in Dataset 2. Both CI and quantitative findings have implications for studying cross-cultural differences in TKS-related symptoms and syndromes in Japan and

Canada. The novel research design used in this study could strengthen cross-cultural survey development and assessment, providing a more nuanced and integrated view of psychopathological phenomena across cultural contexts.

# RÉSUMÉ

Le Taijin Kyofusho (TKS) a longtemps été considéré, dans la littérature occidentale, comme un trouble d'anxiété sociale lié seulement à la culture Japonaise. Un des aspects distinctifs du TKS est la crainte d'insulter ou d'offenser les autres, tandis que la conceptualisation 'occidentale' prévalente du trouble d'anxiété sociale est la crainte de se mettre soi-même dans l'embarras. Cependant, le TKS ne semble pas être aussi intimement lié à la culture japonaise qu'on ne le pensait, et des cas ont déjà été rapportés au sein d'autres cultures. La recherche sur les mécanismes sous-jacents au TKS n'en étant qu'à ses débuts, le Taijin Kyofu Sho Questionnaire (TKSQ : Choy et al., 2008), créé au États-Unis, n'a pas encore été validé. En conséquence, les processus cognitifs et les facteurs contextuels qui façonnent l'expérience du TKS, ainsi que le fonctionnement et l'équivalence interculturelle du questionnaire TKSQ doivent être examinés dans le cadre d'une recherche exhaustive. Cette étude propose un modèle de recherche à méthodes mixtes qui intègre les méthodes qualitatives et quantitatives pour: (1) évaluer les erreurs de traduction, les modèles de questionnaires, et les différents facteurs socio-culturels ayant une influence sur les réponses des participants au questionnaire TKSQ, afin de mieux comprendre le phénomène du TKS au sein de diverses cultures; et (2) déterminer la validité statistique de l'équivalence interculturelle du questionnaire TKSQ. L'entrevue cognitive (Cognitive Interviewing, CI), une méthode de recherche qualitative reconnue et respectée, a été utilisée pour mener cette étude sur des étudiants Japonais ($n$ = 12) et Euro-Canadiens ($n$ = 9) à risque. Dans la phase quantitative de l'étude, deux ensembles de données on été utilisée pour examiner la validité de l'échelle. Tout d'abord, des étudiants au Japon ($n$ = 190) et au Canada ($n$ = 78) ont complété le TKSQ. Une analyse 'Differential Item Funtioning' (DIF) a été menée sur ces échantillons (Ensemble de données 1) dans le but de déceler les sections du questionnaire TKSQ (30 questions) comportant des items biaisés. Une analyse de validation externe fut ensuite menée sur un ensemble de données séparé, laquelle consista en une analyse corrélationnelle des relations entre les TKSQ, TK-S, SIAS, et SCS. Ce second ensemble de données comprenait des échantillons recueillis au Japon ( $n$ = 124) au Canada ($n$ = 114). Les résultats de l'analyse de l'entrevue cognitive démontrent que la principale source d'erreur est associée à la clarté de l'échelle pour l'échantillon d'Euro-Canadiens. Par ailleurs, les erreurs de traduction semblent être un des facteurs les plus problématiques pour les participants Japonais. L'analyse quantitative

a aussi identifié trois items faisant preuve de DIF dans l'ensemble de données 1, bien que le TKSQ soit corrélé positivement avec les échelles TK-S, SIAS, et Interdependent Self-Construal dans les deux groupes culturels dans l'ensemble de données 2. Les résultats obtenus indiquent que les deux types d'analyses, qualitative et DIF, sont liées par des disparités dans l'organisation des réponses des participants au questionnaire. De plus, les résultats ont également d'importantes implications en termes de différences interculturelles dans la compréhension du phénomène du TKS au Japon et au Canada. Pour conclure, dans les contextes interculturels, le modèle de recherche innovateur employé dans cette étude a le potentiel d'améliorer le développement des techniques de sondage et aussi d'offrir une vision plus complète et intégrée du phénomène à l'étude.

## Acknowledgements

There are a great number of people to whom I am grateful. First and foremost, I would like to thank all those who have willingly shared their precious time to act as participants during the process of survey data collection and interviewing in my study.

I would like to express my deepest gratitude to Dr. Laurence Kirmayer and Dr. Andrew Ryder for being my thesis supervisors as well as becoming the most respected mentors. Their immense knowledge, insights, and advice throughout my Master's studies helped me grow and excel as a researcher and as a person. I truly appreciate their excellent guidance, patience, and support in all the time of research and writing of this thesis. I would like to thank them for guiding my Master's thesis project and helping me to develop my background in culture and mental health research.

My sincere thanks also goes to Dr. Jun Sasaki, who served not only as a member of my thesis committee, but also has been a great international collaborator to help me collect data in Japan. I thank him for offering me the opportunity to recruit Japanese participants in Japan and helping to make this cross-cultural research project happen. I would also like to thank the members of Dr. Sasaki's lab at Osaka University as they provided me with continuous feedback and suggestions during the process of interviewing while I was in Japan.

I would also like to thank Keitaro Sasaki and Elizabeth Hébert who played an important role in the qualitative analysis phase of my project. A constant source of advice and clinical expertise, the time they devoted to code the qualitative data enabled me to understand my data from clinical and cross-cultural perspectives. I am also grateful for my fellow labmates, research assistants, and the members of Culture, Health, and Personality lab at Concordia University, who assisted me with participant recruitment, survey administration, data entry and management, and interview transcription and provided me with constructive feedback in all aspects of my research project.

My friends have helped me survive and stay motivated through these difficult years. Both their emotional and academic support helped me overcome setbacks and stay focused on my graduate study. Their friendship has been a source of strength all these years.

Most importantly, none of this would have been possible without the support and love of my family. My family was always there cheering me up and believing in me throughout this endeavor. Their continued support and love are the reason I am where I am today.

## Preface and Contribution of Author

As per McGill thesis regulations, I confirm that this thesis is an original work by Momoka Watanabe who was the primary investigator, responsible for all major areas of concept formulation, data collection, and analysis, as well as the majority of composition of this thesis. The research project, of which this thesis is a part, received research ethics approval from the Research Ethics Committee of Osaka University, Osaka, Japan and the Human Research Ethics Committee of Concordia University, Montreal, Canada. Some of the research conducted for this thesis forms part of an international research collaboration, led by Dr. Jun Sasaki at Osaka University with Dr. Andrew Ryder being the lead collaborator at Concordia University. Qualitative and quantitative methods and the data analysis plan in Chapter 3 were designed by myself. The data analysis in Chapter 4 is my original work, as well as the literature review in Chapter 2 and discussion in Chapter 5. The research described herein was conducted under the supervision of Drs. Laurence Kirmayer, Andrew Ryder, and Jun Sasaki who were involved throughout the project in conceptualization, oversight of data collection and analysis, and thesis editing.

## LIST OF TABLES

# CHAPTER 1: INTRODUCTION

## 1.1. Justification and Purpose

The study of psychopathology in social and cultural context has long been central to the field of cultural psychiatry. Given the need for culturally informed care of the minorities and immigrants, researchers in this field have endeavored to better understanding of sufferers of all backgrounds and thereby improve diagnosis and treatment. Despite greater acknowledgment of cultural issues in recent years, as found for example in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013), the traditional position in psychiatry has been dominated by a Western theoretical framework and nosology. The majority of existing theories in psychopathology are proposed by Westerners based on their experiences and data collected with people in Western societies (Tseng, Bartocci, Rovera, Infante, & De Luca).

Efforts to engage with cultural issues by psychopathology researchers have been complicated; moreover, by methodological debates regarding the relative status of qualitative and quantitative methods within the field and across its related subdisciplines (Henrich, Heine, & Norenzayan, 2010; Matsumoto & Yoo, 2006). The positions taken by researchers vary considerably; however, quantitative methods have dominated mainstream psychopathology research, whereas researchers concerned with local culture and ethnographic contexts have tended to rely on qualitative methods. Some researchers argue for a "best of both worlds" strategy and claim that one should combine quantitative and qualitative methods and engage in mixed-methods research (MMR) (Creswell & Tashakkori, 2007; Tashakkori & Creswell, 2008). A recent approach to understanding psychopathology in the light of culture posits that culture,

mind, and brain form a single system in which these three different levels mutually constitute one another, and therefore psychopathology is best understood through the integration of both quantitative and qualitative approaches (Doucerain, Vargas, & Ryder, 2015; Ryder, Ban, & Chentsova-Dutton, 2011). In this view, examining only one level of people's lives or from one perspective in paradigm is insufficient to understand a complete story of a person's psychopathology. Integrating qualitative and quantitative methods would allow researchers to examine the biological, psychological, and socio-cultural levels of a person's life and provide holistic and multidimensional explanations for the phenomena of the suffering.

From this perspective, the question of whether a specific psychological disorder is universal or culture-specific across racial, ethnic, and cultural groups is arguably outdated. Attention should shift to the question of how to investigate and identify which aspects of a disorder are universal and which aspects are culture-specific. The methodological flexibility of MMR provides researchers tools to deal with these kinds of questions. Furthermore, MMR maximizes the benefits of integrating scientific research and clinical practice, which in turn helps the researchers to gain in-depth understanding of psychopathology and sufferers in specific cultural contexts, as well as to generate hypotheses to test across cultural contexts for future research. Thus, MMR has a significant advantage in the study of culture and comparative designs for the investigation of a common psychological disorder and its symptoms.

Social anxiety is one example of a psychological disorder that is clearly shaped by the social world (Hofmann, Asnaani, & Hinton, 2010; Kirmayer, 1991), and there are well-known cultural variants. *Taijin Kyofusho* (TKS) is one of these culturally-defined social anxiety syndromes. In Japanese, the name refers to the syndrome (*sho*) of fear (*kyofu*) of interpersonal relations (*taijin*) (Dinnel, Kleinknecht, & Tanaka-Matsumi, 2002). TKS has long been

considered a Japanese culture-bound syndrome by the Western scholars (Hofmann & Hinton, 2014). In comparison to other forms of social anxiety, the main distinctive feature of TKS is a fear of adverse effects on *others* through one's performance and interpersonal relations. These concerns about the impact on *others* (in what might be termed an "allocentric focus") are, referred to as "offensive TKS symptoms" (Nagata et al., 2006; Takahashi, 1989). People with TKS have *other*-oriented or allocentric interpersonal focus of fears. They are afraid of doing something or presenting an appearance that will embarrass, offend, and cause discomfort to *others* (Hofmann et al., 2010).

In contrast, the "Western" diagnosis of social anxiety disorder (SAD) involves a fear of adverse effects on the *self* in social and interpersonal situations. People affected by SAD have anxiety pertaining to *self*-oriented or egocentric interpersonal focus. They are concerned about embarrassing and humiliating the *self* by making mistakes and looking bad in front of others (Hofmann et al., 2010; Kleinknecht, Dinnel, Kleinknecht, Hiruma, & Harada, 1997). Despite evidence for their importance in TKS, allocentric interpersonal fears were not included in the criteria for SAD in DSM-IV (American Psychiatric American Psychiatric Association, 2000). Recently, however, the fear of offending others was included as one of the symptom presentations of SAD in DSM-5 (American Psychiatric American Psychiatric Association, 2013).

In both TKS and SAD, sufferers tend to have an early age of onset, fear and avoidance of social situations, and they suffer from chronic distress and significant functional impairment. They exhibit anxiety centered on embarrassment or humiliation in social or performance situations due to a variety of specific fears in their perceived defects (e.g., eye-to-eye contact or staring, difficulty making small talk) and physical presentations of anxiety (e.g., blushing,

trembling, and sweating) (Choy, Schneier, Heimberg, Oh, & Liebowitz, 2008). People with offensive TKS symptoms are categorized with "offensive subtype TKS (OTKS)," and they are specifically afraid of displeasing, hurting, and offending *others* when they fear that others will notice their perceived shortcomings. While many common symptoms of TKS overlap with "traditional" or "Western" SAD, the allocentric fears many TKS sufferers endorse are considered a necessary criterion in the diagnosis of TKS (Nagata et al., 2006).

The term TKS was first introduced by Japanese psychiatrist Shoma Morita in the 1930s (Kasahara, 1987; Takahashi, 1989). Given the high prevalence reported in Japan, along with its "exotic" features, TKS has captured the interest of Western researchers and scholars for several decades. In much of this discussion, TKS has been characterized as a culture-bound syndrome that is specific to Japanese people. However, TKS does not appear to be as culturally specific as previously thought. TKS has been identified in other East Asian countries as well as in Western countries such as Canada, Spain, and the US (e.g., Clarvit, Schneier, & Liebowitz, 1996; Kitanishi, Lee, Choi, & Nakamura, 1999; McNally, Cassiday, & Calamari, 1990).

Moreover, the differences between the DSM-defined SAD and TKS are not especially clear-cut. Both involve many similar symptoms related to social anxiety, but they overlap incompletely. In a study comparing American students and Japanese students using the Social Phobia Scale and Social Interaction Anxiety Scale (Mattick & Clarke, 1998) to measure SAD and a revised *Taijin Kyofusho* Scale (Takahashi, 1989) to assess TKS, more than 50% of co-occurrence between SAD and TKS was found for both groups despite the factor analysis showing the discriminant validity of TKS (Kleinknecht et al., 1997; Mattick & Clarke, 1998; Takahashi, 1989). In another study, students with high levels of TKS symptoms (measured by the same TKS scale used in the aforementioned study) had relatively low scores on the

Liebowitz Social Anxiety Scale (Liebowitz, 1987) in Japan (Tarumi, Ichimiya, Yamada, Umesue, & Kuroki, 2004).

In sum, symptoms of offending others may also be present in SAD in Western cultural contexts. For example, a heterogeneous sample of Canadian patients with SAD reported significantly higher levels of fear of causing discomfort to others than did Canadian students or patients with other anxiety disorders but not SAD (Rector, Kocovski, & Ryder, 2006). Choy et al. (2008) reported that American patients with SAD endorsed a surprisingly high level of features of offensive subtype of TKS symptoms although they did not make a clinical diagnosis of the TKS syndrome in the sample. Studies suggesting the presence of these symptoms outside Japan challenge the cultural exclusivity of TKS. However, some researchers argue that TKS should be understood as a different psychological construct than SAD because TKS sufferers' primary concern is fear of causing distress and offence to others rather than fear of being scrutinized and evaluated, and the specific symptoms often focus on fears due to body odor, gaze, or appearance (Iwase et al., 2000; Kasahara, 1987; Nagata et al., 2006; Ono et al., 2001). For example, correlational analyses found a strong correlation between SAD and TKS for a clinical sample with SAD in Australia, but the prevalence of offensive symptoms in particular was low (8.5%), and none of the sample met full diagnostic criteria for TKS, supporting some distinctiveness between the two diagnoses (Kim, Rapee, & Gaston, 2008).

One of the most commonly used TKS measures, developed by Kleinknecht et al. (1997), is a 31-item self-report, 7-point Likert measure that assesses the extent to which respondents' perceptions reflect their concerns about offending or embarrassing others by doing something or showing an inappropriate appearance (e.g., 1 = exactly true, 7 = totally false). This scale only focuses on allocentric fears and does not allow assessment of offensive TKS symptoms

separately from the severity of fear associated with different interpersonal foci. To address this shortcoming, Choy et al. (2008) developed the TKSQ to assess the severity of offensive TKS symptoms and measure fears pertaining to both allocentric and egocentric interpersonal foci independently. The TKSQ also includes an intermediate option (i.e., fear of making someone uncomfortable) in respect to the severity of fear. The TKSQ has been used in research studies of TKS to systematically evaluate various symptoms of TKS as well as features of SAD and in the clinical assessment of TKS (Asakura et al., 2012). Despite the fact that the construct of TKS originates from Japan, the scale was first developed in English, in the United States, and then in South Korea, using clinical samples of psychiatric patients with a primary diagnosis of DSM-IV SAD. This raises questions about the similarities in social anxiety not only between North American and Japan but also between Japan and Korea.

While Japanese and Korean societies share broadly similar values, beliefs, and norms, a number of studies have documented cultural variations in social anxiety. For instance, Imamura and colleagues (1991) examined cultural variations in self-consciousness and behaviors among Japanese and Korean university students. Japanese participants were found to report significantly higher levels of TKS symptoms and lower levels of self-esteem and self-assertion than their Korean counterparts. The authors speculate that cultural variations in the views of the self and others may explain the differences found in their study, and further note that the patterns of self-esteem and self-assertion among Korean participants are more similar to the patterns found in Western cultural contexts. Specifically, their findings imply that Japanese are more sensitive to needs and expectation of others and are also less self-assertive, which lead to more susceptibility to social anxiety compared with Koreans. An empirical study examining the relationships among identity, public self-consciousness, and interdependent self-construal in the context of social

anxiety among Japanese and Korean university students found that Japanese students were more socially anxious than Korean students (Kim, 2005). In another cross-cultural study, Japanese university students also reported higher scores on interdependent self-concept, public self-awareness, and TKS symptoms than Koreans (Choi, 1993). These findings are an important reminder of cultural variations within different East Asian cultural contexts. The crude "East-West" dichotomy may ignore the distinctive symptom patterns and underlying mechanisms of TKS in Japan compared to Korea.

To summarize, the construct of TKS was first identified in Japan (Kasahara, 1987; Kirmayer, 1991; Ono et al., 2001). A variety of studies have shown that TKS-related symptoms occur among SAD patients in many cultural contexts. The TKSQ scale, which assesses the offensive-type and allocentric symptoms of TKS separately was developed in the United States and evaluated using American and Korean samples. It is therefore of interest to evaluate whether the symptoms included in the TKSQ have the same meaning in Japan and in North America, and to do so using an MMR approach. Doing so can help clarify which elements or aspects of TKS or perhaps questionnaire design influencing the observed differences across cultural contexts. In-depth understanding of symptom experience can help elucidate sociocultural determinants and mechanisms of social anxiety, which in turn may aid clinical assessment and treatment (Kirmayer, 1991; 2001).

## 1.2. Goals and Objectives

The overall objective of the present study is to develop an MMR approach to examining cross-cultural scale equivalence and apply it to a measure of TKS in Japan and Euro-Canadian samples. In addition to quantitative analysis of questionnaire response patterns, the study also examines respondents' discourse about symptoms and the unique social and cultural factors that

inform their responses. The MMR approach used here combines qualitative interview data for examining response processes and patterns and quantitative methods for analyzing statistical equivalence of the scale. In particular, Cognitive interviewing (CI) is utilized to qualitatively reveal non-clinical samples of Japanese and Euro-Canadians' experiences and interpretation of self-reported items on Taijin Kyofu Sho Questionnaire (TKSQ). Specifically, CI was used to: (1) examine generic problems of questionnaire that may influence all-groups; and (2) identify problems of cultural adaptation that may influence the cross-cultural comparability across Japanese and Euro-Canadian participants. Differential item functioning (DIF) analysis was conducted to statistically identify the survey items with biases across cultural groups. Incorporating the CI and DIF analyses in the investigation of TKSQ in two cultural groups can help reconcile some of the inconsistencies and limitations found in the literature. In addition to providing an in-depth understanding of the ways in which social and cultural factors shape the sources, experiences, and expressions of TKS, findings from the present study also illustrate the utility of MMR for cross-cultural research.

# CHAPTER 2: LITERATURE REVIEW

## 2.1. Introduction to *Taijin Kyofusho* (TKS)

### 2.1.1. *Taijin Kyofusho* in Japan

TKS is a social anxiety disorder common in Japan and generally believed to be at least somewhat specific to Japanese cultural contexts. The disorder is characterized by a pattern of symptoms marked by an intense fear of social and interpersonal situations. Specific symptoms of TKS include fear of perceived physical deficits (e.g., imagined ugliness) and physical manifestations of anxiety (e.g., blushing, trembling, emitting inappropriate body odors). People suffering from TKS often avoid interpersonal interactions, avoiding or withdrawing from social situations due to an excessive fear of being embarrassed, humiliated, and of offending others by their physical shortcomings or inappropriate social behavior.

The term TKS was first introduced by Japanese psychiatrist Shoma Morita in the 1930s (Kasahara, 1987; Takahashi, 1989). TKS has been a psychiatric diagnostic category in Japan for nearly 90 years. Social phobia or SAD, in contrast, was only included in the third edition of DSM in 1980. TKS was originally defined as a condition with a wide spectrum of illness severity levels, and included four distinct characteristics: (1) a history of transient adolescent social concerns; (2) neurotic symptoms (most typical); (3) delusional thinking; and (4) phobia, accompanied by schizophrenia, arising as a prodrome of schizophrenia or as a postpsychotic syndrome (Kasahara, 1987). The disorder has been studied extensively in Japan since its introduction; however, formal criteria for TKS were not established until 1995 (Maeda & Nathan, 1999).

The official Japanese diagnostic system describes four TKS subtypes, depending on the source of the sufferer's fear of offending, displeasing, and embarrassing *others*: *sekimen-kyofu* (the fear of blushing), *shubo-kyofu* (the fear of a deformed body), *jikoshisen-kyofu* (the fear of eye-to-eye contact), and *jikoshu-kyofu* (the fear of one's own foul body odor). Of these subtypes, *sekimen-kyofu* is most closely related to the traditional social anxiety disorder (SAD) defined in DSM-IV. *Shubo-kyofu*, on the other hand, is most closely linked to body dysmorphic disorder. While the features of TKS resemble that of SAD, TKS is understood as a broader concept in Japan. Researchers have pointed out because some TKS symptom presentations are similar to symptoms of other disorders such as paranoid personality, delusional disorder, hypochondriasis, body dysmorphic disorder, and even obsessive-compulsive disorder (Kirmayer, 1991; Kirmayer, Young, & Hayton, 1995; Nakamura, Kitanishi, Miyake, Hashimoto, & Kubota, 2002; Yamashita, 1970).

