



A qualitative assessment of cross-cultural adaptation of intermediate measures for schizophrenia in multisite international studies

Jodi M. Gonzalez ^{a,*}, Maureen Rubin ^b, Megan M. Fredrick ^a, Dawn I. Velligan ^a

^a Department of Psychiatry, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA

^b Department of Social Work, University of Texas at San Antonio, San Antonio, TX, USA

ARTICLE INFO

Article history:

Received 13 January 2012

Received in revised form

23 October 2012

Accepted 25 October 2012

Keywords:

Co-primary measures

Test adaptation

Functional capacity measures

Cross-cultural assessment

International clinical trials

ABSTRACT

In this substudy of the Measurement and Treatment Research to Improve Cognition in Schizophrenia we examined qualitative feedback on the cross-cultural adaptability of four intermediate measures of functional outcome (Independent Living Scales, UCSD Performance-Based Skills Assessment, Test of Adaptive Behavior in Schizophrenia, and Cognitive Assessment Interview). Feedback was provided by experienced English-fluent clinical researchers at 31 sites in eight countries familiar with medication trials. Researchers provided feedback on test subscales and items which were rated as having adaptation challenges. They noted the specific concern and made suggestions for adaptation to their culture. We analyzed the qualitative data using a modified Grounded Theory approach guided by the International Testing Commission Guidelines model for test adaptation. For each measure except the Cognitive Assessment Interview (CAI), the majority of subscales were reported to require major adaptations in terms of content and concepts contained in the subscale. In particular, social, financial, transportation and health care systems varied widely across countries—systems which are often used to assess performance capacity in the U.S. We provide suggestions for how to address future international test development and adaptation.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Cognitive deficits and impairments in functional outcome are core features of schizophrenia (Gold and Harvey, 1993; Green, 1996; Velligan et al., 1997). The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative was designed to encourage the development of pharmaceutical agents to enhance cognitive functioning in schizophrenia by developing a process for medication to receive an indication for cognitive enhancement (Nuechterlein et al., 2005; Buchanan et al., 2005). Improvement in performance on neuropsychological tests and on a co-primary measure of functional outcome is required (Buchanan et al., 2005; Green et al., 2008).

As part of the MATRICS initiative (Green et al., 2011) the Independent Living Scales (Loeb, 1996), the UCSD Performance-based Skills Assessment (Patterson et al., 2001), the Test of Adaptive Behavior in Schizophrenia (Velligan et al., 2007) and the Cognitive Assessment Interview (Bilder et al., 2008) have been assessed for their psychometric properties as co-primary measures and found to be reliable and valid instruments. The everyday activities of these measures may not be representative of

everyday activities in other cultures around the world. Given the increase in multi-site international trials, evaluating cultural adaptation of the most commonly used U.S. measures is indicated.

The need for adaptation of instruments in cross-cultural studies is well documented (Van de Vijver and Hambleton, 1996; Nampijja et al., 2010; Jeanrie and Bertrand, 1999). In 1993 the International Test Commission (ITC) identified four domains in test translation; context, test development and adaptation, administration, and documentation/scoring (Van de Vijver and Hambleton, 1996). Jeanrie and Bertrand (1999) expanded the ITC guidelines suggesting there should be three distinct considerations in test adaptation: linguistic equivalence (i.e., comparable words, verb tenses and idioms), content equivalence (i.e., appropriateness of the behaviors and symbols that are referred to in the original test items) and conceptual equivalence (i.e. the meaning of the concept carried by the items, such as whether the concept exists in the target group and the specific meaning of each item).

As part of the MATRICS initiative, the Cross-Cultural Adaptability of Intermediate Measures Study (CIM) surveyed international clinical trial experts in schizophrenia. In their quantitative study, Velligan et al. (2010) provide an overall grade for cultural adaptability for the main tests by country, and also make a number of between group comparisons about the cultural adaptability of the subtests (e.g. India versus Mexico). In general, the

* Corresponding author. Tel.: +1 210 567 5424; fax: +1 210 567 3759.