By definition, TKS refers to a psychiatric disorder, and a vast majority of TKS research and theory used to understand TKS have been based on studies of participants in psychiatric or medical settings. To date, true community-wide prevalence of TKS in Japan is unknown; however, clinical studies have estimated clinical prevalence rates ranging from 7.8% of patients with neurosis seen in a university general psychiatric outpatient clinic to 45.5% of patients with neurosis seeking TKS-specific treatment at a clinic offering Morita therapy (Ono et al., 2001; Takahashi, 1989). Researchers speculate that this wide range of prevalence rates may be due to the differences in sampling and versions of diagnostic criteria used in the studies (Aizawa, 2014; Hosoya, 2010). People who present with clinical symptoms of TKS are predominantly male. The male-female ratio is approximately 3:2, although the percentage of female patients in Japan is rising. Onset of TKS typically occurs during adolescence, affecting the sufferer's personal

growth and development as they have difficulty with the formulation of interpersonal relationships (Stein, 2009; Takahashi, 1989).

### 2.1.2. TKS in Current Research (Culture-Bound vs. Cross-Cultural)

In the international psychiatric literature, TKS has been commonly known as a Japanese culture-bound disorder. TKS attracted researchers in the West due to its high prevalence rates in Japan and its "exotic" features. A culture-bound syndrome is characterized by a set of specific behavioral, affective, and cognitive manifestations observable *only* within a certain culture or society. Culture-bound syndromes are generally limited to specific societies or culture areas, in which certain patterns of troubling experiences are deviant from the usual behavior of the people of that culture and perceived as distress. The conceptualization of the term *culture-bound syndrome* dates back to Yap's research in Hong Kong in the 1950-60s (Yap, 1965). Culture-bound syndromes were then incorporated in the Fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*), which provided a glossary of the selected culture-bound conditions mentioned elsewhere in the manual (4th ed.; DSM-IV; American Psychiatric Association, 1994). The DSM-IV-TR also included culture-bound syndromes as glossary terms for use with culturally and racially diverse patients; the list was still insufficient and limited when used for the assessment and diagnosis of patients with diverse backgrounds.

Furthermore, a number of culture-bound syndromes have been increasingly identified across different cultural groups. For instance, anorexia nervosa and bulimia nervosa were once considered Western culture-bound syndromes, in which a culturally constructed form of fat phobia resulted in pathology in Western countries (Lee, 1996). However, more recent studies have reported that both anorexia nervosa and bulimia nervosa have been observed in every non-Western part of the world as cultural values, norms, and beliefs as well as global influences have

changed eating behavior and body image over the years (Keel & Klump, 2003). Many of the syndromes previously thought to be culture-bound have now been found to be non-culture specific, although they could still be closely link to a particular cultural context if they demonstrate differential prevalence or nuanced symptom variations not identified in existing diagnostic systems. Cooper (2010) suggests that some culture-bound syndromes may be variants of a universally occurring disorder, others may be considered completely distinct disorders, and still others may be treated as being both as they form a distinct subtype of a more universally occurring category of disorder. He further notes, in any case, the existing diagnostic criteria should avoid the ethnocentric assumption that the Western variants disorders are "purer" and "less-culture-bound" than those that occur in non-Western cultural contexts. Some disorders may be the "Western" form of a disorder that should be seen as a mere variant of a more universal category.

The assumption of the cultural specificity of TKS is no exception and has been called into question. Suzuki et al. (2003) argue that TKS has been labeled as culture-bound because of an erroneous introduction of the concept of TKS to the West. TKS has been known as a Japanese culture-bound syndrome since the 1960s (Prince & Tcheng-Laroche, 1987). However, studies have identified the symptoms in different countries such as Korea, Indonesia, Egypt, Spain, Switzerland, and United States (Choy et al., 2008; Elkholy, 2012; Kitanishi et al., 1999; Vriends, Pfaltz, Novianti, & Hadiyono, 2013). More recently, Choy et al. (2008) conducted the first systematic assessment of an offensive subtype of TKS and found a strikingly high level of symptoms associated with TKS in both Korean and American patients with SAD.

Recent empirical studies have shown that underlying mechanisms and processes such as cultural models and self-concept may account for the manifestations of TKS. Group membership

(i.e., Japanese) alone does not explain the different prevalence rates of TKS found in different cultural contexts. For example, modes of self-construal may play a role in TKS symptoms. People with an interdependent self-construal are motivated to maintain culturally appropriate behaviors, thoughts, and feelings for the well-being of others (Markus & Kitayama, 1991). When Japanese people, who are typically considered interdependent, perceive that they have failed to act appropriately, they tend to express the fears of the anticipated adverse effects on *others* rather than on the *self* because the maintenance of social harmony is the primary goal that they should strive for in the society. Furthermore, research on the association between the self-concept and TKS has shown that interestingly while TKS symptoms are more common in Japan, such symptoms are more commonly reported among Euro-Americans who consider themselves highly interdependent (Dinnel et al., 2002).

At the same time, the interdependent self-construal alone was not found to be a strong predictor of TKS symptoms among Japanese university student (Dinnel et al., 2002; Imura et al., 1991; Kim, 2005). Relatedly, in one study, it was demonstrated that differences in a specific cognitive tendency that was apparent in a particular cultural context accounted the increased level of attention focused on others and fear related to TKS when comparing Japanese and Americans (Norasakkunkit, Kitayama, & Uchida, 2012). Once individuals in the Western cultural context selectively attend outward to detect negative social threat, they quickly switch their attention away from social cues and start monitoring and evaluating the *self*. Such cognitive process may explain tendencies of SAD or Western social anxiety. On the other hand, Japanese tend to use holistic cognitive process to pay more attention to the environment than the *self*, resulting in a greater emphasis on negative social cues and *other*-focused evaluation. Therefore, using holistic cognition, other-focused fear and TKS may be more culturally convergent in

Japanese cultural context. Norasakkunkit, Kitayama, and Uchida (2012) then demonstrated that although SAD and TKS tendencies appeared to co-occur among the same individual in both cultural groups, Japanese participants scored higher on TKS and holistic cognition compared to Americans while American participants scored higher on SAD. This study suggests that TKS tendencies may be associated with holistic cognition, whereas SAD tendencies may be associated with decreased levels of holistic cognition. Another study examined the frequency of TKS and SAD and its relation to the role of family environment among young adults in Japan and England (Essau, Sasagawa, Chen, & Sakano, 2012). Consistent with previous findings, participants in Japan scored significantly higher on both TKS and SAD than in England. Family sociability had a consistent effect on both TKS and SAD across the groups, but parental rearing attitudes, specifically, parental care, and overprotection had showed distinct patterns between the two cultural contexts. Hence, the function of family on TKS and SAD may differ depending on cultural background. The study failed to find concrete evidence that TKS has a distinct etiological background but has an implication that TKS may be a variant of SAD given that some familial factors are universal or predictive across cultural contexts, and others may be more culturally specific. These studies suggest that socio-contextual factors beyond membership in a particular cultural group, therefore, likely account for an important amount of variance in the types of symptoms associated with TKS.

Even the relative absence of a non-Western defined condition in the West should not instantly label the condition as culture-bound (Balhara & Goel, 2011). Furthermore, as Cooper (2010) argues, the fact that particular conditions occur in particular cultural contexts does not necessarily mean absence in other contexts. These conditions may be more difficult to observe than those that appear frequently or universally because measurement tools developed from a

universal standpoint may not capture culturally specific aspects of the conditions, where meaning, experience, and manifestation of the conditions may be different. To date, there are no comprehensive data on the prevalence of TKS in Japan and other countries. Cross-national comparisons of TKS prevalence rates have been problematic due to differences in language, measurements, and political, geographic, and complex sociodemographic contexts. Nevertheless, studies have shown that TKS symptoms are observed outside Japan. This suggests that TKS may be a variant of a universally occurring disorder. Therefore, labeling TKS a culture-bound syndrome assuming it is limited to be found in Japanese culture or with Japanese people may not be appropriate.

### 2.1.3. TKS in DSM-5 and Offensive Subtype of *Taijin Kyofusho* (OTKS)

Much of the literature concerned with culture-bound syndromes maintains a culturally relative position that syndromes (and explanatory theories) in all cultural contexts are distinct and valid in their own contexts. In this view, the manifestation of culture-bound syndromes is dependent on the interactions of numerous factors including cultural models, religious beliefs, social norms, economic resources, educational, political, and geographical environments, and so forth (Cole, 2013). Thus, cultural psychologists and psychiatrists are concerned with unpacking the culture-specific levels of psychological phenomena from the universal processes expected across cultures. Many psychological processes and constructs studied in cross-cultural research have both universal and culture-specific aspects. For instance, Heine, Lehman, Markus, and Kitayama (1999) demonstrated how psychological phenomena presumed to be functional universals failed to replicate across cultural contexts. Self-enhancement through experience of success was believe to be a universally common psychological source of motivation by mainstream psychologists for many years. However, in their study, although both Japanese and

American participants were motivated to do well on given tasks, Japanese participants approached the tasks with a self-critical and self-improving orientation through experience of failure instead of self-enhancement through success. The findings reflect the cultural variations in the pathways by which individuals motivate themselves, even as the higher order concept of 'motivation' remains a universal.

In another study examining cultural variations in factors related to mental health outcomes, happiness was found to be universal and valued in across different cultural contexts; however, meanings of happiness, motivations underlying happiness, and predictors of happiness significantly differed between North American and Japanese cultural contexts (Uchida, Norasakkunkit, & Kitayama, 2004). In this study, while Euro-American participants defined and experienced happiness as personal achievement, people with an East Asian cultural background considered happiness as a realization of social harmony. The study suggests that happiness is universal and biologically determined to some extent, but it is not the direct and only outcome of physiological or neurological mechanisms. The meanings and practices in shaping happiness are situated and embedded in specific cultural contexts.

Turning to the study of psychopathology, Ryder et al. (2008) provided an example of cultural variations in the presentation of depressive symptoms in Chinese and North American cultural contexts. Cultural variations in thinking style (externally oriented thinking) contributed to qualitatively distinct symptom presentations (psychological vs. somatic), although both samples were psychiatric outpatients seeking help for depression. These studies endeavor to go beyond epidemiological descriptions of group differences as well as to overcome the limitations of applying Western nosological systems to other cultural contexts.

Given the need to consider the cultural variations reported across the world, DSM-5 represents a significant effort to further understand the relevant contextual factors contributing to the distress expressed by sufferers from diverse backgrounds. Instead of a simple list of culture-bound syndromes, DSM-5 aims to improve diagnosis and assessment by updating the criteria to incorporate cross-cultural variations in symptom presentations as well as providing more detailed and structured resources about cultural concepts of distress. In an attempt to have a new and broader understanding of cultural variations in psychiatric symptoms, DSM-5 made modifications to culturally determined criteria so that there would be more standardized diagnostic criteria that would work across different cultural contexts (American Psychiatric Association, 2013; Lewis-Fernández et al., 2010). One of the most remarkable changes made in light of cultural variations in the DSM-5 is the criteria for social anxiety disorder.

Previously, the DSM diagnostic criteria for social anxiety disorder were limited to the fear of embarrassing *oneself* when an individual performs in front of others. Social anxiety disorder is one of the most prevalent psychiatric disorders according to systematic epidemiological surveys in the general population in many countries (Hofmann et al., 2010; Hofmann & Hinton, 2014). However, studies have shown that the definition was not representative of the ways in which social anxiety is experienced and expressed by the individuals in different cultural contexts. Specifically, in non-Western contexts, social anxiety disorder may be considered more than just shyness or fear of embarrassing *oneself*. Reflecting this diversity, in DSM-5 the diagnostic criteria for social anxiety disorder now include the core feature of TKS, the fear of offending *others*. *Taijin Kyofusho* is also included as one of illustrative examples of cultural concepts of distress in DSM-5 (Hofmann & Hinton, 2014).

While several classifications of TKS exist, the most widely accepted approach classifies the syndrome into two types: the "neurotic" or "common" type and the "offensive" subtype (OTKS). The common type overlaps with social anxiety disorder (SAD) defined in DSM-IV (American Psychiatric Association, 2000). In both the common type of TKS and SAD, individuals suffer from an intense fear of embarrassment and humiliation of the *self* in social interactions due to their perceived physical shortcomings or manifestation of anxious feelings. They are often afraid of face blushing, preoccupied with having stiff or awkward facial expressions, or concerned about negative evaluation by others, and they tend to avoid or withdraw from social situations. OTKS symptoms are considered a more severe form of TKS, encompassing the delusional and borderline forms initially described by Kasahara (1987). OTKS is characterized by fear of embarrassing, offending, and even harming *others* as a result of perceived physical deficits and anxious bodily movements (Kirmayer et al., 1995 1995; Takahashi, 1989). Fears of face blushing, trembling, sweating and making eye contact are considered common to both the common type of TKS and SAD; fears of stiff facial expression, unpleasant body odor, inappropriate staring, intestinal gas, and physical appearance are believed to be specific to OTKS (Choy et al., 2008 Oh, & Liebowitz, 2008).

There has been considerable debate regarding the conceptualization of TKS. Two central questions concern whether TKS is (1) a Japanese culture-specific disorder, or (2) a distinctive form of social anxiety disorder (SAD). To clarify the relationship between features of TKS and the "traditional" or Western notion of SAD, the Taijin Kyofu Sho Questionnaire (TKSQ) was developed by Choy and colleagues (2008), using clinical samples of Korean and American patients with SAD. Because previous studies had not measured offensive TKS symptoms separately from allocentric interpersonal foci, they aimed to assess those features. While the

instrument has proven useful for the systematic assessment of OTKS, limitations of the scale include the lack of pre-existing psychometric evaluation and clinical validation. Furthermore, the scale was first developed in English and then translated into Korean. Although the scale was translated into Japanese in a previous study examining OTKS (Zhou et al., 2014), its psychometric validity and cross-cultural comparability have not been evaluated in Japanese respondents. Given the wide variation reported in the studies of TKS within East Asian samples, one should not assume, "one measure fits all." Most importantly, the psychological construct of TKS was first developed in Japan; however, Choy et al. did not include Japanese participants. The TKSQ, therefore, may have both cultural and methodological biases and requires an in-depth examination of responses of Japanese participants to the scale, which in turn will help researchers gain a better understanding of the phenomenology of TKS among across cultural contexts.

## 2.2. Approaches to the Psychological Study of Culture

### 2.2.1. Mixed-Methods Research (MMR)

Qualitative and quantitative research methods have often been regarded as independent or even mutually exclusive in the psychological sciences (van de Vijver & Chasiotis, 2010). Furthermore, qualitative research has been largely marginalized and dominated by quantitative research over the past five decades (Karasz & Singelis, 2009). Qualitative data are obtained through descriptive reports, transcripts, or any other non-numerical output to uncover multiple realities and truths, whereas quantitative data are generated by numerical information based on the premise that there is an objective truth. Qualitative and quantitative data involve two different types of data collection and analyses, which may give an impression of incompatibility to researchers in both fields. However, some researchers argue for the "best of both worlds"

strategy to enhance understanding of research questions being investigated. Miles and Huberman (1994) note, "The careful measurement, generalizable samples, experimental control, and statistical tools of good quantitative studies are precious assets. When they are combined with the up-close, deep, credible understanding of complex real-world contexts that characterize good qualitative studies, we have a very powerful mix" (p.g., 42).

Specifically, a number of contemporary researchers concerned with the psychological study of culture has advocated combining both methods (e.g., Doucerain et al., 2015; Norenzayan & Heine, 2005). Much cross-cultural psychological research is yet preoccupied with the premise that psychological theories developed in the west are fundamental universals across the world (Henrich et al., 2010). They argue that cultural studies *need* mixed methods research and good cultural research starts with an in-depth examination of context of the people being researched (van de Vijver & Chasiotis, 2010). Given the cultural variations and differences found in underlying common psychological phenomena, one needs to investigate which levels of psychological processes are universal and which levels are culture-specific. This new position suggests a need for more attention to study the concrete and specific contents of local social and cultural processes that shape psychological variables (Karasz & Singelis, 2009). In this view, the question of whether a specific psychological phenomenon is universal or culture-specific across racial, ethnic, and cultural groups is outdated. The focus has shifted to the question of *how* to investigate and identify which aspects of a psychological process are universal and which aspects are culture-specific.

The methodological flexibility of MMR provides researchers with tools to deal with this new kind of inquiry. The study of concrete meaning and individual experiences in light of cultural context is often exploratory and qualitative, but it is open to quantitative and empirical

scrutiny. The qualitative study then becomes a crucial part of the statistical investigation of the overlapping meaning of a construct across cultural contexts when cross-cultural similarities and differences in targeted constructs are examined. Studies of attitudes, behaviors, and cognitions in different cultural contexts often require various types of methods and evidence. In psychological studies of culture, methodological aspects and components function as a means to increase the validity of the studies. An attractive feature of MMR is that research questions are treated as the main force determining the subsequent choice of data-collection and analyses. It is often common in mainstream psychology that methods not only dictate the choice of research design but also which research questions are studied (van de Vijver & Chasiotis, 2010). Furthermore, MMR provides methodological flexibility in studies of both culture-specific aspects that are largely qualitative and culture-comparative aspects that are studied quantitatively. Qualitative methods demonstrate their strength in the context of new discoveries – from this point of view, qualitative and quantitative research methods are therefore not incompatible but rather complementary.

The use of MMR is particularly well suited for the study of psychopathology across cultural contexts. A more recent approach in understanding psychopathology in light of culture posits that culture, mind, and brain form a single system in which the three different levels mutually constitute each other, namely cultural-clinical psychology (Ryder et al., 2011). In this view, psychopathology is shaped by the inseparable interrelation of psychological function, biology, and socio-cultural factors. Theories and research in understand psychopathology and clinical practices in treatment of sufferers, therefore, should be undertaken from an interdisciplinary point of view, and psychopathology should be studied in a multi-method way.

From a practical and community-based intervention standpoint, researchers and professionals in the mental health fields have paid increasing attention to the benefits of MMR. They emphasize the importance of findings offered by MMR as to determine, "what works for whom, under what circumstances, and why" (Hohmann & Shear, 2002). MMR maximizes the contribution of scientific research and clinical practice, which in turn helps researchers to gain in-depth understanding of psychopathology and sufferers in their specific cultural context as well as to generate hypotheses to test within and across cultural contexts for future research.

Furthermore, many of the standard assessments used in psychopathology research and mental health services were developed in the West. When applied to non-Western region of the globe with no cultural adaptation, such assessment tools can present problems. Most mental health assessments begin with a questionnaire or standard survey design (Betancourt & Bolton, 2005). These measures are intended to assess psychopathology constructs in the affected population (across cultural contexts or in an indigenous cultural setting), but frequently use existing "Western" measures that have little or no relation to the non-Western setting of interest (Betancourt & Bolton, 2005; Betancourt et al., 2011). In this manner, constructs being investigated are often created outside the local cultural context. Betancourt et al. (2011) claim that researchers need to alter and modify the existing Western measures and ensure that the constructs measured in the measure match the local interpretations and terminology before they can be considered as an equivalent measurement in the new cultural context. Very few attempts are made to evaluate the validity of the constructs being assessed in cross-cultural mental health research, and standard evaluation process is often quantitative (Betancourt et al., 2011). Such a "standard" approach does not allow researchers to reveal issues and concepts that are important to the local or new cultural context because close-ended survey questions simply do not ask them.

Betancourt et al. (2011) further warn, "without careful attention to locally relevant indicators of distress, current 'standard approaches' to measurement can easily specify constructs that are culturally irrelevant or miss highly relevant symptoms that have cultural and contextual significance" (p. 315).

Qualitative methods are highly useful for addressing these issues as they help researchers generate hypothesis and provide a support for questions such as how and why by examining a construct in depth. For instance, researchers have argued for the use of qualitative methods as the starting point for developing quantitative mental health assessment measures that are appropriate for a particular cultural context (Betancourt & Bolton, 2005; Mollica et al., 2004). Quantitative methods, on the other hand, are useful for confirming hypothesis and answering classic statistical questions (e.g., prevalence of a given disorder in a cultural context, strength of correlations among disorders and symptoms).

In sum, psychopathology across cultural contexts is best understood using MMR. Relying solely on a single method and results from the investigation of a single element of the culture-mind-brain system hinder our understanding of etiology, manifestation, and interpretation of psychopathology. Combining both qualitative and quantitative methods in a complementary fashion can allows researchers to investigate the three levels in their respective methodology, provides a more insightful and accurate scientific knowledge about psychopathology, and enhance understanding of the complexity of the interaction. Thus, MMR has notable advantages in the study of culture and comparative designs in the investigation of a common psychological disorder and its symptoms.

The TKSQ scale in the present study was developed in the West using samples from cultural contexts (i.e., US and Korean) in which the construct of interest (i.e., TKS) had not

originated. In addition, as noted earlier, cultural variations exist in the mechanisms of TKS among Japanese, Korean, and North American cultural contexts. This raises issues of validity and cultural equivalence in measurement. MMR is the best suited approach in the current study because it aims to examine (1) a novel topic that has not been well established (i.e., offensive-type of *Taijin Kyofusho*); (2) a relatively recently-developed instrument (i.e., TKSQ); and (3) two different cultural groups (i.e., Japanese and Euro-Canadians). The remaining sections will present the current study to demonstrate the ways in which cognitive interviewing and differential functioning analyses within a framework of MMR is useful to study psychological constructs and scale equivalence across cultures.

### 2.2.2. Cognitive Interviewing (CI)

Cognitive interviewing (CI) is a qualitative method that allows researchers to critically evaluate the success of an information transfer based on survey materials. Willis (2004) provides a definition of CI:

> Cognitive interviewing is a psychologically-oriented method for empirically studying the way in which individuals mentally process and respond to survey questionnaires. Cognitive interviews can be conducted for the general purpose of enhancing our understanding of how respondents carry out the task of answering survey questions, for the purpose of pretesting questions and determining how they should be modified, prior to survey fielding, to make them more understandable or otherwise easier to answer (p. 106).

CI is a plausible approach for examining cross-cultural differences pertaining to survey design. As populations in many countries continue to diversify, researchers utilizing survey

measurements face an increasingly complex challenge in translating and establishing the cultural equivalence of their instruments.

Quantitative methods commonly used in cross-cultural research have been criticized, as researchers often fail to provide cross-culturally valid measures (Greenfield, 1997). How individuals make sense of their responses to survey questions can indeed impact the very differences observed in cross-cultural studies. Problems with a survey and possible threats to accuracy created from the survey may not be visible in the quantitative data. Consequently, such structural differences and errors may result in inaccurate statistics and an incorrect conclusion about a concept or construct of interest. Johnson (1998) contends that analysis of cross-cultural survey data becomes problematic because question items may not function similarly when a survey is translated or administered to a diverse group of people.

CI is specifically useful in identifying patters of interpretation and patterns of error (e.g., translation) and establishing the conceptual and psychometric equivalence of a survey (e.g., scales measuring psychological disorders) when used within and across sociocultural and political groups (Karasz & Singelis, 2009). This qualitative method also helps researchers gain insight into the explanation for divergent patterns across cultural groups. In the US, for example, CI has been used as a step to develop large population-based surveys such as the 2003 National Center Institute's Tobacco Use Supplement to the Current Population Survey (Willis et al., 2008) to ensure cross-cultural equivalence across different language versions. In other studies, researchers utilized CI to address culturally relevant aspects of surveys with racial and ethnic questions in the US (McKay & de la Puente, 1996). In a CI study conducted to interview individuals from poverty-sicken rural Mississippi and metropolitan Washington, DC regions demonstrated that respondents coming from the rural areas had little access to health care,

thereby resulting in misunderstanding of the survey questions assessing their various health conditions (Miller, 2002).

***Data collection***

Sample size and selection of sample characteristics in CI studies depend on the research goal and the complexity of the survey questions to be evaluated. Large sample sizes are not required because a small number of samples can still provide rich and detailed information. The suggested sample sizes in the survey research literature range from 12 to 50 cases (Sudman, 1983). General guidance calls for sample size in cognitive interviewing range between 3 and 15 (e.g., Willis, 1994, 2004). However, Willis (2004) notes that determining sample size for cognitive interviewing can be more difficult and complex than in the case of purely quantitative research. He lists several approaches to address sample size issues. First of all, he emphasizes, "*Do what you can*." This is a realist approach to qualitative research generally, where *any* qualitative testing is better than *no* testing. Researchers should simply conduct as many interviews as their resources allow and discover what they can.

The second approach is to focus on saturation. The goal of qualitative testing is to obtain meaningful results rather than setting an initial sample size requirement because sample size becomes merely an empirical matter in the sense that one continues to test until no additional meaningful results are generated (Willis, 2015). For instance, Blair and Conrad (2011) studied the impact of sample sizes in cognitive interviewing and demonstrated that problem production followed an increasing trend as more interviews were conducted; however, the emergence of question flaws did not abate even after 90 interviews. On the other hand, another study has reported that 12 interviews were sufficient for saturating the results when interviewing women in West Africa with the aim of evaluating an HIV Disability questionnaire (Guest, Bunce, &

Johnson, 2006). While there is no specific number of participants required in cognitive interviewing, Willis (2015) describes three factors influencing the saturation level: (1) the nature of the questionnaire evaluated (i.e., more interviews required to reach saturation for disparate topics than for well-defined topics); (2) participant variation (i.e., a homogenous group will reach saturation more quickly than a heterogeneous one); and (3) the definition of saturation (i.e., researchers' judgment in deciding when to "stop the presses").

The final means to determine sample size is the quantitative-statistical approach. In this approach, researchers conduct a relatively large number of interviews to include as many question variants as possible, as they would in the case of a purely experimental design in quantitative comparisons. A vital question to cognitive interviewers and researchers is: Which approach is better? An ideal answer may be that more is better; however, this is oftentimes unrealistic due to lack of money, resources, and time. Boeije (2010) points out that we do not seek statistical representation in qualitative research projects. In fact, projects assessing a large number of interviews often end up being underanalyzed, resulting in the failure to conduct an intensive analysis. Unlike quantitative research, researchers are unable to quickly aggregate the data points in qualitative testing. The determination of sample size should depend on researchers' objectives. Even a single individual can provide valuable insights into the major deficits in the question under investigation, as demonstrated empirically in a study conducted by Willis (2004).

Sample selection is normally aimed to include the diverse characteristics of respondents represented in the given survey sample. Willis (2004) argues that the best sample compositions should contain respondents with diverse characteristics so that researchers can identify a wide range of survey question problems. For instance, if the survey asks questions of both men and women, then the CI study should include both genders and perhaps sexual and gender minority

individuals. Nevertheless, researchers can employ purposive sampling methods to select respondents for a particular purpose. Respondents can be chosen from those who are more likely to reveal problems. For example, Ackermann and Blair (2006) claim that education levels may be related to problem presentations and expressions among respondents in CI studies. They argue that respondents with higher education are more likely to articulate and reveal the larger number of problems within a single interview session than those with less formal education. It is important for the researchers to use good judgment in selecting sample sizes and diverse sample compositions so that the interviews can reveal a wide range of problems and give a complete picture of a survey question's performance. Establishing an appropriate number of interviews in the quantitative sense has been an ongoing debate.

The scope of CI is substantially different from the kinds of standardized interviews normally conducted in qualitative research. The researchers conduct intensive face-to-face interviews and ask respondents questions specific to survey problems. Providing the interviewer with a set of anticipated verbal probes is a commonly used strategy, as this allows the interviewer to take on an active and interactive role to probe the respondents during the interview (Willis et al., 2005). Types of verbal probes in CI can vary, but probes typically used in CI studies are based on Tourangeau's four-stage cognitive process model (1984). For example, Willis (2004) provides a list of probing examples such as "What does this term mean to you?" (comprehension), "What were you thinking about while answering?" (memory retrieval), "How sure are you about your answer?" (judgment), and "Was it difficult or easy to select an answer from the options provided?" (response selection). These probes are designed to enable the researchers to capture the ways in which respondents' specific experiences and perceptions influence their answers to the survey question.

***Data Analysis***

CI can be understood as a qualitative approach to a study of construct validity because identifying various difficulties that respondents may experience as well as issues related to cross-cultural comparability are the central focus of CI (Miller, Chepp, Willson, & Padilla, 2014). The data consist of written notes and transcripts concerning the respondents' subjective reports and answers to targeted questions through verbal probing (e.g., comprehension, memory, translation). The analytic strategies used by CI researchers choose to use vary. The use of coding systems was a popular strategy when CI emerged in the 1980s and 1990s (Dean, Caspar, McAvinchey, Reed, & Quiroz, 2007). These early coding systems were developed according to Tourangeau's model of survey response processes that assumed that response errors would fall under the four categories of cognitive processes (Tourangeau, 1984). The Questionnaire Appraisal System (QAS), a questionnaire coding scheme was developed in 1991 using Tourangeau's model (Willis, Royston, & Bercini, 1991). QAS proposed seven specific problem areas to evaluate comprehension (e.g., "Subjects tend not to understand what we mean by vigorous exercise"), memory retrieval (e.g., "Subjects had problems remembering over 12 months"), judgment (e.g., "The question makes it sound like you should report a lot of hours of work each week"), and response selection (e.g., "The given categories don't match the answers that people normally use") (p.166).

The QAS provides questionnaire developers and evaluators with a system for systematic review and improvement of survey materials (Dean et al., 2007). Willis et al. (1991) applied QAS to conduct CI on surveys administered by the US National Center for Health Statistics. Similarly, Presser and Blair (1994) conducted CI and applied QAS to identify survey problems for National Opinion Research. The results of the two studies reported that the most common

response error appeared to be comprehension. 70.5% of problems were identified as comprehension errors in Willis et al.'s study and 58.1% of errors was due to the same domain in Presser and Blair's laboratory (Presser & Blair, 1994; Willis et al., 1991). In another study, 70% of problems belonged to the comprehension and memory retrieval areas (Rothgeb, Willis, & Forsyth, 2007). Comprehension problems constitute the major source of response error, and issues with response selection were the second most common type of response errors reported by all three studies (DeMaio & Landreth, 2004).

These early coding schemes did not take cross-cultural and multilingual validity into consideration, however. A more recent approach in the use of coding schemes is to adapt areas that assess potential cross-language and cross-cultural problems. For example, Dean et al. (2007) developed the QAS-04 to enable researchers to systematically evaluate a survey for issues that may occur in translation and cross-cultural application. QAS-04 incorporates 15 new areas to consider both cross-cultural (e.g., consider reporting measuring units in both the Imperial system and the metric system) and translation issues (e.g., many idioms do not have an equivalent across languages). The authors particularly recommend the use of QAS-04 for facilitating multicultural instrument development.

Although coding schemes provide researchers with a cost- and time-effective tool to analyze CI data, the method of CI analysis is rooted within the paradigm of qualitative methodology (Miller et al., 2014). Miler et al. (2014) argue that complete and comprehensive analyses are needed to identify various aspects of respondents' patterns including response difficulties, construct validity, and comparability. Furthermore, relying solely on response errors during analysis is incomplete because it only tells the researchers half the story about the validity of a survey question (Ridolfo & Schoua-Glusberg, 2011). In order to evaluate the validity of a

survey question or measurement, researchers must identify not only response error but the construct a survey question is intended to measure.

Miller (2009) also emphasizes that responding to a survey question is a complex cognitive process which is strongly influenced by respondents' unique context and lived experiences. One of the most important aspects of CI is the conceptualization of this method as an ethnographical exercise as opposed to strictly cognitive focused interviews (i.e., cognitive interviews used for eyewitness-related purposes in law enforcement). Qualitative data obtained through CI include respondents' beliefs, norms, and everyday practices (Willis, 2004). Researchers are provided with such invaluable information, allowing for an in-depth examination of the standpoints of respondents to understand the ways in which respondents put their own life experiences and narratives into responses. This is particularly crucial in studies examining culturally relevant psychological constructs across different cultural groups.

One analytic method found to be useful in cross-cultural studies is the constant comparative method (CCM). The CCM allows researchers to perform systematic coding and generate a theoretical framework from the raw data (Creswell, 2013). In this inductive analytic method, researchers undertake five steps: (1) conducting interviews; (2) reviewing data; (3) assigning codes to analytic categories to compare across respondents; (4) developing advanced thematic schema to compare across subgroups of respondents; and (5) reaching conclusions (Miller et al., 2014). Over the course of this process of comparing categories, themes, and subgroups, researchers are able to discern the patterns of responses. Ridolpho and Shoua-Glusberg (2011) demonstrated the utility of CCM by conducting cognitive interviews with English-speaking Americans and first-generation Hispanic immigrants in the US to ask about two questions regarding ethnicity and race in their native languages. They found systematic

differences in the ways in which respondents interpreted ethnicity and race and formulated their answers based on their own self-conceptions (e.g., experiences of discrimination, beliefs about social characteristics, personal identity). Furthermore, they identified response variations between two cultural groups in how they understood, mentally processed, and produced an answer to these questions, subsequently leading to different types of data across Hispanics and non-Hispanics.

In sum, the use of CI in cross-cultural studies can be a compelling and rigorous means to obtain a meaningful piece of information for assessing measurement equivalence. Furthermore, CI provides not only useful information for the survey question improvement and establishment of equivalent measures, but invaluable insight into the manner in which individuals' unique sociocultural factors influence response process across cultures and subpopulations (Willis & Miller, 2011). Regardless of these advantages of CI, it is by no means sufficient to evaluate an instrument. CI does not provide direct means to assess the variation in reliability or the instrument's ability to measure different levels of a construct of interest. CI also lacks the support for generalizability of the findings as it typically involves a small, purposive sampling instead of a large-scale random sampling. In order to evaluate the cross-cultural comparability of a scale in a comprehensive manner, one must go further using additional methods that are robust in precisely those areas in which CI is limited.

### 2.2.3. Differential Item Functioning (DIF)

Differential item functioning (DIF) refers to differences in the item functioning between groups (e.g., gender, ethnicity), which are matched on the latent trait (e.g., skills, intelligence) being tested by the set of items (Hambleton, 1991). DIF occurs when individuals from different groups with the same level of the measured characteristic, have a different probability of giving a

particular response to on a survey or questionnaire depending on the group membership. When DIF is detected, total-score differences exist due to possible item biases instead of true differences in the measured attribute.

DIF analysis is one of the quantitative methods well-suited for detecting biased items when testing different cultural groups or different language versions of tests and questionnaires (van de Vijver & Poortinga, 2005). For example, when testing the degree of equivalence between different language versions of a survey using this technique, the presence of DIF indicates unexpected behavior of items and biases unrelated to the intended construct, thereby threatening the validity of the inferences raised regarding the actual similarities and differences between groups. DIF analysis has developed as the statistical techniques to detect item biases have progressed (Benítez & Padilla, 2013). In DIF analysis, the item serves as the most fundamental level of content analysis of a measure being evaluated, and the presence of DIF provides the foundation for the statistical inferences about item bias. If DIF is not present for an item, then no item bias is apparent. However, even if DIF is detected, the presence of DIF does not guarantee the presence of item bias. DIF is necessary, but *not* sufficient to conclude that the item has bias; rather follow-up item bias analyses such as qualitative content analysis and empirical evaluation would be required to determine the presence of item bias as well as the causes of DIF (Zumbo, 1999).

While there is a wide variety of statistical procedures for the detection of DIF, two common approaches are statistical methods and judgment methods (Hidalgo & López-Pina, 2004; Zumbo, 2009). The statistical methods approach focuses on detecting items with DIF for a particular group or groups evaluated and determining which statistical methods identify DIF in a more efficient way, mainly by examining their power and the nominal significance level

(Schmeiser, 1982). In contrast, judgment methods rely solely on expert's review to select items with potential biases and categorize the items according to their structural characteristics (Zumbo, 1999).

For instance, the utility of DIF analysis was demonstrated in a study that was the first to evaluate the cross-cultural equivalence of the English version of the Rosenberg Self-Esteem scale among American, Canadian, and New Zealand university students (Rusticus, Hubley, & Zumbo, 2004). In this study, ordinal logistic regression was performed to identify items displaying significant DIF. Results showed that all items functioned equally across cultural groups, with the exception of one item showing moderate DIF. The findings from this study suggests the need for reassessing the psychometric comparability of the widely used scale even among a group of English-speaking countries. The assumption of "one-size-fits-all" was jeopardized even within the English-speaking Western contexts in which respondents share presumably similar cultural values and response patterns to a survey. Assuming that a test developed in the West is functioning similarly or equally for all cultural groups without providing empirical evidence may be more threatening and detrimental to the interpretation of the results because the observed differences or similarities between the groups could be due to biases irrelevant to the construct of interest.

Previously, the implementation of judgment methods has mainly emphasized the appraisal of expert or the investigator; however, experts' opinions do not reflect the respondents' response-answer processes. Understanding the respondents' narratives allows the investigators to detect group differences and locate the causes of DIF in groups (Benítez & Padilla, 2013). Thus, consideration of respondents' perspectives means an advance over the traditional judgment methods in the field.

In recent years, researchers have attempted to rethink the traditional views of validity in the field of testing and assessment (Zumbo, 1999). Zumbo (1999) argues that the process of traditional validation was limited to simply selecting the most convenient strategies (e.g., content, criterion, construct validity) and relying solely on common statistical techniques (e.g., correlations with the gold standard, factor analysis, convergent and discriminant validity). In contrast, the current view of validity posits that one should gather all the information and evidence possible. Reliability and construct validity are regarded as a necessary condition, and the definition of validity is no longer a set of statistical methodologies but an elaboration of theory and supporting methods. Hence, one needs to provide a more comprehensive validation process such as incorporating qualitative methods.

The goal of DIF analysis has to do with measurement validity; however, researchers are left with a question of how DIF plays a role in the network of validation. There is no direct link between the presence of DIF and fairness. Scheuneman (1981) notes, "The bias that has been found to exist in items, however, has not been clearly tied to bias in tests. In general, while the presence of biased items seems to decrease the likelihood that the test as a whole is unbiased, failure to detect item bias cannot conversely be considered as evidence that bias does not exist but only that, if bias exists the amount is much the same for all items" (p. 20). However, he emphasizes that regardless of this limitation, researchers can look for the relationship between DIF and validity by combining quantitative and qualitative methods. Examining the presence and causes of item bias using the qualitative technique can provide a confirmation and explanation for biased items and can enrich the connection between the DIF analysis and measurement validity.

# CAPTER 3: METHODS

## 3.1. Mixed-Methods Research Design

This study used a partially mixed, concurrent equal status design based on the MMR typology (Leech & Onwuegbuzie, 2009). According to this typology, there are three dimensions in MMR designs: (1) level of mixing (partially vs. fully); (2) time orientation (concurrent vs. sequential); and (3) emphasis of approach (equal vs. dominant). In partially mixed concurrent equal status MMR design, qualitative and quantitative data are collected and analyzed independently before mixing occurs (partially mixed), the two facets occur approximately at the same (concurrent), and both qualitative and quantitative components are given equal weight in addressing the research question (equal). First, the design in the present study was partially mixed because both CI (qualitative) and DIF (quantitative) data collection and analysis were carried out independently, each with its own methodology. Secondly, qualitative data collection and analysis were carried out in parallel and concurrently with the quantitative phase. Finally, each component had approximately equal status within the research overall. The CI analysis was not intended to provide the basis for the DIF analysis. Likewise, DIF analysis was not designed to inform the qualitative interview process, but to provide an integrated element of the interpretation of the shared research question within the analysis stage, just as though the data originated elsewhere. The findings from each phase were drawn together in the current study to address a shared research question, which was to evaluate the cross-cultural equivalence of the TKSQ.

The rationale for employing the concurrent mixed-methods design was to evaluate the scale functioning and cross-cultural equivalence of Taijin Kyofusho Questionnaire (TKSQ) as

well as to explore in-depth experiences and perceptions of the offensive subtype of *Taijin Kyofusho* (OTKS). The current study's methodological approach was to: (1) gather both qualitative and quantitative data concurrently and separately; (2) analyze the quantitative and qualitative data separately; and (3) integrate the quantitative and qualitative data to provide a better understanding of research inquiry by interpretation of the mixed data sets (Creswell, 2013).

The current study aimed to explore participants' understanding, experiences, and interpretation of *Taijin Kyofusho* among Japanese and Euro-Canadian participants by conducting (1) cognitive interviews to elucidate the patterns and processes considered in responding to a scale measuring the offensive subtype of *Taijin Kyofusho*, and (2) DIF analysis to evaluate cross-cultural equivalence of the Taijin Kyofu Sho Questionnaire (TKSQ) to identify items with biases. Specifically, the CI findings were anticipated to identify generic problems of the TKSQ that affect respondents' answering patterns and issues with translation and cultural adaptation that may impede cross-cultural comparability. DIF results were conducted in order to evaluate the psychometric properties of the scale and biased items that may impede the validity of the measure. The strengths and benefits of both data sources are complementary. In concurrent mixed-methods design, neither of the two datasets is designed to inform the collection or process of the other dataset. Integrating the two different datasets strengthens the methodology and allows for a more nuanced interpretation of the results.

### 3.2. The Taijin Kyofu Sho Questionnaire (TKSQ; Choy et al., 2008)

The *Taijin Kyofu Sho Questionnaire* (TKSQ; Choy et al., 2008) is a 30-item self-report measure that assesses the severity of fear with respect to 10 physical/behavioral symptom characteristics of TKS and social anxiety in each of three interpersonal foci: (1) embarrassing the self; (2) making someone uncomfortable; and (3) offending someone. Respondents are instructed

to rate each statement on a 4-point Likert scale as it applies to them (1 = I would not be fearful at all; 2 = I would be mildly fearful; 3 = I would be moderately fearful; 4 = I would be extremely fearful). Of the 10 TKS symptoms, five are categorized as offensive TKS symptoms (i.e., fears of stiffening facial expression, emitting body odor, staring, releasing intestinal gas, and physical appearance). The remaining five TKS symptoms are common features of both TKS and social anxiety disorder (i.e., fears of blushing, body trembling, voice trembling, sweating, and making eye contact), which are referred to as common symptoms. The scale therefore comprises a total of six subscales: (1) fear of embarrassment to self due to common symptoms (E-CS); (2) fear of discomfort to others due to common symptoms (D-CS); (3) fear of offense to others due to common symptoms (O-CS); (4) fear of embarrassment to self due to offensive TKS symptoms (E-TKS); (5) fear of discomfort to others due to offensive TKS symptoms (D-TKS); and (6) fear of offense to others due to offensive TKS symptoms (O-TKS). Each subscale is the sum of five items (e.g., E-CS is the sum of scores related to embarrassment for the five common symptoms). The scale thus yields six subscale scores and one total score. The list of items and subscales is presented in Table 1.

The TKSQ was developed first in English and Korean. The current study used a Japanese version of the scale, which was previously translated by Zhou et al. (2014) with the authors' permission. Zhou et al. (2014) translated the scale into Japanese by following the Translation, Review, Adjudication, Pretesting, and Documentation (TRAPID) translation procedure (Harkness, 2003). They formed a translation committee consisting of bilingual researchers and translation reviewers. Three translators worked on the translation of the entire scale individually and then reviewed each other's translation versions. Then, reviewers carefully examined all the items and identified items with problems. The committee met to discuss all items and problems

over the internet, holding multiple meetings to make modifications and improvement in the translation to ensure the accuracy of the translation. The study authors reviewed the translated version and made any necessary adjustments; then the translators reviewed this version independently and discussed any further concerns with the reviewers. Finally, the authors made additional modifications and adjustments and finalized the translation after obtaining the final agreement with all the translators and reviewers. Although the English and Japanese versions showed satisfactory reliabilities in their respective studies, the scales have not been validated.