E-mail address: gonzalezjm1@uthscsa.edu (J.M. Gonzalez).

measures were reported as not appropriate for cross-cultural use in their current form and the Cognitive Assessment Interview was rated as the most cross culturally appropriate (Velligan et al., 2010). India, China and Mexico presented the greatest challenges in adaptation, as compared to Spain, Germany, Russia and Argentina. However, the quantitative findings did not address how to proceed in terms of understanding the linguistic, content and conceptual equivalence issues in the original measure, nor whether the test should be adapted at a development, scoring or administration level. Qualitative feedback is an important second step in determining why an item may not perform well in another culture. This feedback can also explicate general guidelines about what aspects of functional outcome measures are particularly challenging to adapt cross-culturally. At times so many modifications would be required that while a test could be used within a given country with its own standardization and norms, the test would no longer be considered valid to make outcome comparisons across countries.

The current paper reports on the qualitative feedback provided by respondents in the CIM project. We provide an overall rating by subscale for the degree of modification recommended for adaptation, and provide detail on the specific conceptual adaptations recommended for functional outcome measures. Specific countries identified in the paper can utilize and incorporate the feedback in development for these specific measures; other countries can also use the qualitative information to guide any test adaptation or development for functional instruments in schizophrenia.

2. Methods

2.1. Design

As part of the CIM study, the Cross-Cultural Adaptation Rating Scale (C-CARS) was developed to survey clinical trials experts on issues of cultural adaptability of functional outcome measures used in schizophrenia research. The ITC guidelines and literature related to cultural adaptation were used to inform item development (e.g. Tanzer and Sim, 1999; Cook et al., 2005). Nine items were created to rate the extent to which each subscale of each intermediate measure of functional outcome was appropriate to the culture overall as well as by gender, socioeconomic status, ethnicity and area of residence (urban versus rural). Area of residence was a primary focus due to the greater number of rural citizens in the participating countries (see Table 1). Each item on the C-CARS is rated on a seven point scale with anchor points ranging from “least useful” to “most useful”. Higher scores indicate better cultural adaptability. If the rater scored four or below for any item, he or she was asked to provide a free-form explanation regarding the problems the item may have in cultural adaptation and suggestions to improve

Table 1

Population of urban and rural areas 2011, by countries participating in the CIM project. Source: <<http://esa.un.org/unpd/wup/CD-ROM/Urban-Rural-Population.htm>> (downloaded 6/29/2012); United Nations, Department of Economic and Social Affairs, Population Division (2012). World Urbanization Prospects: The 2011 Revision, CD-ROM Edition.

Country	Urban	Rural
Argentina	93%	7%
United States	82	18
Spain	77	23
Mexico	78	22
Germany	74	26
Russia	74	26
China	51	49
India	31	69

Urban population=De facto population living in areas classified as urban according to the criteria used by each area or country.

Rural population=De facto population living in areas classified as rural. Data refer to July 1, 2011.

item fit with the cultural context. Responses therefore referred back to a specific item within a subscale and ranged in length from a phrase to a full paragraph.

A qualitative summary of the free-form feedback is the subject of the current study. In summarizing the results, we were guided by the ITC guidelines for test development and adaptation and focused particularly on the content and conceptual equivalence of items and subscales. Study data were collected from May 2009 to January 2010.

2.2. Respondents

Pharmaceutical companies engaged in international research in schizophrenia identified participating sites. Two sites were selected for countries judged to be culturally more similar to the U.S. in Western/European cultural influences (i.e., Germany and Spain) and five sites were selected for countries considered less similar to the U.S. (but may also have European cultural influences), i.e., Argentina, China, India, Mexico, Russia. Thirty one sites were identified in eight countries as the U.S. also participated. Typically one principal investigator and one research assistant participated. Study participants were bilingual (English and local language). Fifty five participants responded (Argentina=eight, China=seven, Germany=four, India=eight, Mexico=eight, Russia=11, Spain=four, United States=five) out of 64 requested responders.

2.3. Instruments

The Cognitive Assessment Interview (CAI; Bilder et al., 2008) is a 10-item semi-structured interview developed from the Clinical Global Impression of Cognition in Schizophrenia Scale (Bilder et al., 2003) and the Schizophrenia Cognition Rating Scale (Keefe et al., 2006). The CAI can be administered with the patient alone or the patient and caregiver. The rater uses all available information to rate cognition on a series of Likert scales. Questions include: “Do you have difficulty keeping figures in mind while paying bills?” and “Do you have trouble coming up with alternatives when your plans are disturbed?”