## 3.3. Sampling and Data Collection

### 3.3.1. Quantitative Data

#### *Samples*

For the Japanese sample, humanities and social science students enrolled in psychology courses at Osaka University were recruited through course instructors and word-of-mouth to complete a set of screening questionnaires. Japanese participants were all born in Japan, self-identified as having Japanese heritage, and were from the greater Osaka Metropolitan area in Japan. Euro-Canadians attended an English-language university in Montreal, were born in Canada or moved to Canada before the age of six, attended school at an English-language school board for all levels of education, and self-identified as White/Caucasian with European ancestry. Both Osaka and Montreal are the second largest cities in each country (population sizes in Osaka and Montreal are 2.66 million and 1.66 million, respectively), and participants were attending a large public university in their respective regions. Participants at both sites were recruited using purposive sampling within a convenience sampling frame. Euro-Canadian respondents were recruited through the Psychology Research Participant Pool system at Concordia University in Montreal. Respondents from both sites filled out a paper-and-pencil survey packet. The survey

included informed consent, demographic information, and the TKSQ. In addition, on the packet, respondents were asked to indicate their willingness to be contacted for the future study, which involves interviewing for the qualitative part of the study. Quantitative analysis was conducted using two different datasets (Dataset 1 and Dataset 2). First, DIF analysis of TKSQ was conducted using Dataset 1, comprising the data collected at the screening phase of qualitative data collection described above. Dataset 1 was a sample of 190 Japanese (86 females, 94 males, mean age = 18.76, *SD* = 0.84) and 78 Euro-Canadian participants (9 females, 69 males, mean age = 22.32, *SD* = 4.85) who met the same criteria described in the qualitative data collection. Demographics of Japanese and Euro-Canadian samples are shown in Table 2. Second, additional external validity analysis was conducted using samples drawn from dataset 2, which was collected for a larger project carried out at the same sites. In this dataset, 124 Japanese university students (63 females, 53 males, 3 missing, mean age = 21.03, *SD* = 4.62 completed a paper-and-pencil version of a survey questionnaire packet. Canadian participants who met the criteria used in qualitative data collection were drawn from an ongoing larger research project in which students were recruited through the university participant pool system. A total of 114 Euro-Canadian participants (15 females, 97 males, 2 missing, mean age = 22.12, *SD* = 4.8) completed a survey questionnaire online. Demographics of Japanese and Euro-Canadian samples for Dataset 2 are shown in Table 3.

***Material***

The *Taijin Kyofu Sho Questionnaire* (TKSQ; Choy et al., 2008), as introduced in the qualitative data collection was, the subject of DIF analysis and external validity analysis.

The *Taijin Kyofusho Scale* (TK-S; Kleinknecht et al., 1997) is a 31-item questionnaire that assesses symptoms found to discriminate TKS patients in Japan from non-clinical samples.

The items reflect the definitional symptoms of TKS (Kirmayer, 1991; Takahashi, 1989). The scale asks about the degree to which respondents are concerned that they would do something to bring embarrassment or offense to others. Respondents rated each statement on a 7-point Likert scale as it applies to them (ranging from 1 = totally false to 7 = exactly true). An example item is "I am afraid that my presence will offend others."

The *Social Interaction Anxiety Scale* (SIAS; Mattick & Clarke, 1998) is a 20-item self-report measure that assesses social anxiety during interaction with others. The scale is widely used and designed to reflect SAD or the "Western" definition of social anxiety. Responses are made on a 5-point likert scale (ranging from 0 = Not at all characteristic or true of me to 4 = Extremely characteristic or true of me). A sample item from the scale is "I worry about expressing myself in case I appear awkward."

The *Self-Construal Scale* (SCS; Singelis, 1994) is a 30-item measure that assesses two dimensions of self-construal using a 7-point agreement Likert-type scale (ranging from 1 = Completely Disagree to 7 = Completely Agree). The SCS consists of two 15-item subscales: independent self-construal and interdependent self-construal. A sample independent self-construal item is "I try to do what is best for me, regardless of how that might affect others." An example from the interdependent self-construal item is "It is important for me to maintain harmony within my group."

### 3.3.2. Qualitative Data

#### *Sample*

Interviews were conducted in Japan and Canada using each respondent's native language, either Japanese or English. The sample consisted of 12 Japanese (5 females, 7 males, mean age = 18.67, SD = .65) and 9 Euro-Canadian (9 females, mean age = 22.00, SD = 4.00) undergraduate

students. They were chosen to be interviewed as they met the criteria to be at-risk respondents for endorsing *Taijin Kyofusho*. Among participants who agreed to be recruited for interview part of the study, only those who scored above the mean on TKSQ at the initial screening data collection were contacted. Of those who met the criteria and were contacted for the interview, response rate was 85% for Japanese and 43% for Euro-Canadian samples. Demographic characteristics of Japanese and Euro-Canadian samples are shown in Table 4.

***Procedure***

Of those who agreed to participate in the interview part of the study on the initial screening questionnaire (Dataset 1), respondents who scored above the mean on TKSQ were contacted to schedule an interview. When those selected respondents agreed to participate, they were further informed about the purpose of the study. Upon arrival for the interview, the participants were asked to sign an informed consent. The interviews were audio recorded with the participants' consent. Following this, respondents were asked to complete the 30-item TKSQ. The interviewer immediately calculated the total score for each symptom domain on TKSQ and chose the two highest scored domains and the lowest scored domain for the interview. Following a script, respondents were then interviewed with a series of open-ended questions in which they were asked to recall the thinking processes they used to arrive at their answers to the TKSQ items. Interviews lasted between 60 and 75 minutes. Each respondent was compensated in the form of cash (1,000 yen in Japan; 15 dollars in Canada) at the conclusion of the interview.

***Cognitive Interviewing***

Cognitive interviewing is an in-depth, face-to-face, semi-structured interview method to elicit survey respondents understanding of instrument items (Willis, 2004). In the present study, CI was used to explore the meaning of items in the TKSQ. An interviewer bilingual in Japanese

and English conducted CI at both sites. The interviews utilized a retrospective verbal probing technique commonly used in CI research (Willis, 2004; Willis et al., 2005). In retrospective verbal probing, respondents were probed as a debriefing activity after the completion of questionnaire administration as opposed to concurrent verbal probing in which probes are administered concurrently during the interview. Because the current study focused on the two highest-scored domains to interview, respondents first needed to complete the survey, and the interviewer had to identify those items before probing or without disrupting the interview. Therefore, respondents answered the questions on a paper-and-pencil questionnaire sheet first, and then they were debriefed and probed retrospectively to explain each answer. Follow-up questions were based on comprehension, judgment, memory retrieval, response selection, and elaborative/ethnographic cognitive probing objectives. Typical questions pertaining to the objectives included, "What does the term 'embarrassment' mean to you?", "What went on in your mind when you answered this question?", "Having 1 through 4 answer options, why did you select your answer? How sure are you of your answer?", and "Can you tell me more about that?"

If a respondent showed difficulty or uncertainty giving an answer, the interviewer probed: "Can you give me an example to explain?" Specifically, for the fear in each interpersonal focus respondents were also asked if and how they differentiated between embarrassing the self, making someone else uncomfortable, and offending someone. They were asked to paraphrase the question and describe the situations in their own words that did not appear in the original question. This probing protocol was designed to gain a better understanding of the ways in which respondents interpreted and conceptualized the question and considered their own unique experiences in providing an answer. The semi-structured nature of

the verbal probing allows an interviewer to focus on the content as well as the direction of the interviews. At the same time, this technique also facilitates open-ended answers to allow participants to respond to the questions in their own words.

The design of the interview protocol was first evaluated in collaboration with the research supervisors and colleagues, including five clinical psychologists, a psychiatrist, and a qualitative psychology researcher who served as experts, translators, reviewers, and adjudicators in each site. Two pilot interviews in each site were conducted to determine logistic and procedural problems that may influence the participants in each site as well as to assess the cross-cultural equivalence and appropriateness of the question structures to address the research questions across the two sites. Several modifications were made in the protocol after piloting the interview guide. Specifically, because the cognitive interviewing protocol was available only in English, it was translated into Japanese. Experienced bilingual and Japanese-speaking survey researchers, including the bilingual interviewer, discussed linguistic issues (e.g., terms, expressions, idioms) that could potentially undermine item equivalence. Concerns and recommendations for modification were then discussed in conference meetings over the internet between Japan and Canada. The research team then carried out the translation and ensured equivalence between the English and Japanese versions of the interview protocols.

When selecting interviewers for CI, Willis (2004) suggests that interpersonal skills, technical abilities, familiarity with questionnaire problems, understanding of respondents' narratives, and awareness of the subtleties in the data should be taken into account. Interviewers are responsible for both recognizing the respondents' point of view and eliciting meaningful information from the respondents needed to identify emerging patterns and themes. Furthermore, an interpretivist approach to cognitive interviewing suggests that the interviewer also serves as a

data analyst who can provide an in-depth examination of respondents' stories (Miller et al., 2014). In effect, the interviewer becomes equivalent to other qualitative researchers (Miller et al., 2014). The interview process, then, should be understood as a complex interviewer-respondent interaction, in which interviewers needed to decide how to present themselves, situate the interaction with the participant, and build rapport (Fontana, 2002).

The current study focused on the evaluation of both the questionnaire functioning and phenomenology of TKS as it pertains to respondents' life experience. A bilingual interviewer conducted the interviews in both sites. The interviewer played an integral and active role in collecting and interpreting the narrative. Clearly, understanding both Japanese and English, the social and cultural contexts of the participants, and qualitative research methods in mental health was important in the current study. The selection of the interviewer in this study was determined based on these considerations. The bilingual interviewer in this study had: (1) an educational background in psychology; (2) research experience in both quantitative and qualitative research methods; (3) work experience in mental health; and (4) lived experience in Japan and North America. The background and experiences of the interviewer may have helped to obtain valuable data and made the analysis process more efficient.

## 3.4. Data Analysis

### 3.4.1. Quantitative Data

To address the issue of cross-cultural comparability of TSKQ between Japanese and Euro-Canadian participants in this study, DIF analysis and correlation analyses were conducted using two different data sets. All analyses were conducted using SPSS version 22 (IBM, 2014). Before conducting DIF analysis, the study first examined the scale reliability for the samples drawn from the cognitive interviewing screening phase (Dataset 1). Quantitative measurement

studies concern the design and evaluation of a scale or test for the assessment of latent constructs (e.g., intelligence, attitude, behavior), which cannot be directly and explicitly measured with a single item but a set of items or scores. The goal in quantitative measurement studies is to obtain observed scores as close to the real score as possible. In order to assess such constructs, measures have to demonstrate validity. Reliability, the degree to which the measurements are free from error, is a necessary condition for a measure to be valid. Reliable measurements minimize random errors (Tekera & Doganb, 2015). In the present study, internal consistency, a measure of how well the items on a test capture the same construct. Cronbach's alpha and inter-item correlations, which are the most widely-used methods of measuring internal consistency reliability, were assessed in the current study.

First, DIF analyses were performed using the Dataset 1. Specifically, DIF analyses were conducted to assess whether particular *Taijin Kyofusho* symptoms (individual TKSQ survey items) behaved differently across the two cultural groups. The patterns of respondents' answers across groups were based on the scores given to each survey question. Thus, DIF analysis focused on the scoring format rather than the question format. The two most widely used scoring formats for surveys and measures are binary (e.g., correct/incorrect or yes/no format) and ordinal (e.g., Likert-type format). For the past several decades, researchers in the field have developed and refined statistical procedures to detect the incomparability of survey questions.

One issue is that, historically, most of the standard DIF statistical methods, however, have focused on binary responses. The current study used a procedure developed by Zumbo (1999) because it provides a natural extension of the methodology to detect item-level measurement bias for ordinally-scored items. This approach uses three-step ordinal logistic regression modeling that allows researchers to detect uniform and non-uniform DIF

simultaneously with a 2 degree-of-freedom Chi-squared ($\chi 2$) test as well as the DIF effect size using Nagelkerke's pseudo $R^2$ values. Uniform DIF is identified when the test item is systematically more difficult for one group than it is for another even after matching respondents on ability or trait. Non-uniform DIF is identified when there is an interaction between group assignment, ability/trait level, and item responses. That is, the shift in item difficulty is not consistent across the ability levels because the difference switches direction for one group compared to the other at some point across the trait continuum. This technique is advantageous because it requires much smaller sample sizes, less data manipulation, and inexpensive software compared to other techniques (e.g., Item Response Theory methods) (Banks, 2015; Holland & Wainer, 1993; Zumbo, 1999). The approach has been successfully used in other cross-national, cross-language psychopathology studies (e.g., Ryder et al., 2008).

In this study, ordinal logistic regression (OLR) DIF detection was applied separately to items related to the six subscales. In this hierarchical sequential regression modeling, predictors are used to test the probability of observing a given level of an item response. In order to detect DIF, predictors include the total score as a trait/ability measure, a grouping variable (e.g., cultural group), and the interaction between trait/ability and group. In the current study, individual TKSQ items were entered as dependent variables, and outcome variables were added in the three-step modeling equation: (1) total score (the conditioning variable); (2) total score + cultural group (the grouping variable); (3) total score + cultural group + (total score × cultural group interaction term). The two-degree-of-freedom (2-df) Chi-squared difference ($\chi 2$ difference), accompanied by a large effect size using the Nagelkerke $R^2$ from step 1 to step 2 and from step 2 to step 3, indicates the presence of uniform DIF and non-uniform DIF respectively. The sequential technique of the DIF process enables the researchers to

simultaneously determine both forms of DIF. Following Zumbo (1999), the criteria used to flag items with DIF were a statistically significant 2-df $\chi 2$ test ($p \leq .01$) between steps and large $R^2$ difference (defined as $R^2 \geq .13$).

Second, correlation analysis was performed using the Dataset 2. Specifically, Spearman rank-order correlations were computed to examine the relationships among TKSQ, TK-S, SIAS, and SCS measures (Independent Self-Construal and Independent Self-Construal subscales) separately for Japanese and Euro-Canadian groups. Internal consistency analyses (i.e., Cronbach's alpha and inter-item correlations) were also conducted to evaluate the reliability of all scales.

### 3.4.2. Qualitative Data

All interview data were first transcribed to transform verbal data into a textual discourse, allowing the analysis of the data to be usable and efficient. Next, the Question Appraisal System (QAS-04) was applied to code the transcribed data. QAS-99 was first developed in 1999 (Willis & Lessler, 1999) to identify potential issues with survey questions through an expert review process and later updated as the QAS-04 (Dean et al., 2007 Reed, & Quiroz, 2007). It uses a checklist model and is useful for coding cognitive interviewing results. The QAS-04 consists of eleven major coding categories, each containing several subcategories, for a total of 52 codes (see Appendix C). In this updated version, a series of practical guidelines to establish item and concept equivalence across languages and cultural groups were taken into consideration.

For instance, a category for cross-cultural considerations, consisting of seven codes addressing concept and normative equivalences, was included. The seven cross-cultural consideration codes are: Reference Periods (8a); Knowledge (8b); Measuring Units (8c); Assumptions (8d); Response Categories (8e); Name Format (8f); and Politeness (8g). These seven codes identify concepts or understanding of the question that may vary across cultures. The code Assumptions (8d), for example, can be applied to identify local or emic population differences that may influence responses about religious beliefs (e.g., Christian or Muslim) or sports (e.g., European football vs. American football). QAS-04 allows researchers to systematically analyze both cultural elements that give rise to varying responses to the items and social-structural elements that highlight the background context that the question must capture. After reviewing all the transcripts for Japanese and Euro-Canadian data separately, problems that emerged from the data were noted. Next, each written comment was coded according to the QAS-04 manual. Then, frequency of coded problems was calculated. Further qualitative analyses

were conducted by summarizing the overall qualitative trends to elucidate the types of problems that each cultural group experienced.

# CHAPTER 4: RESULTS

## 4.1. Quantitative Data

***Dataset 1***

### 4.1.1. Preliminary Analyses

Multiple studies have shown the sub-scales to have either acceptable internal consistency (Singelis, 1994; Singelis and Sharkey, 1995; Sato and McCann, 1997; Norasakkunkit and Kalick, 2002), or low reliability estimates (Okazaki, 2000; Levine et al., 2003). In the present study, the Cronbach alpha coefficient for the independent self-construal sub-scale was 0.51 for the Indonesian sample and 0.55 for the Swiss sample. For interdependent self-construal sub-scale the Cronbach alpha coefficient for the Indonesian sample was 0.53 and for the Swiss sample was 0.62. These Cronbach alpha coefficients indicate moderate internal consistency.

For the Dataset 1, Missing data analysis showed that 0.7% of the responses were missing from the combined dataset. Euro-Canadian data had no missing responses, but 1.0% (2 cases) of the responses were missing from the Japanese data. Therefore, Little's Missing Completely At Random (MCAR) test was performed for the Japanese data only (Little, 1988). Little's MCAR test is one of the most commonly used techniques to assess whether missing data are missing completely at random. If the probability that missing data are independent of both observable and unobservable variables of interest. Following the MCAR test procedure, Chi-square tests between the missing variables and observed variables in the data set allowed for a test of missingness assumption. A large *p* value ( > .05) indicates one fails to reject the null hypothesis. In the current study, results showed that responses were missing completely at random, $\chi^2$ (27) =

9.957, $p = .99$, suggesting the weak evidence against the null hypothesis. Because missing data were not extensive, listwise deletion, the most commonly used missing data technique in OLR was used (Banks, 2015).

As part of the standardization process, it is important to reduce the effect of extreme values or outliers in the data set because the data can be heavily influenced by the possible spurious values. Winsorizing is a method of transforming any value above the ninety-fifth percentile into ninety-fifth percentile, and values below the fifth percentile rank into fifth percentile, improving the estimates by making the outlier look like other values (Ghosh & Vogt, 2012). To carry out winsorizing, the values are first converted to the z-scores, and then values above or below z-score of 3.3 are identified and modified. Mahalanobis Distance is one of the most widely used statistical procedures to detect outliers across variables (multivariate outliers) (De Maesschalck, Jouan-Rimbaud, & Massart, 2000; Mahalanobis, 1936). It measures the distance between a point and a distribution, in which the distance is zero if the point is at the mean of the distribution. Mahalanobis Distance takes into account the correlation in the data, and the variance of each variable and the covariance between variables. Mahalanobis Distance procedure first converts the data into standardized data and compute the ordinary Euclidean distance for the transformed data. SPSS performs linear regression to provide Mahalanobis $D^2$ values. If any cases in the date set has a Mahalanobis $D^2$ with a $p$ value > .001, then those cases are referred to as multivariate outliers. In the current study, two univariate outliers were identified on the O-TKS and D-TKS subscales for Euro-Canadian data and winsorized to 3.3 standard deviations from the mean. No multivariate outliers were observed using the Mahalanobis distances for the subscales.

### 4.1.2. Scale Reliability and Inter-Item Correlations

The internal consistencies of the TKSQ in Dataset 1 are shown separately for the Japanese and the Euro-Canadian groups in Table 5. Both groups demonstrated a similar pattern of correlations among the measures. Cronbach's alpha coefficients of the TKSQ total scores and six subscales reached satisfactory levels for both groups. The analysis of the average inter-item correlation is useful for evaluating the effect of item redundancy in relation to the number of items in a scale, especially during the stage of scale development process. The average inter-item correlation computes the average of all paired item-correlations. This internal consistency measure compares correlations all pairs of items that measure the same construct by assessing the extent to which answers to one test item correlate with answers to other test items. As the number of items increases, α also increases. However, if items are worded too similarly, this can increase the average inter-item correlation, which in turn artificially increases Cronbach's alpha (DeVellis, 2012). Although some similarities among items is needed to capture the construct being measured, redundancy or slight variations in wording provide very little information about the construct. Low average inter-item correlations do not necessarily imply a high degree of measurement error, but may hinder the fact that the items measure distinct constructs that have a weak relationship.

Researchers use different standards to evaluate average inter-item correlations, taking a heuristic approach to examine the effect of number of items given the lack of "hard" statistical criteria (DeVellis, 2012). Robinson et al. (1991) recommended an average inter-item correlations of .30 or better as satisfactory. Clark and Watson (1995) initially advocated the average inter-item correlations range of .15 to .50, but later advocate a range of .40 to .50 for narrowly defined constructs. Piedmont suggests the average inter-item correlation for a set of items should range

between .20 and .40 (Piedmont, 2014). The average inter-item correlations for the measures in the present study reached the satisfactory levels advocated by the most researchers with a of *r* = .33 to .40 for the Japanese group, suggesting that although the items are reasonably homogenous, they do have a great deal of unique variance. In contrast, the Euro-Canadian group had two items with average inter-item correlations higher than .40 (*r*s = .45, .41), meaning that those items may be capturing a small portion of the construct or an over presentation of the same construct domain.

### 4.1.3. DIF Analysis

The analysis found three items (10) with substantial DIF (items 6b, 8a, and 9a) following the criteria suggested by Zumbo (1999). The Chi-square statistic test showed that the item-by-group interactions were statistically significant for 19 items (ranged from $\chi^2$ (2) = 9.09 to 116.14, $p \leq .01$). The effect sizes for these 18 items ranged from $R^2$ =.00 to .23. Of these 18 items, three had effect sizes larger than .13. Therefore, only these three items were analyzed further to identify the type of DIF. In OLR terminology, DIF is a test of any kind of DIF or simultaneous test of uniform *and* non-uniform DIF; however, this sequential OLR modeling allows one to also the measure *R*-squared differences between Step #1 and Step #2 to identify the presence of uniform DIF, and between Step #2 and Step #3 to identify non-uniform DIF. For item 6b, 8a, and 9a, the differences in *R*-squared from Step #1 to Step #2 were quite large for the three items ($R^2$ =.128, .119, and .229, respectively), and from Step #2 and #3 were small ($R^2$ =.004, .016, and .001, respectively) suggesting that DIF was predominantly uniform. The summary of the DIF analyses for all the TKSQ items are presented in Table 6.

***Dataset 2***

### 4.1.4. Preliminary Analysis

For Dataset 2, missing data for the measures used in the correlation analysis was not extensive for either group: Japanese = 0.3% and Euro-Canadian = 0.4%. Approximately 0.4% of responses were missing from the combined data set. Little's MCAR test was conducted for the Japanese and Euro-Canadian participants separately. Results showed that missing data for both groups were completely at random, $\chi^2$ (1643) =1639.72, $p$=.52 for Japanese participants and $\chi^2$ (91) =97.55, $p$=.30 for Euro-Canadians. Neither univariate nor multivariate outliers were identified. Because there were few missing data points, imputations were not conducted, but mean item scores were calculated in the syntax for participants who completed at least two-thirds of each multi-item measure. Means and standard deviations are presented in Table 7.

### 4.1.5. Scale Reliability and Inter-item Correlation

All measures' internal consistencies reached satisfactory levels in the same range as estimates reported in previous studies. Cronbach's alpha coefficients and average inter-item correlations are presented in Table 8.

### 4.1.6. Correlation Analysis

Table 9 contains the correlations among TKSQ, TK-S, SIAS, independent self-construal and interdependent self-construal by cultural groups. The results suggest that correlations among the TKSQ, TK-S, SIAS, and interdependent self-construal were statistically significant and were greater or equal to $r$ (122) = .33, $p$ < .05, two-tailed for Japanese participants and $r$ (112) = .23, $p$ < .05, two-tailed for Euro-Canadians. While interdependent self-construal was consistently associated with higher levels of all social anxiety measures for the Japanese group, it was only positively correlated with the TKSQ in Euro-Canadians. SIAS showed stronger correlations with

both TKSQ and TK-S for the Euro-Canadians ($r$s = .50 and .78, both $p$s < .01) than the Japanese participants ($r$s = .47 and .67, both $p$s < .01). In summary, the results suggest that in both Japanese and Euro-Canadian samples, TKSQ scores were significantly correlated with other measures of social anxiety as well as interdependent self-concept. Specifically, the interdependent self-construal was found to be a predictor of social anxiety irrespective of the interpersonal focus of fear (i.e., other vs. self) for the Japanese group.

## 4.2. Qualitative Data

### 4.2.1. Demographic characteristics and data collection background

The qualitative component of the study analyzed a total of 21 interviews (12 Japanese and 9 Euro-Canadian participants). The self-reported demographic characteristics of Japanese and Euro-Canadian groups are presented in Table 4. There were several important differences in the participant characteristics across the two groups. Gender distribution was not balanced across the two groups. All the participants for Euro-Canadian group were female, whereas there were more male participants in the Japanese group. On average, Euro-Canadian group was older (mean = 22.00, *SD* = 4.00, range from 19 to 32) than the Japanese group (mean = 18.67, *SD* = 0.65, range from 18 to 20). Japanese participants were recruited from introductory psychology courses during the first two months of the academic year, and were mostly first year university students. In contrast, many of the Euro-Canadian students were psychology majors who may have had more experience in answering surveys and participating in interview research in psychology. There was no declared major data available for Japanese students because they were first or second year students enrolled in humanities and social sciences faculties and therefore had not yet selected majors.

### 4.2.2. Assessment of TKSQ

The data analysis focused on identifying the presence of problems with TKSQ items, which could affect the function of the TKSQ scale across two cultural groups. This analysis used a problem-feature focused coding scheme, the QAS-04 designed for recording questionnaire problems. Respondents' narratives were also reviewed to compare the interpretation patterns of both groups. Table 10 presents the frequency of symptom domains reviewed by the Japanese and Euro-Canadian participants. The analysis could not be conducted for all 30 items because not all

items were reviewed by the participants. Some symptom domains were not reviewed by any of the participants because those domains were not among the highest two scored domains. For example, items within symptom domains 9 (eye contact) and 10 (physical appearance) were not reviewed by any of the participants in the two groups. In addition, for Euro-Canadian group, symptom domains 4 (voice trembling) and 5 (blushing) were not reviewed. None of the Japanese participants reviewed symptom domain 3 (body trembling). Therefore, only the symptom domains that were reviewed by at least one participant were analyzed. The majority of symptom domains reviewed were *offensive symptoms* for both groups (Japanese: 16 out of 24, 66.7%; Euro-Canadian: 16 out of 18, 88.9%).

The frequency analysis presented in Table 11 shows that there were no problems identified in Reading, Instructions, Knowledge/Memory, Sensitivity/Bias, Cross-Cultural Considerations, and Cross-Question categories. Clarity and Assumptions categories were the most frequently identified problem categories for both groups. Cross-Cultural Considerations and Cross-Question problems were not found given the nature of the scale structure. Translation issues were also identified for the Japanese group. Therefore, further qualitative observations were conducted for the rest of the problem categories: Clarity, Assumptions, Response Categories, Translation, and Cross-Question.

Willis and Zahnd (2007) suggest that one should avoid statistical analysis of QAS results but simply present descriptive information gained from the narratives. While it is possible to quantify CI data, it may be misleading rather than informative. Given the qualitative nature of research inquiry investigated in CI, the obtained statistical effects may be influenced by a number of confounding factors. CI often relies on a mix of observed and implicit information (Willis et al., 2005). In most cases, the written comments of CI interviewers — who take

informal notes as the participants answer probes during the interview — are also meaningful data. Willis (2004) argues that, in practice, readers of CI results are not interested in the detailed process. Rather, they would simply desire a user-friendly reports presenting (a) what was found to be a problem and (b) what can be done about it (Willis et al., 2005). Following Willis' (2004) guidelines for informal analysis of interview results, this study produced a descriptive report which aggregates the results from the coding scheme, informal handwritten notes for each participant, and the results across interviews, in order to provide an integrated, clear, and comprehensive summary of the study's complex results. The next section provides a written summary for each problem category identified using the QAS-04 coding system.

***Clarity***

The clarity code was applied each time participants asked the interviewer for clarification or expressed difficulty or confusion understanding a term or sentence. Clarity code was applied when (1) question is vague because there are multiple ways in which respondents can interpret it; (2) what is to be included and excluded is unclear; (3) question contains undefined or unclear common terms or words; and (4) question simply lacks enough information given to prompt a meaningful response. Specifically, items asking about the fear of causing discomfort to and the fear of offending someone due to symptoms in the presence of others appeared to be problematic for many participants in both groups, because there were multiple ways to interpret "someone" and "others," or to decide who was to be included or excluded. Some of the participants even questioned the interviewer as to whom the "someone" and "others" referred, and asked for clarification while answering the survey. For instance, regarding domain 8 (intestinal gas), one Euro-Canadian participant asked the interviewer, "do you mean 'others' as strangers or some other people? Which context exactly is it asking for?" One Japanese participant asked what the

verb "embarrassing" means. Although it is not a technical or academic verb, the participant explained it is not commonly used to describe people's feelings in the spoken language; therefore, it was still difficult to understand. The participant further expressed that the verb felt too literal to express his feelings of anxiety. Similarly, one Euro-Canadian participant mentioned that "releasing intestinal gas" is not really used in daily conversation, and it could sound too technical or even medical to some extent. She pointed out that "farting" sounds more natural and appropriate in this context. Furthermore, she made a similar point for the use of "perspire" for the symptom domain 5, suggesting an additional comment about choice of the word in a question regarding a domain other than the ones she was interviewed for at the end of the interview, is yet important to address as a potential issue in this category.

There were some instances in which Euro-Canadian participants had difficulty with differentiating between the *self* and *others* across the symptom domains when they were asked about the fear pertaining to making someone uncomfortable and offending someone. For example, regarding item 8b (releasing intestinal gas), one participant first explained that she would be fearful of making someone uncomfortable, because causing discomfort to others can in return make herself feel uncomfortable; however, she became unsure later if she would be fearful of making someone uncomfortable more than making herself uncomfortable when releasing intestinal gas. Similarly, for symptom domain 6 (body odor), one participant became confused about what exactly she was afraid of, whether she was afraid of her own feeling or feelings of others when she was asked to explain the difference. Initially, she rated herself as "extremely fearful" of making someone uncomfortable and offending someone due to body odor, but she changed her answer to "moderately fearful" during the course of the interview when she realized

she was afraid of making herself uncomfortable or embarrassed more than making other people feel uncomfortable or offended.

***Assumptions***

Assumption issues were found to be common in both groups. This category is somewhat similar to a clarity problem, but the code is intended to capture the assumption of a constant pattern of behavior or inappropriate assumptions made about respondents. Based on the QAS-04 coding manual, each instance in which a participant mentioned, "it depends" was recorded as an assumption problem. This problem was particularly pronounced for symptom domains 1, 3, 4, 5, 6, and 8 (i.e., questions pertaining to "*in the presence of others*"), in which many respondents in both groups expressed that their fear level would *depend* on who is around them. Many participants explained the level of closeness or types of relationship to the person present might change the severity of fear. Most respondents mentioned that their fear level would decrease in the presence of close friends and family members. They defined "others" as acquaintances, distant friends (including classmates), and strangers that they encounter in public.

Seven Euro-Canadian students defined "others" as strangers in public transportation settings. They mentioned they were fearful of causing discomfort to or offending these strangers, suggesting that they interpreted "others" and "someone" as being the same people. Two Japanese participants explained that the existence and severity of fear depended on whether or not they defined "others" *and* "someone" as the same people, and that their answers would differ according to the situation and types of relationships involved. For instance, when asked about symptom domain 1 (blushing), a Japanese participant mentioned that if she were to blush in front of both strangers and family members at the same time, she would only be extremely fearful of embarrassing or causing discomfort to her family members. She explained that she would feel

extremely guilty or apologetic for causing her family members to feel embarrassed rather than proud of her performance in front of "others." However, she mentioned that she would not be fearful of embarrassing or making strangers uncomfortable if she was to blush in front of strangers only. Interestingly, none of the Euro-Canadian participants expressed such concern for family members, although they appeared to have similar concerns for their friends.

A Japanese participant noted that the size of audience or number of people around her would change her fear level if her facial expression were to stiffen. She mentioned that she would feel more fearful to show her facial expression stiffening in front of a person in one-on-one situations compared to a larger audience situation. She thought that the large audience would not be so bothered or offended by her facial expression stiffening because they would understand that she felt nervous in a public speaking situation. However, in a one-on-one context, she would be afraid of making a bad impression or even disrespecting the person she engaged in close conversation by showing her nervous or uncomfortable emotional state. Similar comments and patterns of interpretation were made by other Japanese participants for symptom domains 1 (blushing) and 4 (voice trembling).

Other inappropriate assumptions were made about items asking about the distinction between the fear of making someone uncomfortable and offending someone across different domains. One Japanese participant mentioned that there was no difference between the two, and he did not see the need to have two different options. He noted that people who are feeling uncomfortable are already feeling offended, and thus, both feelings should be classified as one single "negative" feeling. In contrast, another Japanese participant thought that causing discomfort to someone was more detrimental and dangerous than offending someone. She suggested that the feeling of discomfort was more "uncomfortable" than the feeling of offence.

She added that those who felt offended often expressed this negative feeling more explicitly than those who felt uncomfortable. Feelings of discomfort are generally more suppressed or less expressed; therefore, she would be more afraid of making someone feel uncomfortable or suppressed than offended.

One Japanese participant stated that it was *unlikely* that she would offend anyone by blushing, although she gave the "extremely fearful" response choice. She noted that offending someone involves intention or controllability. She explained that blushing is a natural body reaction, not an act that one has intention to do or can control, and that it should therefore be tolerated and accepted. She would still be extremely fearful, stating that while it was highly unlikely that she would offend someone by blushing, if someone were offended it would feel like "the end of the world." Her answers, then, were based not on the likelihood or probability of offending someone, but on the possible severity of the offense.

The Euro-Canadian participants who reviewed domain 8 (intestinal gas) had different interpretations regarding the intention behind or controllability of, the offense. They believed releasing intestinal gas is a faux pas or violation of social norms and the values taught by their parents if it involves intention and voluntary action. Similarly, other Euro-Canadian participants expressed emitting body odor (domain 6) and staring (domain 7) deserves public censure given that it is controllable. Because these participants believed such social norms and upbringing are universal and expected in their respective society, it is highly *likely* that others would feel offended due to releasing intestinal gas, emitting body odor, or inappropriate staring.

***Response Categories***

Problems with Response categories were coded when participants expressed trouble in finding an accurate choice to select. The majority of the participants in both groups thought that

the number of response categories and descriptions provided in the survey were reasonable and easy to choose among. While, five Japanese participants expressed difficulty with some question response options, none of the Euro-Canadian participants expressed such concern. For instance, one Japanese participant mentioned that "fearful" (恐れる) does not quite capture her anxiety or fear, rather "uneasy" (不安になる) or "unpleasant" (嫌，嫌な感じ) may describe her emotional state better. Another Japanese participant pointed out that there was no neutral option, which forced him to choose either *mildly fearful* or *moderately fearful*. In addition, two participants had difficulty with differentiating between "mildly" and "moderately" fearful response options in the Japanese version.

***Translation***

Because the TKSQ was originally developed in English, potential translation problems were examined only for the Japanese group. Overall, Japanese participants interpreted the terms and wording used in the questionnaire in the same manner as Euro-Canadian participants did. However, one serious translation issue was found in the term "embarrassing." Both groups associated embarrassment with emotional distress and negative feelings. Specifically, they defined embarrassment as feeling shy, nervous, ashamed, bad, sad, and overwhelmed but some discrepancies appeared to reflect possible translation errors. The recurring interpretations of "embarrassing myself" that were specific to the Japanese group include feeling rushed, pressured, puzzled, confused, surprised, panicked, or experiencing a loss of memory, speech (e.g., mumbling, difficulty with articulation, feeling the loss of muscle control in mouth) and movement (e.g., loss of sensation, paralysis). Eleven out of the 12 Japanese participants defined at least one of these experiences as embarrassment. None of these interpretations were mentioned by the Euro-Canadian participants.

These feelings (e.g., shy, nervous, bad) and interpretations of embarrassment were often accompanied by the experience of intense physical or bodily reactions including both offensive and common TKS symptom domains when Japanese participants were asked to provide an explanation of the answer they gave on the questionnaire. For instance, one participant told the interviewer about her real life experience when she blacked out because she once blushed in public. That was the feeling of embarrassment in reaction to blushing. She did not feel "embarrassed" because of the fact that she blacked out. In her own words, she defined embarrassment as realizing that she had a panic attack and inability to speak due to blushing. Another participant defined it as an emotional state in which he feels surprised, confused and puzzled when he is uncertain about what to do in a certain situation, particularly about the symptom domain 8 (intestinal gas). He further explained that he would first feel surprised at himself accidentally and involuntarily releasing intestinal gas in the presence of others. He would then think that not only he does not want to "embarrass" himself, but also he did not want to "embarrass" his friend because of him releasing intestinal gas, meaning he did not want the friend to feel confused and uncertain about how to deal with the situation. Similarly, a participant also mentioned that she would feel lost at first if she were to emit body odors in the presence of her friends (symptom domain 6), then she would immediately feel rushed and pressured to find the ways to remove the odor. These differences in the interpretation of embarrassment among Japanese and English participants may be due to an inappropriate translation.

# CHAPTER 5: DISCUSSION

## 5.1. Significance of Findings

### 5.1.1. Quantitative Findings

The results of DIF analysis showed that three out of 30 items, or 10% of the TKSQ, demonstrated DIF. Although the probability of Japanese and Euro-Canadian participants of endorsing an item should depend only on their level of the latent variable (i.e., the degree of fear pertaining to different interpersonal foci due to certain symptoms), responses to these three TKSQ items also depended on membership in a cultural group or different language versions used. These findings imply that scores on these three items do not mean the same thing for all participants across the groups or language versions. Specifically, the items that displayed DIF were 6b (D-TKS, offensive symptom domain, fear of discomfort to others due to body odor), 8a (E-TKS, offensive symptom domain, fear of embarrassment due to intestinal gas), and 9a (E-CS, common symptom domain, fear of embarrassment due to eye contact).

The results suggest the presence of possible problems in the adaptation of the Japanese version used in the current study and that the two versions are not equivalent. One can conclude that the items with DIF may contain different conceptual elements that carry different meanings and interpretations because of instrumentation differences or different experiences related to the context or culture. Given the reduced validity of the scale demonstrated in this study, researchers and users of both English and Japanese versions of TKSQ should be cautious when choosing to use this instrument. Possible revisions and modifications should be considered, focusing on items identified as problematic.

Correlation analyses indicated that the TKSQ was associated with both the TK-S and the SIAS for both cultural groups, suggesting that TKS might not only be a Japanese culture-bound

disorder. The SIAS as a measure of SAD showed somewhat stronger correlations with the TKSQ and the TK-S for Euro-Canadians compared with Japanese participants, suggesting perhaps that the distinction between SAD and TKS may not be quite as clear-cut for Euro-Canadians. Furthermore, interdependent self-construal was associated with TKSQ in both groups, although it was not correlated with TK-S and SIAS among Euro-Canadians. These findings indicate a need to investigate underlying mechanisms contributing to interpersonal fear, and TKS-like presentations might occur in many different cultural contexts worldwide by people construing themselves as interdependent.

Unfortunately, the extant literature examining TKS in different cultural contexts is largely limited to identifying group differences using Western-derived diagnostic interviews and instruments. Little is known about the specific mechanisms contributing to observed cultural variations. Recently, researchers have identified specific cultural variations in mechanisms of social anxiety that may account for the differences in interpersonal focus. For example, recent efforts to investigate the cultural variations of social anxiety have begun to reveal how particular cultural values and orientations (e.g., collectivism in East Asia or individualism in North America) that shape the way in which the person defines or construes the self (e.g., interdependent self-construal in collectivistic contexts and independent self-construal in individualistic contexts) may play a role in determining the objects and expressions of social threat (Kirmayer, 1991; Kleinknecht et al., 1997; Markus & Kitayama, 1991; Zhang, Yu, Draguns, Zhang, & Tang, 2000). Dinnel et al. (2002) found that Japanese participants were more interdependent than American participants, and TKS symptoms were more likely to be expressed by Japanese participants who scored lower on independence and higher on interdependence (Dinnel et al., 2002; Kleinknecht, Dinnel, Tanouye-Wilson, & Lonner, 1994). Likewise, people

with more interdependent self-construal reported more symptoms of SAD and TKS than those with independent self-construal in Western cultural contexts (Choy et al., 2008; Dinnel et al., 2002; Kleinknecht et al., 1997; Rector et al., 2006). These variations reflect complex interactions of culture and psychopathology. The psychological study of mind-culture links requires careful analysis of both individual and different cultural and contextual cues (Ryder et al., 2011).

### 5.1.2. Qualitative Findings

CI provided information that could not be obtained through traditional psychometric evaluation of the survey. The involvement of in-depth interviews and consideration of the narratives of participants represents an advance over traditional psychometric methods based purely on statistical evaluation of scale reliability and validity. The findings of CI have yielded evidence about instrument issues identified in the data analysis that would not have been uncovered without the specific CI probes employed. For example, most participants asked the interviewer for clarification and pointed out the possible inappropriate assumptions made in the questions and response categories in the TKSQ. The ambiguous usage of the terms “someone” and “others” confused the participants and assumed constant behavior that may have created the differences in the response patterns. Specifying what is meant by “someone” or “others” may help participants provide more consistent answers. Similarly, the response categories of the TKSQ could be revised to capture more accurate response and to allow participants to choose a neutral option.

Some participants in both groups did not appear to distinguish the focus of fear adequately. Especially when they were asked to explain the differences and/or similarities between embarrassing the self and making someone uncomfortable or feel offended, the focus of the fear for the majority of participants was ego-centric rather than allocentric. Participants often

changed their answer as they became unsure about their initial response during the course of the interview. This may be due to an assumption error. Although the items are intended to capture three different levels of fear, which are assumed to be different from each other, perhaps for most participants the root of fear may be self-presentation or performance. That is, even in the case of "offensive type symptoms" participants may not be concerned about the adverse effects on *others* entirely, but concerned about the negative consequences on the *self* through the effects on *others.* This observation is consistent with the notion that TKS patients have a so-called "double face" (Yamashita, 1970). Hence, the fear of offending others and narcissistic aspect of TKS, often manifested in the fear of hurting or harming the self (e.g., receiving negative evaluation) coexist in TKS participants (Mukai, 2002).

Qualitative and quantitative methods can provide fundamentally different information about potential problems with survey instruments. In the present study, items involving the Japanese translation of "embarrassment" in the TKSQ survey did not appear to have a consistent meaning across Japanese and Euro-Canadian participants. Embarrassment is generally considered a distinct, universal emotion closely related to social anxiety, shyness, and shame across cultures (Edelmann et al., 1989; Ekman, 1992). The core component of embarrassment is an emotional state involving a concern with one's public image and negative evaluation by others. There have been inconsistencies and confusions in the translation of embarrassment in Japanese literature. General English-Japanese dictionaries translate the English term "embarrassment" to *Touwaku* (当惑) or perhaps *Konwaku* (困惑). Some use *Haji* (恥) or shame and *Touwaku* interchangeably (Hirabayashi, 2010). Others define *Touwaku* as embarrassment with an element of shyness (Kojima & Furukawa, 2012). While there is more than one translation available for embarrassment in Japanese, *Touwaku* has been the most popular choice

in the psychological literature to ensure distinction from shame. Watanabe, Okazaki, and Crystal (1998) explain that the Japanese translation of *embarrassment* originally derived from the definition defined in American English, "a feeling of being self-conscious, confused, and ill at ease," distinguishing from the term *shame*, "a painful feeling of having lost the respect of others because of the improper behavior, incompetence, etc. of oneself or another" (Webster's New World Dictionary of American English – Third College Edition, 1998). The Japanese version of the TKSQ in the current study also used *Touwaku* to mean embarrassment. However, the term *Touwaku* in Japanese can mean a state in which one feels confused, bewildered, and perplexed (e.g., Hirabayashi, 2010). The state of being confused and perplexed was a recurring interpretation made by Japanese participants in the current study. Interpretation of *Touwaku* made by Japanese participants appears to be more closely related to cognitive state rather than emotional state. None of the Euro-Canadians interpreted the feeling of embarrassment in the same way as the Japanese participants.

Japanese participants also expressed a series of extreme physiological responses accompanying their interpretations of embarrassment such as blacking out, or the loss of memory, speech, and movement. In contrast, Euro-Canadians reported only minor physical reactions such as sweating and did not report the extreme physiological reactions observed in the Japanese group. Japanese university students described emotional experiences related to self-conscious emotions; they associated the experience of *Touwaku* with physiological responses (e.g., bounding pulse, rapid heartbeat) although they did not report such responses for the experience of shame (Arimitsu, 2015). Arimitsu (2015) claims that Japanese understanding of shame (*Haji*) is perhaps closer to or equivalent of English definition of embarrassment, warning researchers to be cautious when using *Haji* to refer to shame. Although the current study did not test for the

differential understanding of the terms among embarrassment, shame, and shyness, how Japanese people represent, display, and experience embarrassment was substantially different from Euro-Canadians, which could be due to this translation error.

Although it is relatively rare that researchers report translation as an error component when evaluating an existing or well-established survey in cross-cultural research, the lack of adequate translation may cause substantial response problems. Willis and Zahnd (2007) demonstrated that many of the well-established surveys did not work well for Korean-speaking respondents due to translation issues in their empirical study investigating cross-cultural questionnaire design of California Health Interview Survey. Translation may create error variance or systematic bias in results due to subtle differences in meaning and interpretations of particular features.