The Independent Living Scales (ILS) (Loeb, 1996) is a performance-based test of instrumental activities of daily living and has five subscales: Memory/Orientation—e.g., person is asked to remember the name of a new doctor and the time of an appointment when asked later; Managing Money—e.g., person is asked to make out a check/money order to a utility company; Managing Home and Transportation—e.g., person is asked how they would go about getting repairs made to their home; Health and Safety—e.g., the person is asked what they would do if they cut their hand and it was bleeding profusely, and Social Adjustment—e.g., person is asked to name two reasons why it is important to have relationships.

The UCSD Performance Based Skills Assessment (UPSA; Patterson et al., 2001) assesses performance on functional tasks. Five skill areas are considered: General Comprehension—the person is asked to read a newspaper article about the opening of a water park, to remember information, and plan a trip there; Finance—the person is asked to pay a bill, make change, etc. Social/Communications—the person is asked to read a letter from their doctor about an appointment, and to remember the information when the letter is removed; Transportation—the person is asked to plan a bus route to specific destinations and to answer questions about it; Household Chores—the person is asked to write a shopping list for items needed to prepare a specific dish.

The Test of Adaptive Behavior in Schizophrenia (TABS; Velligan et al., 2007) assesses goal-directed adaptive behavior and comprises six subscales: Medication Management—the person is asked to fill a medication container based upon instructions on pill bottles; Empty Bathroom—the person is asked what would be needed to stock an empty bathroom to get ready every day; Shopping Skills—the person is asked how he or she would get to the store by using a map, to remember a grocery list, and to pay for items with a set amount of money; Clothes Closet—the person is asked to select appropriate clothing for specific activities; Work and Productivity—the person is asked to make packets of fliers and stack them for mailing, and Social Skills—basic skills such as voice volume and eye contact are rated during the assessment.

2.4. Data analysis

We used a modified grounded theory approach for the data analysis (Charmaz, 2006). First, detailed notes were taken separately by two researchers (M.R. and J.G.) while reading the qualitative data. The researchers then met to discuss their findings of the main themes in the data and overall impressions of test adaptability. The raters used the ITC model (Van de Vijver and Hambleton, 1996; Jeanrie and Bertrand, 1999) for adaptation in structuring the feedback, i.e. assessing whether the test items and subscales met criteria for content equivalence and conceptual equivalence. New constructs and themes were identified when the content could not be mapped onto existing equivalence domains. Coded segments were then organized into the preliminary categories. The researchers discussed the content at this stage until both were satisfied that data saturation and

reliability had been achieved. Initially coded segments were further refined into the categories presented in the results.

3. Results

There were 55 participants who provided feedback. The results are structured by the areas of adaptation assessed in the C-CARS: general cultural differences, socioeconomic and area of residence (urban versus rural) differences, and gender differences. Within those areas, specific domains of adaptation emerged and the content and concept equivalence for those domains are described. The extent of content in particular topics needing adaptation is a consequence of the emphasis placed by the study participants on those topics. For example, participants focused more on urban versus rural differences than gender, so there is more content reported on urban versus rural differences. The quotes are intended to provide an example of how respondents might view a measure as non-equivalent, rather than to make any generalizations about the country.

3.1. General cultural differences

General cultural differences refer to adaptations to the scale that would be pertinent for most persons with schizophrenia. Because of the content overlap of SES and area of residence with general cultural differences, the feedback is integrated in this section. Comments were organized into eight domains of adaptation.

Health insurance, health care and medication. Respondents noted adaptation difficulties with the concept of health insurance and health care. The ILS Health and Safety subscale asks questions about appropriate ways to contact a pharmacy for a refill of medication, the UPSA Communication Skills subscale asks test takers to answer questions regarding a letter from a medical doctor regarding an upcoming appointment, and the ILS Managing Money subscale asks questions about health insurance and paying medical bills. Some countries reported there was universal health care in their country while others reported that a substantial number of their citizens had no access to health insurance. For example, Argentina noted “health insurance is not used in Argentina” and India noted “very few people know about health insurance”. Or, “In Spain there are private medical insurance companies but since there is universal health care, their use is more limited than in the U.S.” This feedback raised questions regarding the content and concept appropriateness of items.

All countries reported medication management in their country differed from the TABS Self Medication task, which involves putting medications in a pill box and answering questions about the prescription bottle. In other countries, medications were most often dispensed in their original, non-personalized packs from the pharmaceutical company, rather than personalized bottles. Pharmacies did not refill medications, rather patients needed to return to the clinic or a hospital setting to obtain refills. For example, Germany noted: “Medications in Germany are supplied in branded blisters and are not dispensed individually-labeled by the pharmacy. There are no refills.” China noted that patients were not expected to manage their own medication nor did they usually know the name of the medications they were prescribed. This feedback indicated problems with test content as well as the concept of patient medication management. Respondents noted the adaptation concerns in this category would be particularly pertinent for persons of low SES and from rural areas.

Financial transactions. Money management and the type of government programs available varied widely. The UPSA Financial Skills and ILS Managing Money subscales involve tasks such as filling out a

money order or writing a check to a utility company. As expected, all countries noted content inequivalence for U.S. currency. The use of money orders or checkbooks was unusual and the cash system was primary. Argentina noted “Cultural difficulty because generally it is not the patient who takes care of these kinds of things.” Mexico noted conceptual difficulties with the ILS Managing Money subscale, stating “There is a lack of Social Security, the services or benefits that people can receive from this don't apply in Mexico... we don't have money orders”.

In all countries but Germany and Spain individuals of low SES and rural areas were identified as being unfamiliar with non-cash financial transactions. Respondents in Russia noted “credit cards and driver's licenses are not valid for lower SES” (they are used as props in the ILS Memory and Orientation subtest).

Utilities. In most countries except for Germany and to some extent Spain and Russia, the management of utilities differed, particularly in the UPSA Financial Skills subscale where test takers are to answer questions regarding paying a utility bill. Utilities were scarcer as were energy bills. In Mexico, respondents noted that payments are made in cash and utility bills were not generated. China noted, “People pay less attention to the details on the bill, or the real bills contain less information than asked in the test”. Respondents in Argentina noted a “large proportion don't have access to services provided by a corporation”. Overall, there was content and concept inequivalence noted with the potential adaptation of these subscales.

Emergency numbers and emergency situations. Raters reported that emergency situations and systems in their country differed substantially. In the ILS Health and Safety subscale, respondents are asked how to handle medical emergencies. India noted, “The choice of actions on help seeking are made by patients and close relatives collectively in the Indian socio-cultural milieu. Hence presuming standard line of actions and involvement of tertiary support may give a misleading picture.” In terms of emergency situations in the U.S., the rapidity of response to potential emergency situations was viewed as atypical in other cultures. For example, one ILS Health and Safety item asks how test taker should respond after experiencing a 10 lb weight loss in four weeks. In the U.S., the highest functioning response was to go to the doctor. However, respondents from other countries noted it would take more symptoms to evoke concern. Equivalence problems were mostly centered on content.

Home. The TABS Empty Bathroom subscale asks test takers to prepare an empty bathroom for use every day. Except for Germany and Spain, the configuration of a bathroom differed, such as the amount of items contained in a typical bathroom. Some countries noted that not all test takers would have bathrooms in their homes and particularly not the set-up of the subtest bathrooms. Respondents in China noted “They have no personal experience of what is needed for a bathroom and think of less items and categories.” There were content and conceptual difficulties with this subscale.

Social interactions. Fewer adaptation difficulties were noted for subtests involving social interactions. The ILS Social Adjustment Scale assesses mood and attitude toward social relations. India echoed the general feedback of this scale, noting “this scale is universal providing the deeper meaning comes through”. For UPSA Communication Skills and TABS Social Skills, clinical assessors rate the test takers' social skills, including eye contact, gestures and initiation of conversation. Russia noted that cultural subgroups might differ with respect to appropriate levels of eye contact. Other countries also reported that cultural groups can differ in social interaction norms. China noted “speech rate, volume, gesture, and initiative are a totally personalized style, it's hard to point out which is good or bad.” Adaptation challenges involved primarily content equivalence.

Food. Respondents noted wide cultural differences in common foods and in some cases how grocery stores were configured. In the UPSA Household Management subtest, participants are asked to make a shopping list for a recipe, based upon items currently in their home. Respondents in China noted “We never prepare the rice dessert like described, it should be modified to Chinese cooking”. The concept of dessert also was less familiar in China. In the TABS Grocery Store subtest, aisles are depicted on poster boards and subjects are asked the prices of items and to remember a short grocery list. India, China and Mexico suggested altering the grocery store depictions. Content differences were the primary concern on these subtests.