### 5.1.3. Synthesis of Findings

To reiterate, the present study was not designed to investigate the cause of DIF, which might be done by using CI in a sequential design. Nevertheless, the qualitative and quantitative results complemented one another. In particular, both results identified items that were influenced by translation or by cultural assumptions that needed clarification, as well as divergent questionnaire design. Further evidence from the qualitative findings might explain the cause of the DIF for the two items with DIF as reflected in the qualitative findings. Item 6b and 8a were identified as DIF items and concerned the fear of embarrassment due to emitting body odors or releasing gas in the presence of others, which was viewed as problematic by participants in CI, possibly due to interpretation differences in understanding of the presence of others for both groups and specifically, the translation of the term “embarrassment” for the Japanese group. DIF analysis provided relatively little information about participants’ understanding of

"embarrassment" and "others" as well as subjective experience of differing fear pertaining to the symptoms, but the results from the qualitative part bridged this gap. Taken together these results suggest that revision and clarifications of these terms might lead to items that capture a more accurate and appropriate meaning of "embarrassment" for Japanese participants and of "others" for both groups.

The study also highlighted the unique contributions of the quantitative analysis. CI has no direct way to evaluate variance in reliability as it relates to measuring different levels of TKS. Reliability analysis showed the TKSQ to be a reliable scale within each cultural group. Correlation results were somewhat complex that SIAS was found to be more strongly correlated with TKSQ and TK-S for Euro-Canadians than did for Japanese sample. This suggests that greater effort needs to be made and decision should be made in careful consideration when refining the scale or revise the response categories to capture more accurate degree of "being fearful" for the Japanese group.

Due to the small sample size, not all 30 items were reviewed by the participants in CI part of the study; therefore, not all problems may not have been identified by the CI. However, DIF analysis partially compensated for this "missing data" and signaled item 9a as a potential threat to item bias given that item 9a was not reviewed and included in CI analysis. DIF detected for the item 9a, assessing the fear of making someone uncomfortable due to making eye-to-eye contact with another person, could be because of cultural differences in the meaning and consequences of eye contact. In Western cultures, making eye contact during social interaction is considered a sign of confidence and respect; however, in Japan, *avoidance* of eye contact is considered a sign of respect and humbleness (Sue & Sue, 1977). In an empirical study, Japanese people tended to use information from the eyes, while Americans relied on the mouth when

recognizing emotion in faces (Hawrysh & Zaichkowsky, 1990). These differences in understanding the meaning of "eye contact" and potential consequences of violation of the cultural norm across Japanese and Euro-Canadians might cause the DIF; in other words, the question may not capture the same psychological construct across the groups. In sum, the three items demonstrated DIF in the current study may limit the cross-cultural validity of the TKSQ.

## 5.2. Limitations

### 5.2.1. Quantitative data

It is important to note that recruitment and initial screening varied between the two groups. Euro-Canadian students were recruited using the university's online participant pool, whereas Japanese students were informed about the study verbally by their instructors. The differences in gender and age distribution and recruitment procedures across two sites may have influenced the outcomes of cognitive interviews. The sample size for each group was relatively small and unequal. Specifically, sample sizes for Japanese and Euro-Canadian participants were 198 and 78, respectively. In order to have satisfactory statistical power for ordinal logistic regression DIF detection, it is sometimes recommended that sample sizes be close to 200 (French & Maller, 2007). However, sample size in current study is comparable to that observed previous studies that analyzed DIF using samples under 200 (e.g., Thissen, Steinberg, & Gerrard, 1986; Zhou et al., 2014). Small sample sizes are also common in studies that examine translated tests (Fidalgo & Madeira, 2008; Muniz, Hambleton, & Xing, 2001). With small samples, the techniques used to deal with missing data may be critical. Listwise deletion has been widely used in DIF analysis; however, it may not be appropriate because there may not be sufficient observations at each trait/ability level to match the two groups (Banks, 2015).

In a review of missing data techniques in 16 DIF studies, listwise deletion was the most commonly used technique as it is easy and quickly performed in SPSS, although the impact on power tends to be smaller than other techniques (e.g., multiple imputation) (Banks, 2015). Banks found that power impact for DIF analysis tended to be larger when using modern techniques such as lowest score imputation than listwise deletion. Lowest score imputation missing data technique is recommended for small sample sizes. Unequal sample size and the technique used to deal with missing data in the current study may have impeded the results of DIF detection. For instance, in an empirical study that that analyzed the effect of sample sizes in detection of DIF, when there were large differences in sample sizes across groups, DIF detection was impeded the two sets of analyses (Hauger & Sireci, 2008).

Empirical studies that analyze DIF usually recommend the use of more than one statistical procedure to increase confidence in the results and overcome problems that may be caused by inherent limitations of particular statistical methods (e.g., labeling DIF false positive or missing true DIF) (Hidalgo & López-Pina, 2004). Exploring items with DIF at an early stage of the research is only a first step. Although no further investigation is necessary once DIF is identified, but this should not stop researchers to endeavor to look for other potentially problematic items that were not detected.

### 5.2.2. Qualitative data

The limitations of the present study include the use of a small sample of university students which may not reflect the full range of people with TKS. Due to the small sample size, not all the items were reviewed by the participants. In addition, respondents were interviewed only two highest scored-domains and one lowest-scored domain, a total of three symptom domains (a total of nine questions); however, qualitative analysis only focused on the two

highest scored-domains. Six items per participant were included in the analysis. Although TSKQ is not a tool to diagnose the presence of clinical TKS, a symptom is considered present if the respondents rate two or three (report moderate or extremely fearful) in at least one of the three interpersonal foci (embarrassing the self, making others uncomfortable, or offending others) in TKSQ (Choy, et al., 2008). Because one of the objectives of the study was to understand the construct and meaning of TKS among the two cultural groups, only the domains in which participants rated themselves highly fearful were relevant and included in the analysis. Many participants indicated that certain symptoms did not apply to them simply because they had never experienced them. While this is not a large number of CIs per item, it should be noted that the most commonly reported survey problem types are logical or structural issues, and they are sample size independent (Willis, 2015). If a survey question makes an erroneous assumption that influences a specific class of participants, then problems addressed by even one member of the class may be taken in to account. Likewise, the present study examined only a subset of items in the TKSQ. The sample was mostly limited to non-clinical, undergraduate psychology students, and the Euro-Canadian sample consisted of female students only. Given time and resources, a larger sample size and longer interview duration would yield a more thorough consideration of the measure. Conducting CI is resource intensive. The small sample size and open-ended question and answer approach obtained in CI surely makes it difficult to draw firm conclusions about causation and the strength of the relationship about items such as one might derive from path analysis and correlation analysis in quantitative methods.

Nevertheless, within the limitations of time and resources, the current study did make an effort to interview groups of people most relevant to the study by screening and selecting participants who endorsed multiple TKS symptoms. A revised version of the TKSQ, responding

to the issues raised in this study, should in future be evaluated using a larger sample. Furthermore, the whole notion of "small" is not appropriate in this current study because the participants of interest are those with social anxiety, who may not be easily sampled from a "target population." Specifically, men report higher levels of TKS and are found to be more socially anxious in Japan (Takahashi, 1989), but studies consistently suggest that men are more inexpressive and stoic in responding to stressors (Thoits, 1995). Men more often report controlling their emotional distress, not engaging in problem-solving efforts. In contrast, women have an emotional and expressive style of responding to psychological distress. Women more often seek social support and express their feelings. Irrespective of gender, socially anxious people may be more likely to avoid reporting and sharing their fears and anxiety with others. The nature of the recruitment of socially anxious participants in the current study, therefore, may have resulted in a "small number" of participants who agreed to come in for interviews.

Fortunately, in this kind of interview-based qualitative research, even a small number of respondents elaborate and represent meaningful experience-context links (Crouch & McKenzie, 2006). Small sample size is sufficient to bridge the gap between the respondents' personal experiences and the social context that envelops them, and every case should be taken into account. In qualitative studies, the guiding principle generally used in determining a sufficient sample size the idea of saturation. Saturation means reaching the point where the collection of new data (e.g., themes, issues relevant to the research question) does not shed any further light on the issues under investigation (Creswell, 2013; Tashakkori & Creswell, 2008). Thus, determining the sample size should be based on the concept of saturation. Thompson reviewed fifty qualitative studies conducted during the 2002-2004, and found that while sample size ranged from 5 to 350 with an average of 31, data saturation generally occurred between 10

and 30 interviews (Thompson, 2010). The sample size in the current study was small, but it should be noted that the results were developed through the quality or saturation of data when there was no further meaningful information relevant about the TKSQ.

Another limitation of the current study is the lack of inter-rater reliability in the coding of CI results given that the data was coded by the interviewer only. Although during the course of the data aggregation and synthesis process, two clinical psychology experts in each site were involved in reviewing the composite results, coding of the QAS-04 mostly relied on one analyst. To the extent that the results are substantially influenced by the subjective judgment of one analyst, study results must be interpreted with this limitation in mind.

For example, Benitez and Padilla (2014) took a step forward in combining quantitative and qualitative methods to study sources of DIF across American and Spanish participants. They demonstrated a new way of addressing the investigation of the causes of DIF within an MMR framework, which involved cognitive interviewing to reveal respondents' interpretations. Specifically, they first identified items with DIF by conducting statistical analyses (e.g., ordinal logistic regression) and then utilized cognitive interviewing to investigate the interpretations considered by the respondents. The results has proven a mixed method approach in studying DIF can make a difference to the investigation of cross-cultural scale equivalence. However, the use of CI is useful for researchers hoping to gain detailed information about items and respondents. The use of CI reinforces the need to address the in-depth issues of meaning in the development and refinement of scales prior to administering a measure across cultural contexts.

### 5.2.3. Cross-Cultural Site Equivalence

The present study attempted to ensure equivalence across sites in research procedure and data collection. Unfortunately, in the present study, the modes of data collection were not

completely compatible. Biases and difference may have been caused by procedural difference despite the efforts made in the study to achieve equivalence. In North America, online research participant pool system is common at educational institutions including universities, and students are encouraged to participate in psychological studies in exchange of extra credit in their enrolled courses. In Japan, the use of participant pool system has been an ongoing debate because of ethical concerns. Although some universities and laboratories allow researchers to utilize the system, most of them are still reluctant to choose the option. In the present study, the use of a participant pool system and online survey were not available for Japanese group; therefore, the study was advertised by word of mouth in Japan, and the participation was voluntary without any compensation. Euro-Canadian students were recruited through online participant pool system and received extra credits at the completion of the survey. Quantitative data collection undertook a paper-and-pencil form of the survey for both Japanese and Euro-Canadian data although online survey is more common for Euro-Canadian students. Score differences observed and even the qualitative data obtained in the present study may have been partly biased by the incompatibility in the participant recruitment and data collection procedure.

### 5.3. Contributions of the Study

The present study took a step beyond the usual post hoc statistical group comparisons to investigate the scale equivalence by incorporating a qualitative approach. Combining qualitative and quantitative methods allows the shortcomings of one method to be compensated by the strengths of another, thereby yielding more valid data. The Taijin Kyofu Sho Questionnaire (TKSQ) was developed to address the distinct symptoms of TKS. Although the TKSQ had no published psychometrics in Japanese and Euro-Canadian populations, and it has not been

clinically validated, the present study provided some insight into the potential and limitations of the TKSQ to assess individuals' experience of *Taijin Kyofusho* symptoms in Japan and Canada.

Although cross-cultural and linguistic comparisons have become increasingly common, researchers generally have paid insufficient attention to the issues of item bias and overall scale equivalence when comparing groups across cultural contexts. It is common for researchers utilizing survey questionnaires to not test equivalence directly without taking into account the possible bias threat in the survey used. If the construct being measured is not equivalent across cultural groups and countries, the inferences made from the results may not be valid. In recent years, many efforts have been made to ensure scale validity; however, the majority of such research has only focused on psychometric issues such as testing reliability and validity of measurements developed in English language in the West. Whether the construct being measured, and the instrument used to measure that construct, are consistent and fair to all participating groups with diverse cultural backgrounds remains an open question. The findings of the present study suggests the benefits of MMR in the cross-cultural study of a psychological construct over other methodologies and lead to the conclusion that the results could not be obtained by applying a single method.

### 5.4. Directions for Future Research

The findings of the present study suggest further steps to validate the TKSQ. To deepen our understanding of the cross-cultural validity of the TKSQ, it is necessary to extend this study to compare participants from more diverse populations, including clinical samples, English-speaking participants from different cultures, and participants from other East Asian countries. The results from this study have limited generalizability. A larger scale study should be conducted to take into consideration the within-culture or -country variations and to obtain more

generalizable results in both qualitative and quantitative methods. It is worthwhile to investigate DIF between diverse participants and different language versions. For example, DIF could be conducted to detect item biases with respect to age, gender, and educational status within a particular cultural group.

For a more comprehensive research design, one could interview and obtain quantitative data from four groups: (1) monolingual Japanese in Japanese, (2) bilingual Japanese in Japanese, (3) monolingual Euro-Canadian in English; and (4) bilingual Japanese in English. This sampling design would allow researchers to evaluate translation issues and determine whether and how cultural variations exist in respondent effects across cultural contexts. Second, the development of new DIF techniques may be useful to test data with a multilevel structure. For instance, clinical researchers may be interested in examining symptoms of TKS and test biases may change across time within a cultural group (e.g., symptom scores over time nested within patients). Currently, available DIF detection techniques do not allow researchers to analyze nested data.

Although statistical DIF detection procedures has been the core of item bias analysis so far, substantive qualitative analysis is of special importance as it brings up possibilities to eliminate bias in the future. CI focuses on investigating what is best understood by, and most concerning to, the participants, whereas quantitative methods focus on the measurements' ability to evaluate the different levels of fear pertaining to TKS among Japanese and Euro-Canadians in the present study. Hence, findings from each method to some extent address different questions about the instrument. Considering the optimal ordering of MMR, for the future studies, a sequential design may be appropriate to approach the evaluation of the scale. For example, researchers could conduct DIF testing first to identify items with biases and then conduct CI to

focus on those items to locate the cause of DIF in a sequential manner. The developers and users of the instrument are then left with the task of integrating these pieces of information and making a decision about the ways to refine the questionnaire to address the errors and concerns in the TKSQ. Finally, ensuring valid cross-cultural and linguistic comparisons should take into consideration the evaluation of all levels of equivalence: construct equivalence, structural equivalence, measurement unit equivalence, and scaler equivalence across groups (van de Vijver & Chasiotis, 2010). Future research should be directed at searching for evidence of all levels of equivalence in the psychological study of culture through the use of mixed-methods research approach to gain insights that cannot be generated with a single method.

## 5.5. Final Conclusion

The present study presented an approach to the evaluation of the cross-cultural equivalence of a psychological measurement by applying both quantitative and qualitative methods – mixed-methods research (MMR). Specifically, the study evaluated the scale functioning of Taijin Kyofu Sho Questionnaire (TKSQ) by conducting cognitive interviewing (CI) to elicit the interpretations made by the participants and differential item functioning (DIF) analysis to detect statistical item bias across Japan and Canada. This study was not designed to substitute for full psychometric evaluation of the TKSQ or provide a full description of appropriate use of each of the methods. Rather, the study sought to identify some ways in which qualitative and quantitative methods can be used together allow researchers to evaluate a questionnaire and identify problems that may cause response biases across groups, a goal motivated by the methodological shortcomings generally apparent in the cross-cultural assessment literature.

Qualitative and quantitative methods each have strengths and weaknesses. Qualitative methods enable researchers to have the advantages of flexibility and conduct in-depth analysis, whereas quantitative methods are tightly constrained by the data that are collected, hypotheses that are defined, and analyses that are planned in advance. On the other hand, quantitative results are generalizable to the target population and replicable, it is difficult to generalize and replicate the qualitative results. The MMR utilized in this study facilitated the integration of the results provided by both methods to better understand the "complete" story rather than relying on "one side" of the story.

The analysis of DIF results using the Dataset 1 showed that 10% of all items in TKSQ demonstrated statistical item bias. External validity study using the Dataset 2 indicated that the TKSQ was found to have significant positive correlations with TK-S, SIAS, and Interdependent Self-Construal for both Japanese and Euro-Canadian groups. However, it was found that SIAS had stronger correlations with the TKSQ and TK-S for Euro-Canadians than did Japanese participants. Results of CI showed that response errors were found to be clarity, assumptions, response categories, and translation areas, and different response errors and patterns were identified across the Japanese and Euro-Canadian groups.

Critically, each of these methods showed different types of question problems or located a different facet of the problems underlying the TKSQ the other method had missed. The results of these analyses suggest that the TKSQ was not found to be psychometrically sound and demonstrate adequate validity across two cultural groups. The findings of this study also suggest the potential benefits of MMR to evaluate and refine a scale assessing a latent psychological construct across two cultural groups. The use of MMR provides a tool to combine depth and breadth of knowledge and explore the construct of interest thoroughly across cultural contexts. In

sum, as reflected in the results in both qualitative and qualitative data in the study, the use of CI and DIF together have demonstrated value for identifying items with biases, evaluating cross-cultural scale equivalence, and also providing a more integrated view of the phenomenon of the fear of offending *others*, *Taijin Kyofusho*.

References

Ackermann, A. C., & Blair, J. (2006). *Efficient respondent selection for cognitive interviewing.* Paper presented at the Annual Meeting of the American Association of Public Opinion Research, Hollywood, FL.

Aizawa, N. (2014). 子どもの社会不安に関する心理発達的研究について : 研究ノート (神戸大学大学院人間発達環境学研究科, Trans.) *神戸大学大学院人間発達環境学研究科研究紀要* (Vol. 7, pp. 149-156): 神戸大学.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders fifth edition (DSM-5)* (5th ed.). Washington, D.C.: American Psychiatric Association.

Arimitsu, K. (2015). 自己意識的感情の経験的定義の言語間比較 1. *感情心理学研究, 22*(2), 53-59. doi: 10.4092/jsre.22.53

Asakura, S., Inoue, T., Kitagawa, N., Hasegawa, M., Fujii, Y., Kako, Y., . . . Koyama, T. (2012). Social Anxiety/Taijin-Kyofu Scale (SATS): Development and Psychometric Evaluation of a New Instrument. *Psychopathology, 45*(2), 96-101.

Balhara, Y. P. S., & Goel, R. (2011). Culture bound syndromes: Need to relook, relabel and include? *Asian Journal of Psychiatry, 4*(1), 75. doi: http://dx.doi.org/10.1016/j.ajp.2011.01.007

Banks, K. (2015). An Introduction to Missing Data in the Context of Differential Item Functioning. *Practical Assessment, Research & Evaluation, 20*(12), 2.

Benítez, I., & Padilla, J.-L. (2013). Analysis of Nonequivalent Assessments Across Different Linguistic Groups Using a Mixed Methods Approach: Understanding the Causes of

References

Differential Item Functioning by Cognitive Interviewing. *Journal of Mixed Methods Research*. doi: 10.1177/1558689813488245

Betancourt, T., & Bolton, P. (2005). Using qualitative methods to develop a locally-derived measure of psychosocial problems in Acholi war-affected children. *World Vision Uganda*.

Betancourt, T. S., Meyers-Ohki, S. E., Stevenson, A., Ingabire, C., Kanyanganzi, F., Munyana, M., . . . Beardslee, W. R. (2011). Using Mixed-Methods Research to Adapt and Evaluate a Family Strengthening Intervention in Rwanda. *African journal of traumatic stress, 2*(1), 32-45.

Blair, J., & Conrad, F. G. (2011). Sample Size for Cognitive Interview Pretesting. *Public Opinion Quarterly, 75*(4), 636-658. doi: 10.1093/poq/nfr035

Boeije, H. (2010). Analysis in qualitative research.

Choi, S. (1993). 韓国人の我々意識と日本人の我々意識の比較. *韓国心理学会 93 年次大会学術発表論文集*.

Choy, Y., Schneier, F. R., Heimberg, R. G., Oh, K. S., & Liebowitz, M. R. (2008). Features of the offensive subtype of Taijin-Kyofu-Sho in US and Korean patients with DSM-IV social anxiety disorder. *Depression and Anxiety, 25*(3), 230-240. doi: 10.1002/da.20295

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309.

Clarvit, S. R., Schneier, F. R., & Liebowitz, M. R. (1996). The offensive subtype of Taijin-kyofu-sho in New York City: the phenomenology and treatment of a social anxiety disorder. *Journal of Clinical Psychiatry, 57*(11), 523-527.

Cole, L. E. (2013). Japanese Culture-Bound Disorders: The Relationship between" Taijin Kyofusho, Hikikomori," and Shame. *Online Submission*.

References

Cooper, R. (2010). Are culture-bound syndromes as real as universally-occurring disorders? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 41*(4), 325-332. doi: http://dx.doi.org/10.1016/j.shpsc.2010.10.003

Creswell, J. W. (2013). *Research Design: Qualitative, quantitative, and mixed methods approaches*: Sage publications.

Creswell, J. W., & Tashakkori, A. (2007). Editorial: Differing Perspectives on Mixed Methods Research. *Journal of Mixed Methods Research, 1*(4), 303-308. doi: 10.1177/1558689807306132

Crouch, M., & McKenzie, H. (2006). The logic of small samples in interview-based qualitative research. Social science information, 45(4), 483-499. De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems, 50*(1), 1-18.

Dean, E., Caspar, R., McAvinchey, G., Reed, L., & Quiroz, R. (2007). Developing a Low‐Cost Technique for Parallel Cross‐Cultural Instrument Development: The Question Appraisal System (QAS‐04). *International Journal of Social Research Methodology, 10*(3), 227-241.

DeMaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results. *Methods for testing and evaluating survey questionnaires*, 89-108.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26): Sage publications.

Dinnel, D. L., Kleinknecht, R. A., & Tanaka-Matsumi, J. (2002). A cross-cultural comparison of social phobia symptoms. *Journal of Psychopathology and Behavioral Assessment, 24*(2), 75-84.

References

Doucerain, M., Vargas, S., & Ryder, A. G. (2015). Mixed-methods research: integrating qualitative and quantitative approaches to the psychological study of culture. . In N. Zane, G. Bernal, & F. Leong (Eds.), *Culturally-Informed Evidence-Based Practice for Ethnic Minorities: Challenges and Solutions.* Washington D.C.: American Psychological Association.