Transportation. The concept of daily transportation skills was acceptable in urban areas but modes of transportation differed and all respondents noted that subjects would be unfamiliar with U.S. transportation systems. In the U.S., taxis may be more often used due to inaccessible public transport and airplanes are also more common given the size of the U.S. territory and lack of accessible cross-country transportation. In the ILS Managing Home and Transportation and UPSA Transportation subtests, transportation includes buses, bus schedules, taxi cabs and the airport. In other countries, bus schedules were either not used or not as complex as the bus schedule example used in the UPSA Transportation subtest. Argentina noted “bus schedules are unusual, maps not usually available and not familiar to the culture” and China noted, “Some people, especially some with schizophrenia, never go to an airport”. Not surprisingly, in rural areas public transportation was less available and transportation was described as less complex of a task than the U.S.-based subtests.

Major SES and area of residence concerns. All countries except Germany and Russia mentioned concerns about the literacy requirement of subtests. In addition to simple literacy, respondents reported concerns about the complexity of the written information. For example the ILS Managing Home and Transportation asks subjects to read a map and use a telephone book to look up a number. Spain noted that persons of lower SES “may be near illiterate and never have used a map.” With regard to UPSA Transportation, Argentina noted it was “difficult to understand directions on maps, low literacy will increase difficulties”. For the CAI Verbal Learning and Memory, Argentina stated the subtest “may not work in individuals in rural areas because they usually don’t read or write”.

High SES concerns. Russian respondents noted adaptability concerns with a high SES group on the TABS Work Productivity subscale, noting “people can refuse to do simple work because of arrogance”. They also noted that on the TABS Social Skills a high SES group “may be more arrogant”.

In summary, there were substantial cultural differences indicating content inequivalence for all subtests on the intermediate scales. Even more challenging to the process of adaptation, respondents in multiple countries identified numerous subtests with conceptual inequivalence. There was significant overlap in rater’s opinions of how SES and area of residence influenced cultural adaptability. Both of these areas of adaptation had a major impact on the adaptability of the subscales, particularly in terms of familiarity with day to day activities. Literacy was also noted as limited in lower SES groups. With the exception of Argentina, all other countries have a higher rural population than the U.S. (23%–70% versus 18% of U.S.). These differences suggest that without substantial adaptation the subtests would not be appropriate measures of day to day functional activities.

3.2. Gender

There was less participant feedback regarding perceived gender differences contained in the measures. Respondents in all

countries except for Argentina reported that gender differences in daily functioning could impact subtest adaptability. In general, non-U.S. countries described more traditional gender roles with respect to social interactions, financial knowledge, cooking and grocery shopping. Regarding the UPSA Household Management subtest, all countries except for Spain noted that a subtest on cooking would not be acceptable for some men in their country. For example, China noted “some males do not cook in China”. Mexico noted “most men aren’t familiar with recipes or pantries”. For the TABS Grocery Store subtest, India and Russia mentioned concerns about being asked to simulate shopping, with India noting “males do not generally go grocery shopping”. In terms of assessing financial skills in the ILS Managing Money, India noted “most of the items are not appropriate for females”. Interestingly, Russia noted that men pay bills less often than women and so may be less familiar with bill paying. In Russia, men were reported to be less “sensitive” and “expressive” than females which might affect subtests such as the TABS Social Skills.

3.3. Adaptability of test administration

In general, there were few concerns in the domain of test administration and concerns were not consistent across any two countries. This suggests that generally the props, number of items, time to take tests and scaling were well received across countries.

Table 2 provides an overview of the results by subscale and provides a recommendation for the overall level of adaptation needed for each subscale. Within a country, the level of adaptation could differ.

4. Discussion

In this qualitative analysis of intermediate measures of functioning in schizophrenia all countries participating noted major adaptation challenges for the U.S. measures, both in content and concepts assessed. The CAI, an interview and non-performance based measure, was viewed as requiring the fewest adaptations. The results suggest substantial test adaptation should occur prior to implementing tests for cross cultural use.

One of the primary challenges in adapting tests is ensuring conceptual equivalence (Jeanrie and Bertrand, 1999). It appears specific daily functioning tasks vary greatly from country to country. For example, in the U.S. the consensus of mental health professionals is that functional individuals with serious mental illness are able to manage their medication, albeit with supports such as pill boxes, reminders, etc.—the TABS Self Medication subscale (Velligan et al., 2007). However, this concept was unfamiliar to other countries. To adapt this subscale, experts would need to consider how a person with schizophrenia is expected to manage his/her illness and what is considered the most adaptive level of illness management in their country. Or, in other countries the concept of self-management of illness may be flawed. Subtests or items could be removed or an equivalent concept could be inserted into the subtest/item. An equivalent concept for daily functioning could be assessing cooking skills, the ability to assist with the family business or maintaining a clean home.