Edelmann, R. J., Asendorpf, J., Contarello, A., Zammuner, V., Georgas, J., & Villanueva, C. (1989). Self-reported expression of embarrassment in five European cultures. *Journal of Cross-Cultural Psychology, 20*(4), 357-371.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200.

Elkholy, H. A. (2012). Social Anxiety Disorder vs. Taijin Kyofusho: An Egyptian Perspective (第 107 回日本精神神経学会総会 (シンポジウム)). *精神神経学雑誌, 114*(42), 482-484.

Essau, C. A., Sasagawa, S., Chen, J., & Sakano, Y. (2012). Taijin Kyofusho and Social Phobia Symptoms in Young Adults in England and in Japan. *Journal of Cross-Cultural Psychology, 43*(2), 219-232. doi: 10.1177/0022022110386372

Fidalgo, A. M., & Madeira, J. M. (2008). Generalized Mantel-Haenszel Methods for Differential Item Functioning Detection. *Educational and Psychological Measurement, 68*(6), 940-958. doi: 10.1177/0013164408315265

Fontana, A. (2002). Postmodern trends in interviewing. *Handbook of interview research: Context & method*, 161-175.

French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement, 67*(3), 373-393.

References

Ghosh, D., & Vogt, A. (2012). *Outliers: An evaluation of methodologies.* Paper presented at the Joint Statistical Meetings.

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*(10), 1115-1124. doi: 10.1037/0003-066X.52.10.1115

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods, 18*(1), 59-82. doi: 10.1177/1525822x05279903

Hambleton, R. K. (1991). *Fundamentals of item response theory* (Vol. 2): Sage publications.

Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods*. Hoboken, NJ: John Wiley.

Hauger, J. B., & Sireci, S. G. (2008). Detecting Differential Item Functioning Across Examinees Tested in Their Dominant Language and Examinees Tested in a Second Language. *International Journal of Testing, 8*(3), 237-250. doi: 10.1080/15305050802262183

Hawrysh, B. M., & Zaichkowsky, J. L. (1990). Cultural Approaches to Negotiations: Understanding the Japanese. *International Marketing Review, 7*(2), null. doi: doi:10.1108/EUM0000000001530

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*(4), 766.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83; discussion 83-135. doi: 10.1017/s0140525x0999152x

References

Hidalgo, M. D., & López-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement, 64*(6), 903-915.

Hirabayashi, H. (2010). 感情の社会化. In K. Takemura, H. Kitamura, & C. Sumiyoshi (Eds.), *感情と思考の科学事典*: 朝倉書店.

Hofmann, S. G., Asnaani, A., & Hinton, D. E. (2010). Cultural aspects in social anxiety and social anxiety disorder. *Depression and Anxiety, 27*(12), 1117-1127. doi: 10.1002/da.20759

Hofmann, S. G., & Hinton, D. E. (2014). Cross-Cultural Aspects of Anxiety Disorders. *Current Psychiatry Reports, 16*(6), 1-5. doi: 10.1007/s11920-014-0450-3

Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry*.

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Hosoya, H. (2010). *社会不安障害*. Tokyo: 新興医学出版社.

IBM. (2014). Released 2014. IBM SPSS Statistics for Windows (Version 22). Armonk, NY: IBM Corp.

Imura, O., Nakamura, T., Shimabukuro, T., Tanahara, K., & Beku, U. (1991). 日本と韓国の青年の意識と行動に関する比較文化的研究--デモグラフィック要因と対人関係意識項目の分析. [A Cross-Cultural Study of Consciousness and Behavior between Japanese and Korean Students]. *琉球大学教育学部紀要 第一部・第二部*(38), 347-360.

References

Iwase, M., Nakao, K., Takaishi, J., Yorifuji, K., Ikezawa, K., & Takeda, M. (2000). An empirical classification of social anxiety: performance, interpersonal and offensive. *Psychiatry and Clinical Neurosciences, 54*(1), 67-75. doi: 10.1046/j.1440-1819.2000.00639.x

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

Karasz, A., & Singelis, T. M. (2009). Qualitative and Mixed Methods Research in Cross-cultural Psychology: Introduction to the Special Issue. *Journal of Cross-Cultural Psychology, 40*(6), 909-916. doi: 10.1177/0022022109349172

Kasahara, Y. (1987). *Social phobia in Japan.* Paper presented at the Social Phobia in Japan and Korea: Proceedings of the First Cultural Psychiatry Symposium Between Japan and Korea. Seoul, Korea: East Asian Academy of Cultural Psychiatry.

Keel, P. K., & Klump, K. L. (2003). Are eating disorders culture-bound syndromes? Implications for conceptualizing their etiology. *Psychological Bulletin, 129*(5), 747.

Kim, J., Rapee, R. M., & Gaston, J. E. (2008). Symptoms of offensive type Taijin‐Kyofusho among Australian social phobics. *Depression and Anxiety, 25*(7), 601-608.

Kim, M. (2005). 韓国と日本の大学生における対人不安と同一性，公的自己意識，相互依存的自己との関係. [The Relationship of Social Anxiety with Identity, Public Self-Consciousness, and Interdependent Self: A Comparative Study of Korean and Japanese University Students]. *パーソナリティ研究, 14*(1), 42-53. doi: 10.2132/personality.14.42

Kirmayer, L. J. (1991). The Place of Culture in Psychiatric Nosology: Taijin Kyofusho and DSM-III-R. *The Journal of nervous and mental disease, 179*(1).

References

Kirmayer, L. J. (2001). Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of Clinical Psychiatry, 62*, 22-30.

Kirmayer, L. J., Young, A., & Hayton, B. C. (1995). The cultural context of anxiety disorders. *Psychiatric Clinics of North America, 18*(3), 503-521.

Kitanishi, K., Lee, S., Choi, Y., & Nakamura, K. (1999). スペインにおける対人恐怖の 1 事例. [One case report of Taijin Kyofusho in Spain ]. *精神医学, 41*(8), 891-895. doi: doi:10.11477/mf.1405904829

Kleinknecht, R., Dinnel, D., Tanouye-Wilson, S., & Lonner, W. (1994). Cultural variation in social anxiety and phobia: A study of taijin kyofusho. *The Behavior Therapist (17, 8*, 175-178.

Kleinknecht, R. A., Dinnel, D. L., Kleinknecht, E. E., Hiruma, N., & Harada, N. (1997). Cultural factors in social anxiety: A comparison of social phobia symptoms and Taijin Kyofusho. *Journal of Anxiety Disorders, 11*(2), 157-177. doi: http://dx.doi.org/10.1016/S0887-6185(97)00004-2

Kojima, Y., & Furukawa, Y. (2012). 感情区別と感情制御との関係. *Bulletin of Aichi University of Education (Educational Sciences), 61*, 127-135.

Lee, S. (1996). Reconsidering the status of anorexia nervosa as a western culture-bound syndrome. *Social Science and Medicine, 42*(1), 21-34.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & quantity, 43*(2), 265-275.

Lewis-Fernández, R., Hinton, D. E., Laria, A. J., Patterson, E. H., Hofmann, S. G., Craske, M. G., . . . Liao, B. (2010). Culture and the anxiety disorders: recommendations for DSM-V. *DA Depression and Anxiety, 27*(2), 212-229.

References

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry, 22*, 141-173.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association, 83*(404), 1198-1202. doi: 10.2307/2290157

Maeda, F., & Nathan, J. H. (1999). Understanding taijin kyofusho through its treatment, Morita therapy. *Journal of Psychosomatic Research, 46*(6), 525-530.

Mahalanobis, P. C. (1936). *On the generalised distance in statistics.* Paper presented at the Proceedings National Institute of Science, India.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224.

Matsumoto, D., & Yoo, S. H. (2006). Toward a New Generation of Cross-Cultural Research. *Perspectives on Psychological Science, 1*(3), 234-250. doi: 10.1111/j.1745-6916.2006.00014.x

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety1. *Behaviour Research and Therapy, 36*(4), 455-470. doi: http://dx.doi.org/10.1016/S0005-7967(97)10031-6

McKay, R. B., & de la Puente, M. (1996). Cognitive testing of racial and ethnic questions for the CPS supplement. *Monthly Lab. Rev., 119*, 8.

McNally, R. J., Cassiday, K. L., & Calamari, J. E. (1990). Taijin-kyofu-sho in a black American woman: Behavioral treatment of a "culture-bound" anxiety disorder. *Journal of Anxiety Disorders, 4*(1), 83-87. doi: http://dx.doi.org/10.1016/0887-6185(90)90025-5

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*: Sage.

References

Miller, K. (2002). *The role of social location in question response: Rural poor experience answering general health questions.* . Paper presented at the Annual Meeting of the American Association for Public Opinion research, St. Pete Beach, FL.

Miller, K. (2009). *Cognitive interviewing*. Paper presented at the Workshop on Question Evaluation Methods, Hyattsville, Maryland.

Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (2014). *Cognitive Interviewing Methodology*: John Wiley & Sons.

Mollica, R. F., Cardozo, B. L., Osofsky, H. J., Raphael, B., Ager, A., & Salama, P. (2004). Mental health in complex emergencies. *The Lancet, 364*(9450), 2058-2067.

Mukai, Y. (2002). 対人不安の生起・維持プロセスの理論モデルに関する展望: 回避的行動と自己愛, 他者への関心の葛藤という観点から *東京大学大学院教育学研究科紀要* (Vol. 41, pp. 312-326). Tokyo: University of Tokyo.

Muniz, J., Hambleton, R. K., & Xing, D. (2001). Small Sample Studies to Detect Flaws in Item Translations. *International Journal of Testing, 1*(2), 115-135. doi: 10.1207/S15327574IJT0102_2

Nagata, T., van Vliet, I., Yamada, H., Kataoka, K., Iketani, T., & Kiriike, N. (2006). An open trial of paroxetine for the “offensive subtype” of taijin kyofusho and social anxiety disorder. *Depression and Anxiety, 23*(3), 168-174. doi: 10.1002/da.20153

Nakamura, K., Kitanishi, K., Miyake, Y., Hashimoto, K., & Kubota, M. (2002). The neurotic versus delusional subtype of taijin-kyofu-sho: their DSM diagnoses. *Psychiatry and Clinical Neurosciences, 56*(6), 595-601. doi: 10.1046/j.1440-1819.2002.01061.x

References

Norasakkunkit, V., Kitayama, S., & Uchida Y. (2012). Social anxiety and holistic cognition: Selffocused social anxiety in the United States and Other-focused social anxiety in Japan. *Journal of Cross-Cultural Psychology*, *43*, 742-757.

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: what are they and how can we know? *Psychological Bulletin, 131*(5), 763-784. doi: 10.1037/0033-2909.131.5.763

Ono, Y., Yoshimura, K., Yamauchi, K., Asai, M., Young, J., Fujuhara, S., & Kitamura, T. (2001). Taijin Kyofusho in a Japanese Community Population. *Transcultural psychiatry, 38*(4), 506-514. doi: 10.1177/136346150103800408

Piedmont, R. (2014). Inter-item Correlations. In A. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3303-3304): Springer Netherlands.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results. *Sociological Methodology, 24*(5), 73-104.

Prince, R., & Tcheng-Laroche, F. (1987). Culture-bound syndromes and international disease classifications. *Culture, Medicine and Psychiatry, 11*(1), 3-52.

Rector, N., Kocovski, N., & Ryder, A. (2006). Social Anxiety and the Fear of Causing Discomfort to Others. *Cognitive Therapy and Research, 30*(3), 279-296. doi: 10.1007/s10608-006-9050-9

Ridolfo, H., & Schoua-Glusberg, A. (2011). Analyzing cognitive interview data using the constant comparative method of analysis to understand cross-cultural patterns in survey data. *Field Methods, 23*(4), 420-438.

Rothgeb, J., Willis, G., & Forsyth, B. (2007). Questionnaire pretesting methods: Do different techniques and different organizations produce similar results? *Bulletin de Méthodologie Sociologique, 96*(1), 5-31.

References

Rusticus, S. A., Hubley, A. M., & Zumbo, B. D. (2004). *Cross-national comparability of the Rosenberg self-esteem scale.* Paper presented at the Poster presented at the annual convention of the American Psychological Association, Honolulu, HI.

Ryder, A. G., Ban, L. M., & Chentsova-Dutton, Y. E. (2011). Towards a Cultural–Clinical Psychology. *Social and Personality Psychology Compass, 5*(12), 960-975. doi: 10.1111/j.1751-9004.2011.00404.x

Ryder, A. G., Yang, J., Zhu, X., Yao, S., Yi, J., Heine, S. J., & Bagby, R. M. (2008). The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? *Journal of Abnormal Psychology, 117*(2), 300-313. doi: 10.1037/0021-843x.117.2.300

Schmeiser, C. B. (1982). Use of experimental design in statistical item bias studies. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 64-96). Baltimore, MD: Johns Hopkins University Press.

Singelis, T. M. (1994). The Measurement of Independent and Interdependent Self-Construals. *Personality and Social Psychology Bulletin, 20*(5), 580-591. doi: 10.1177/0146167294205014

Stein, D. J. (2009). Social anxiety disorder in the West and in the East. *Annals of Clinical Psychiatry, 21*(2), 109-117.

Sudman, S. (1983). Applied Sampling *Handbook of Survey research* (pp. 145-194): Academic Press Inc.

Sue, D. W., & Sue, D. (1977). Barriers to effective cross-cultural counseling. *Journal of Counseling Psychology, 24*(5), 420-429. doi: 10.1037/0022-0167.24.5.420

References

Suzuki, K., Takei, N., Kawai, M., Minabe, Y., & Mori, N. (2003). Is Taijin Kyofusho a Culture-Bound Syndrome? *American Journal of Psychiatry, 160*(7), 1358-1358. doi: doi:10.1176/appi.ajp.160.7.1358

Takahashi, T. (1989). Social phobia syndrome in Japan. *Comprehensive Psychiatry, 30*(1), 45-52.

Tarumi, S., Ichimiya, A., Yamada, S., Umesue, M., & Kuroki, T. (2004). Taijin Kyofusho in university students: patterns of fear and predispositions to the offensive variant. *Transcultural psychiatry, 41*(4), 533-546.

Tashakkori, A., & Creswell, J. W. (2008). Editorial: Mixed Methodology Across Disciplines. *Journal of Mixed Methods Research, 2*(1), 3-6. doi: 10.1177/1558689807309913

Tekera, G. T., & Doganb, N. (2015). The Effects of Testlets on Reliability and Differential Item Functioning. *Educational Sciences: Theory & Practice, 4*, 1-12.

Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin, 99*(1), 118.

Thoits, P. A. (1995). Stress, Coping, and Social Support Processes: Where Are We? What Next? *Journal of Health and Social Behavior, 35*, 53-79. doi: 10.2307/2626957

Thompson, S. B. (2010). Grounded theory - sample size and validity. Retrieved from: http://www.buseco.monash.edu.au/research/studentdocs/mgt.pdf, 2002.

Tourangeau, R. (1984). Cognitive sciences and survey methods. *Cognitive aspects of survey methodology: Building a bridge between disciplines*, 73-100.

Tseng, W.-S., Bartocci, G., Rovera, G., Infante, V., & De Luca, V. The future of cultural psychiatry.

References

Uchida, Y., Norasakkunkit, V., & Kitayama, S. (2004). Cultural constructions of happiness: theory and emprical evidence. *Journal of Happiness Studies, 5*(3), 223-239. doi: 10.1007/s10902-004-8785-9

van de Vijver, F. J., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. *Adapting educational and psychological tests for cross-cultural assessment*, 39-63.

van de Vijver, F. J. R., & Chasiotis, A. (2010). Making Methods Meet: Mixed Designs in Cross-Cultural Research *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 455-473): John Wiley & Sons, Inc.

Vriends, N., Pfaltz, M. C., Novianti, P., & Hadiyono, J. (2013). Taijin Kyofusho and Social Anxiety and Their Clinical Relevance in Indonesia and Switzerland. *Frontiers in Psychology, 4*, 3. doi: 10.3389/fpsyg.2013.00003

Watanabe, H., Okazaki, Y., & Crystal, D. S. (1998). 日本の大学生における 「はずかしさ」の生起に関する研. *愛媛大学教育学部紀要, 44*(2), 17-381998.

Willis, G. B. (2015). *Analysis of the Cognitive Interview in Questionnaire Design* (1 ed.). New York: Oxford University Press.

Willis, G. B., 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: official journal of the Society for Research on Nicotine and Tobacco, 10*(6), 1075.

Willis, G. B., Lawrence, D., Thompson, F., Kudela, M., Levin, K., & Miller, K. (2005). *The use of cognitive interviewing to evaluate translated survey questions: Lessons learned.* Paper

presented at the Proceedings of the Federal Committee on Statistical Methodology Research Conference; Arlington, VA.

Willis, G. B., & Zahnd, E. (2007). Questionnaire design from a cross-cultural perspective: an empirical investigation of Koreans and non-Koreans. *Journal of Health Care for the Poor and Underserved, 18*(4), 197-217.

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*: Sage Publications.

Willis, G. B., & Lessler, J. T. (1999). Question appraisal system QAS-99. *Rockville, MD: Research Triangle Institute*.

Willis, G. B., & Miller, K. (2011). Cross-Cultural Cognitive Interviewing: Seeking Comparability and Enhancing Understanding. *Field Methods, 23*(4), 331-341. doi: 10.1177/1525822x11416092

Willis, G. B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*.

Yamashita, I. (1970). 対人恐怖について. *精神医学, 12*, 5-14.

Yap, P. M. (1965). KORO--A CULTURE-BOUND DEPERSONALIZATION SYNDROME. *British Journal of Psychiatry, 111*, 43-50.

Zhang, A., Yu, L., Draguns, J., Zhang, J., & Tang, D. (2000). Sociocultural contexts of anthropophobia: a sample of Chinese youth. *Social Psychiatry and Psychiatric Epidemiology, 35*(9), 418-426.

Zhou, B., Lacroix, F., Sasaki, J., Peng, Y., Wang, X., & Ryder, A. G. (2014). Unpacking Cultural Variations in Social Anxiety and the Offensive-Type of Taijin Kyofusho

Through the Indirect Effects of Intolerance of Uncertainty and Self-Construals. *Journal of Cross-Cultural Psychology, 45*(10), 1561-1578. doi: 10.1177/0022022114548483

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). *Ottawa: National Defense Headquarters*.

Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. *The concept of validity: Revisions, new directions and applications*, 65-82.

Table 1

*The Taijin Kyofu Sho Questionnaire (TKSQ)*

| Item | | Question | Answer (4-point rating) |
|---|---|---|---|
| *Offensive Symptoms* | | | |
| *E-TKS* | 2a | If my facial expression were to stiffen, | of embarrassing myself. |
| | 6a | If I were to emit body odors in the presence of others, | of embarrassing myself. |
| | 7a | If I were to stare at someone's body part(s), | of embarrassing myself. |
| | 8a | If I were to release intestinal gas in the presence of others, | of embarrassing myself. |
| | 10a | Due to my physical appearance, | of embarrassing myself. |
| *D-TKS* | 2b | If my facial expression were to stiffen, | of making someone uncomfortable. |
| | 6b | If I were to emit body odors in the presence of others, | of making someone uncomfortable. |
| | 7b | If I were to stare at someone's body part(s), | of making someone uncomfortable. |
| | 8b | If I were to release intestinal gas in the presence of others, | of making someone uncomfortable. |
| | 10b | Due to my physical appearance, | of making someone uncomfortable. |
| *O-TKS* | 2c | If my facial expression were to stiffen, | of offending someone. |
| | 6c | If I were to emit body odors in the presence of others, | of offending someone. |
| | 7c | If I were to stare at someone's body part(s), | of offending someone. |
| | 8c | If I were to release intestinal gas in the presence of others, | of offending someone. |
| | 10c | Due to my physical appearance, | of offending someone. |
| *Common Symptoms* | | | |
| *E-CS* | 1a | If I were to blush in front of others, | of embarrassing myself. |
| | 3a | If my head, hands and/or feet were to tremble in front of others, | of embarrassing myself. |
| | 4a | If my voice were to tremble in front of others, | of embarrassing myself. |
| | 5a | If I were to sweat or perspire in front of others, | of embarrassing myself. |
| | 9a | If I were to make eye contact with someone, | of embarrassing myself. |
| *D-CS* | 1b | If I were to blush in front of others, | of making someone uncomfortable. |
| | 3b | If my head, hands and/or feet were to tremble in front of others, | of making someone uncomfortable. |
| | 4b | If my voice were to tremble in front of others, | of making someone uncomfortable. |
| | 5b | If I were to sweat or perspire in front of others, | of making someone uncomfortable. |
| | 9b | If I were to make eye contact with someone, | of making someone uncomfortable. |
| *O-CS* | 1c | If I were to blush in front of others, | of offending someone. |
| | 3c | If my head, hands and/or feet were to tremble in front of others, | of offending someone. |
| | 4c | If my voice were to tremble in front of others, | of offending someone. |
| | 5c | If I were to sweat or perspire in front of others, | of offending someone. |
| | 9c | If I were to make eye contact with someone, | of offending someone. |

*Note*. Respondents rated the severity of fear for each item on a 4-point likert scale (1 = I would not be fearful at all; 2 = I would be mildly fearful; 3 = I would be moderately fearful; 4 = I would be extremely fearful.

Table 2

Table 2

*Demographics of Japanese and Euro-Canadian samples (Dataset 1)*

| | Japanese ($n$ = 190 ) | | Euro-Canadian ($n$ = 78 ) | |
|---|---|---|---|---|
| Variables | *N* | | *N* | |
| Gender | | | | |
| Male | 94 | 49.5 | 9 | 11.5 |
| Female | 86 | 5.5 | 69 | 88.5 |
| Field of Study | | | | |
| Humanities and Social Sciences | 186 | 97.9 | 75 | 96.1 |
| | Mean | (SD) | Mean | (SD) |
| Age | 18.76 | .84 | 22.32 | 4.85 |

Table 3

Table 3
*Demographics of Japanese and Euro-Canadian samples (Dataset 2)*

| | Japanese ($n$ = 124 ) | | Euro-Canadian ($n$ = 114 ) | |
|---|---|---|---|---|
| Variables | *N* | *%* | *N* | *%* |
| Gender | | | | |
| Male | 53 | 42.7 | 97 | 85.1 |
| Female | 63 | 54.8 | 15 | 13.2 |
| Missing | 3 | 2.4 | 2 | 1.8 |
| Field of Study | | | | |
| Humanities and Social Sciences | 124 | 100.0 | 112 | 98.2 |
| Other | | | 2 | 0.8 |
| Declared Major | N/A | N/A | | |
| Psychology | | | | |
| Exercise Science | N/A | N/A | | |
| | Mean | (SD) | Mean | (SD) |
| Age | 21.03 | 4.62 | 22.12 | 4.87 |

Table 4
*Demographics of Japanese and Euro-Canadian samples (CI)*

| | Japanese ($n$ = 12 ) | | Euro-Canadian ($n$ = 9 ) | |
|---|---|---|---|---|
| Variables | *N* | % | *N* | % |
| Gender | | | | |
| Male | 7 | 58.3 | 0 | 0.0 |
| Female | 5 | 41.7 | 9 | 100.0 |
| Field of Study | | | | |
| Humanities and Social Sciences | 12 | 100.0 | 9 | 100.0 |
| Declared Major | N/A | N/A | 7 | 78.8 |
| Psychology | | | | |
| Exercise Science | N/A | N/A | 2 | 22.2 |
| | Mean | (SD) | Mean | (SD) |
| Age | 18.67 | .65 | 22.00 | 4.00 |

Table 5
*Reliability of Taijin Kyofu Sho Questionnaire (TKSQ) Subscales by Group*

| | Japanese | | Euro-Canadian | |
|---|---|---|---|---|
| | α | Inter-Item *r* | α | Inter-Item *r* |
| TKSQ Total | .95 | .37 | .94 | .39 |
| TKSQ Subscales | | | | |
| *Common Symptoms* | | | | |
| Embarrassment items (E-CS) | .71 | .33 | .76 | .38 |
| Discomfort items (D-CS) | .73 | .34 | .78 | .41 |
| Offend items (O-CS) | .76 | .40 | .76 | .39 |
| *Offensive Symptoms* | | | | |
| Embarrassment items (E-TKS) | .77 | .40 | .72 | .34 |
| Discomfort items (D-TKS) | .71 | .32 | .71 | .34 |
| Offend items (O-TKS) | .76 | .39 | .79 | .45 |

*Note.* α = Cronbach's α. E-CS = fear of embarrassing oneself due to common symptoms; D-CS = fear of discomfort to others due to common symptoms; O-CS = fear of offending others due to common symptoms; E-TKS = fear of embarrassing oneself due to offensive TKS symptoms; D-TKS = fear of discomfort to others due to TKS symptoms; O-TKS = fear of offending others due to offensive symptoms.

Table 6

Table 6

*Summary of the Differential Item Functioning analysis by TKSQ subscale*

| | Step #1 | | Step #2 | | Step #3 | | DIF test | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total Score | | Uniform | | Non-Uniform | | | | |
| Item | $R^2$ | $\chi^2$ | $R^2$ | $\chi^2$ | $R^2$ | $\chi^2$ | $\chi^2(2)$ | $p$ | $R^2$ |
| *E-TKS* | | | | | | | | | |
| 2a | .372 | 111.149 | .375 | 112.319 | .377 | 112.956 | 1.81 | .41 | .01 |
| 6a | .252 | 71.206 | .357 | 107.466 | .361 | 109.094 | 37.89 | .01 | .11 |
| 7a | .470 | 155.809 | .473 | 157.216 | .473 | 157.271 | 1.46 | .49 | .00 |
| 8a | .419 | 132.426 | .538 | 186.118 | .554 | 194.058 | 61.63 | .01 | *.13* |
| 10a | .265 | 75.277 | .287 | 82.701 | .290 | 83.812 | 8.53 | .02 | .03 |
| *D-TKS* | | | | | | | | | |
| 2b | .311 | 86.179 | .359 | 102.427 | .360 | 102.457 | 16.28 | .01 | .05 |
| 6b | .491 | 163.595 | .619 | 229.712 | .620 | 23.411 | 66.82 | .01 | *.13* |
| 7b | .495 | 164.931 | .548 | 19.318 | .548 | 19.382 | 25.45 | .01 | .05 |
| 8b | .395 | 122.781 | .396 | 123.042 | .442 | 141.728 | 18.95 | .01 | .05 |
| 10b | .560 | 191.115 | .562 | 192.234 | .565 | 193.814 | 2.70 | .26 | .01 |
| *O-TKS* | | | | | | | | | |
| 2c | .434 | 135.494 | .445 | 14.229 | .452 | 142.936 | 7.44 | .03 | .02 |
| 6c | .553 | 194.128 | .568 | 202.036 | .568 | 202.112 | 7.98 | .02 | .02 |
| 7c | .364 | 108.992 | .400 | 122.888 | .402 | 123.681 | 14.69 | .01 | .04 |
| 8c | .484 | 161.409 | .485 | 161.998 | .503 | 17.498 | 9.09 | .02 | .02 |
| 10c | .554 | 19.037 | .576 | 201.120 | .576 | 201.165 | 11.13 | .01 | .02 |
| *E-CS* | | | | | | | | | |
| 1a | .334 | 99.285 | .334 | 99.315 | .340 | 101.400 | 2.11 | .35 | .01 |
| 3a | .484 | 157.397 | .490 | 159.872 | .490 | 16.293 | 2.90 | .24 | .01 |
| 4a | .506 | 169.950 | .509 | 171.308 | .512 | 173.043 | 3.09 | .22 | .01 |
| 5a | .460 | 149.855 | .574 | 205.136 | .574 | 205.137 | 55.28 | .01 | .11 |
| 9a | .415 | 131.709 | .644 | 247.061 | .645 | 247.853 | 116.14 | .01 | *.23* |
| *D-CS* | | | | | | | | | |
| 1b | .389 | 119.664 | .439 | 14.042 | .450 | 144.517 | 24.85 | .01 | .06 |
| 3b | .587 | 21.190 | .597 | 215.671 | .603 | 219.444 | 9.25 | .01 | .02 |
| 4b | .319 | 93.857 | .399 | 123.648 | .402 | 125.063 | 31.21 | .01 | .08 |
| 5b | .473 | 153.514 | .501 | 166.280 | .501 | 166.332 | 12.82 | .01 | .03 |
| 9b | .417 | 121.328 | .434 | 127.475 | .434 | 127.543 | 6.22 | .05 | .02 |
| *O-CS* | | | | | | | | | |
| 1c | .367 | 11.210 | .412 | 127.755 | .416 | 129.311 | 19.10 | .01 | .05 |
| 3c | .451 | 143.585 | .511 | 17.135 | .515 | 171.906 | 28.32 | .01 | .06 |
| 4c | .579 | 202.825 | .601 | 214.461 | .605 | 216.678 | 13.85 | .01 | .03 |
| 5c | .576 | 202.614 | .576 | 202.884 | .577 | 203.455 | .84 | .66 | .00 |
| 9c | .460 | 144.686 | .491 | 157.900 | .495 | 159.405 | 14.72 | .01 | .03 |

*Note*. $p \leq .01$. Italicized values (6b, 8a, and 9a) have the effect sizes ($R^2$ difference) $\leq .13$

Table 7

Table 7

*Means and Standard Deviations of All Scales by Group*

| | Japanese (*N*=124) | | Euro-Canadian (*N*=114) | |
|---|---|---|---|---|
| | *M* | (SD) | *M* | (SD) |
| TKSQ | 2.36 | .69 | 1.77 | .55 |
| TKS | 3.512 | 1.21 | 2.30 | 1.24 |
| SIAS | 1.79 | .74 | 1.26 | .79 |
| Independent Self-Construal | 4.58 | .68 | 5.03 | .69 |
| Interdependent Self-Construal | 4.54 | .64 | 4.72 | .66 |

Note. TKSQ = Taijin Kyofu Sho Questionnaire (Choy et al., 2008); TKS = Taijin Kyofusho Scale (Kleinknecht et al., 1994); SIAS = Social Interaction Anxiety Scale (Mattick & Clarke, 1998); and Independent Self-Construal and Interdependent Self-Construal are subscales of Self-Construal Scale (Singelis, 1994).

Table 8

Table 8

*Reliability of All Scales by Group*

| | Japanese | | Euro-Canadian | |
|---|---|---|---|---|
| | α | Inter-Item *r* | α | Inter-Item *r* |
| TKSQ | .96 | .46 | .96 | .46 |
| TK-S | .96 | .45 | .97 | .53 |
| SIAS | .91 | .34 | .94 | .45 |
| Independent SC | .74 | .17 | .77 | .19 |
| Interdependent SC | .72 | .16 | .77 | .19 |

Note. α = Cronbach's α. Inter-term *r* = average inter-item correlations.

*Note*. $p \leq .01$. Italicized values (6b, 8a, and 9a) have the effect sizes ($R^2$ difference) ≤ .13

Table 9

Table 9
*Correlations Among Variables*

| Variables | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. TKSQ | - | .55** | .50** | .00 | .23* |
| 2. TK-S | .63** | - | .78** | -.26** | .17 |
| 3. SIAS | .47** | .67** | - | -.35** | .15 |
| 4. Independent Self-Construal | -.13 | -.21* | -.36** | - | .26* |
| 5. Interdependent Self-Construal | .33** | .39** | .21* | .10 | - |

Note. *$p < .05$; **$p < .01$. Upper diagonal contains coefficients for Euro-Canadian participants, and lower diagonal for Japanese participants.

Table 10

Table 10

*Frequency of symptoms reviewed by Cultural Group*

| | All groups | | Japanese | | Euro-Canadian | |
|---|---|---|---|---|---|---|
| Symptom domains | *n* | | *n* | | *n* | |
| *Offensive Symptoms* | | | | | | |
| 2. Stiff facial expression | 5 | 12.0 | 3 | 12.5 | 2 | 11.1 |
| 6. Body odor | 11 | 26.2 | 6 | 25.0 | 5 | 27.8 |
| 7. Staring | 3 | 7.1 | 1 | 4.2 | 2 | 11.1 |
| 8. Intestinal gas | 13 | 31.0 | 6 | 25.0 | 7 | 38.9 |
| 10. Physical appearance | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| *Common Symptoms* | | | | | | |
| 1. Blushing | 2 | 4.8 | 2 | 8.3 | 0 | 0.00 |
| 3. Body trembling | 1 | 2.4 | 0 | 0.00 | 1 | 5.6 |
| 4. Voice trembling | 4 | 9.5 | 3 | 12.5 | 1 | 5.6 |
| 5. Sweating | 3 | 7.1 | 3 | 12.5 | 0 | 0.00 |
| 9. Eye contact | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Total | 42 | 100.0 | 24 | 100.0 | 18 | 100.0 |

*Note.* 12 Japanese and 9 Euro-Canadians participants reviewed 2 domains each. Each symptom domain contains three items (i.e., a. fear of embarrassment, b. discomfort to someone, and c. offending someone).

Table 11

Table 11
*Frequency of QAS-04 Problem Codes by Cultural Group*

| Code | All groups | | Japanese | | Euro-Canadian | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % |
| 1. Reading | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 2. Instructions | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 3. Clarity | 38 | 33.0 | 16 | 22.5 | 22 | 50.0 |
| 4. Assumptions | 35 | 30.4 | 20 | 28.2 | 15 | 34.1 |
| 5. Knowledge/Memory | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 6. Sensitivity/Bias | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 7. Response Categories | 20 | 17.4 | 13 | 18.3 | 7 | 15.9 |
| 8. Cross-Cultural Considerations | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 9. Translations | 22 | 19.1 | 22 | 31.0 | 0 | 0.0 |
| 10. Cross-Questions | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 11. Other Problems | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| Total Codes | 115 | 100.0 | 71 | 100 | 44 | 100.0 |

# Appendix A
## Taijin Kyofu Sho Questionnaire (Japanese)

Appendix A: TKSQ (Japanese)

次の文章を読んで、１～４の選択肢のうち、あなたに最も当てはまるものを選んでください。

■１から９は下記の選択肢から答えてください。

| |
|---|
| １＝**まったく**恐れないだろう ２＝**少し**恐れるだろう ３＝**まあまあ**恐れるだろう ４＝**かなり**恐れるだろう |

| | | |
|---|---|---|
| **1)** | **もし人前で赤面してしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |
| **2)** | **もし表情がこわばってしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |
| **3)** | **もし人前で頭、手、足などが震えてしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |
| **4)** | **もし人前で声が震えてしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |
| **5)** | **もし人前で汗をかいてしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |
| **6)** | **もし人がいるときに体臭を放ってしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |
| **7)** | **もし他の人の体の部分をじっと見てしまったら、** | |
| a) | 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) | 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) | 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |

**8) もし人がいるときにおならをしてしまったら、**

| | |
|---|---|
| a) 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |

**9) もし人と目を合わせてしまったら、**

| | |
|---|---|
| a) 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |

**■１０は下記の選択肢から選んでください。**

１ = **まったく**恐れない　２ =**少し**恐れる　３ =**まあまあ**恐れる　４ =**かなり**恐れる

**10) 自分の見た目のせいで、**

| | |
|---|---|
| a) 自分自身が当惑してしまうことを・・・ | １ ２ ３ ４ |
| b) 誰かに気まずい思いをさせてしまうことを・・・ | １ ２ ３ ４ |
| c) 誰かを不快にさせてしまうことを・・・ | １ ２ ３ ４ |

# Appendix B
# Taijin Kyofu Sho Questionnaire (English)

Please read the following statements and circle one of the options below:

| (1) I would **not** be fearful at all | (2) I would be **mildly** fearful | (3) I would be **moderately** fearful | (4) I would be **extremely** fearful |
|---|---|---|---|

**11) If I were to blush in front of others,**

d) _____ of embarrassing myself. 1 2 3 4

e) _____ of making someone uncomfortable. 1 2 3 4

f) _____ of offending someone. 1 2 3 4

**12) If my facial expression were to stiffen,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**13) If my head, hands and/or feet were to tremble in front of others,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**14) If my voice were to tremble in front of others,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**15) If I were to sweat or perspire in front of others,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**16) If I were to emit body odors in the presence of others,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**17) If I were to stare at someone's body part(s),**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**18) If I were to release intestinal gas in the presence of others,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

**19) If I were to make eye contact with someone,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

Please read the following statements and circle one of the options below:

| (1) I am **not** be fearful at all | (2) I am **mildly** fearful | (3) I am **moderately** fearful | (4) I am **extremely** fearful |
|---|---|---|---|

**20) Due to my physical appearance,**

a. _____ of embarrassing myself. 1 2 3 4

b. _____ of making someone uncomfortable. 1 2 3 4

c. _____ of offending someone. 1 2 3 4

Appendix C
Question Appraisal System - 04 (QAS -04)

Instructions. Use one form for each question to be reviewed. In reviewing each question:

1. Write or type in question number. Attach question.
2. Proceed through the form – Circle or highlight YES or No for each problem type.
3. Whenever a YES is circled, write detailed notes on this form that describe the problem.

| | | |
|---|---|---|
| **Step 1. Reading: Determine if it is difficult for the interviewers to read the question uniformly to all respondents or if the reading level is appropriate.** | | |
| 1a. What to read: Interviewer may have difficulty determining what parts of the question should be read. | **Yes** | **No** |
| 1b. Missing information: Information the interviewer needs to administer the question is not contained in the question. | **Yes** | **No** |
| 1c. How to read: Question is not fully scripted and therefore difficult to read. | **Yes** | **No** |
| **Step 2. Instructions: Look for problems with any introductions, instructions or explanations from the respondent's point of view.** | | |
| 2a. Conflicting or inaccurate instructions, introductions or explanations. | **Yes** | **No** |
| 2b. Complicated instructions, introductions or explanations. | **Yes** | **No** |
| 2c. Missing or inconsistent instructions for don't know and refused answers. | **Yes** | **No** |
| **Step 3. Clarity: Identify problems related to communicating the intent or meaning of the question to the respondent.** | | |
| 3a. Wording: Question is lengthy, awkward, ungrammatical or contains complicated syntax. | **Yes** | **No** |
| 3b. Technical term(s) are undefined, unclear or complex. | **Yes** | **No** |
| 3c. Vague: There are multiple ways to interpret the question or to decide what is to be included or excluded. | **Yes** | **No** |
| 3d. Reference periods are missing, not well specified, or in conflict. | **Yes** | **No** |
| 3e. Passive voice: Question is written in passive voice. Active voice is clearer both in source language and in translation. | **Yes** | **No** |
| **Step 4. Assumptions: Determine if there are problems with assumptions made or the underlying logic.** | | |
| 4a. Inappropriate assumptions are made about the respondent or about his/her living situation. | **Yes** | **No** |
| 4b. Assumes constant behavior or experience for situations that vary. | **Yes** | **No** |
| 4c. Double-barreled: Contains more than one implicit question. | **Yes** | **No** |
| **Step 5. Knowledge/Memory: Check whether respondents are likely to not know or have trouble remembering information.** | | |
| 5a. Knowledge may not exist: Respondent is unlikely to know the answer to a factual question. | **Yes** | **No** |
| 5b. Attitude may not exist: Respondent is unlikely to have formed the attitude being asked about. | **Yes** | **No** |
| 5c. Recall failure: Respondent may not remember the information asked for. | **Yes** | **No** |
| 5d. Computation problem: The question requires a difficult mental calculation. | **Yes** | **No** |
| **Step 6. Sensitivity/Bias: Assess questions for sensitive nature or wording, and for bias** | | |
| 6a. Sensitive content (general): The question asks about a topic that is embarrassing, very private, or that involves illegal behavior. If question will be applied across cultures, it may be sensitive in some cultures but not others. | **Yes** | **No** |
| 6b. Sensitive wording (specific): Given that the general topic is sensitive, the wording should be improved to minimize sensitivity. | **Yes** | **No** |
| 6c. Socially acceptable response is implied by the question. If question will be | **Yes** | **No** |

| | | |
|---|---|---|
| applied across cultures, social acceptability could vary. | | |
| **Step 7. Response Categories: Assess the adequacy of the range of responses to be recorded.** | | |
| 7a. Open-ended question that is inappropriate or difficult. | **Yes** | **No** |
| 7b. Mismatch between question and response categories. | **Yes** | **No** |
| 7c. Technical term(s) are undefined, unclear or complex. | **Yes** | **No** |
| 7d. Vague response categories are subject to multiple interpretations. | **Yes** | **No** |
| 7e. Overlapping response categories. | **Yes** | **No** |
| 7f. Missing eligible responses in response categories. | **Yes** | **No** |
| 7g. Illogical order of response categories. | **Yes** | **No** |
| **Step 8. Cross-Cultural Considerations: Assess questions for inappropriate or ineffective cross-cultural references.** | | |
| 8a. Reference periods: The reference period uses seasons, American MM/DD/ YYYY format, or may be otherwise ambiguous or unusual in other cultures. | **Yes** | **No** |
| 8b. Knowledge may not exist: Respondent is unlikely to know the answer to a factual question because he/she is not familiar with the American culture. Example: health insurance. | **Yes** | **No** |
| 8c. Measuring units: Measuring units are from English system. If surveying Latin Americans or western European populations, the metric system should be used. | **Yes** | **No** |
| 8d. Assumptions: The question includes culturally inappropriate assumptions or graphics. All statements related to sports, drugs, foods, drinks, activities, meal time, music, family ties, holidays, religion, books, magazines, school system, health system and history should be evaluated. | **Yes** | **No** |
| 8e. Response categories: There is no equivalent concept or rating scale in foreign language. Avoid rating scales with more than five categories. | **Yes** | **No** |
| 8f. Name format: Response categories lack a space for other types of names. | **Yes** | **No** |
| 8g. Courtesy and politeness can differ between cultures. Consider adding a 'Please' before commands like, 'Do not include …,' 'Mark every …,' 'List all …' Consider using 'could' instead of 'should' if possible. Some commands or instructions might be perceived as rude, and respondents could change their attitude towards participating. | **Yes** | **No** |
| **Step 9. Potential Translation Problems: Identify problematic question characteristics.** | | |
| 9a. Double negatives: This type of construction is hard to translate and can easily cause misunderstandings in other languages. | **Yes** | **No** |
| 9b. Idioms: Many idioms do not have an equivalent in other languages. | **Yes** | **No** |
| 9c. Acronyms: The acronyms have no meaning in other languages. Consider providing an explanation with the acronym. | **Yes** | **No** |
| 9d. Unclear use of the term 'you': 'You' not defined as plural, singular, feminine, masculine, formal, informal—a necessary step for translation. | **Yes** | **No** |
| 9e. Time adverbs: Question or response categories use adverbs to describe time, such as recently, lately, usually. Consider specifying time frame with number of days, weeks, etc. | **Yes** | **No** |
| 9f. No equivalent term or concept in foreign language: Text may require an additional explanation. | **Yes** | **No** |
| 9g. References applicable only to English: Toll-free numbers, Web sites, contact information, books and other references are only available in the source language. Consider verifying which services or references are available in the target language. Also consider using numbers instead of letter on phone numbers. | **Yes** | **No** |

| | | |
|---|---|---|
| 9h. Adjectives modifying other adjectives: Using adjectives to modify other adjectives (e.g., 'house warming party,' which must be literally translated as 'a party in celebration of the purchase of a home in which guests take presents for the new home owner') is an uncommon grammatical usage in languages other than English. Consider paraphrasing and clearly define each term. | **Yes** | **No** |
| **Step 10. Cross-Question: Look for cross-question problems in the entire questionnaire.** | | |
| 10a. Question placement: The questions are not positioned in the most adequate section or order. | **Yes** | **No** |
| 10b. Data collection mode: Sensitive question may be more effective if it was administered through another data collection mode. | **Yes** | **No** |
| 10c. Inconsistency with other questions: Wording or response categories lack consistency. | **Yes** | **No** |
| 10d. Content of previous question affects meaning: Does the content of the previous question or section affect the interpretation of the current question. | **Yes** | **No** |
| 10e. Skip pattern problem: Skip pattern is illogical or inadequate. | **Yes** | **No** |
| 10f. Formatting: Layout or formatting is difficult to follow. | **Yes** | **No** |
| **Step 11. Other Problems** | | |
| 11a. Questions contain irrelevant information. | **Yes** | **No** |
| 11b. Inappropriate reading level. | **Yes** | **No** |
| 11c. Other problems. | **Yes** | **No** |