Content equivalence involves culturally appropriate use of behaviors and symbols, such as rupees in India rather than dollars or discussing attending a soccer match rather than a water park for Argentina. An example of content and conceptual inequivalence was for the UPSA Household Management subscale. Respondents from India and China noted that to make a rice dessert based on a recipe would be conceptually inequivalent

Table 2
Revisions suggestions for MATRICS co-primary measure subtests, averaging across sites.

Test	Subscale	Main conceptual challenges	Degree of modification
CAI	Working memory	Different financial and payment systems	Low
CAI	Attention/Vigilance	Lack of familiarity with maps; rural and low SES less familiar with examples	Low
CAI	Verbal learning and memory	Literacy	Low
CAI	Reasoning and problem solving	A subset of rural individuals may have difficulties with examples	Low
CAI	Speed of processing	Illiterate individuals cannot “read instructions”	Low
CAI	Social cognition	Task may be too complicated, possible gender differences	Low
ILS	Health and safety	Different utility and health service systems; perceived severity of health symptoms	Medium
ILS	Managing home and transportation	Different utility, health service and transportation systems, more challenges with low SES individuals	High
ILS	Managing money	Different financial and payment systems, different social service systems	High
ILS	Memory and orientation	Challenges in content and concept with rural and lower SES individuals	Medium
ILS	Social adjustment	Level of social activity may be lower	Low
TABS	Self medication	Different methods of medication dissemination, packaging and medication management, subtest too complex, literacy	High
TABS	Work and productivity	May pose challenges for illiterate individuals, low SES, males, persons with schizophrenia may not be expected to carry out office tasks	Medium
TABS	Empty bathroom	Bathroom set-up differs, too many items in bathroom	High
TABS	Grocery store	Lack of familiarity with maps, males unfamiliar with tasks, grocery shopping tasks may differ	Medium
TABS	Clothing	Limited access to a wide variety of clothing; appropriate dress differs considerably across cultures and groups within the culture	Medium
TABS	Social skills	Consider differences by geographic location, gender and ethnicity	Low
UPSA	Communication skills	Complexity of written information, different access to health service systems, some differences with social service systems	High
UPSA	Comprehension and planning	Typical leisure activities and how they are structured differ	Medium
UPSA	Financial skills	Lack of familiarity with bill paying due to lack of responsibility for bill paying or being from rural or low SES group	High
UPSA	Household management	Gender differences in cooking skills	Medium
UPSA	Transportation	Too complex, different transportation systems	High

The degree of modification was based on rater agreement of the overall number of comments, number of items in the subscale recommended for conceptual modification and the quality of the feedback.

SES=socioeconomic status.

because the level of education required to read and to follow recipe instructions was too high for their population. Further, the type of dessert was not familiar to some non-U.S. countries and would require using a dessert typical to their country (content equivalence).

The CAI interview and social skills subscales rely more on rater judgment. For example, ratings of good eye contact in India can be rated using culturally consistent norms, where the anchors for what involves good eye contact can rather easily be adjusted. However, there is a caveat with the use of the CAI as a co-primary measure of outcome in schizophrenia. Recent findings (Green et al., 2011) showed that the CAI scores did not correlate well with other measures of functional outcome, or to cognitive measures that predict functional outcome in schizophrenia.

The respondents also noted strong concerns about items working well across various cultural subgroups. Other than Argentina, other countries participating often had much higher numbers of rural citizens than the U.S. In those countries, two test versions may be indicated if day to day functioning is quite dissimilar. In addition, respondents described more traditional and differentiated gender roles in non-U.S. countries which impacted how daily functioning tasks were operationalized. It is important to note that differences between the U.S. and other countries do not indicate that the U.S. system is considered superior or the gold standard.

Socioeconomic level and cultural traditions often overlap and we particularly found this to be the case in the expert feedback from this study. Cultural traditions are certainly impacted by socioeconomic status and in what type of social and work

environment people live. In test development, it is important that each subgroup is well-represented and quantitative analyses should occur in conjunction with qualitative analyses to ensure impressions of experts are consistent with actual group differences and not stereotypes. Even with this, it may be unclear whether a particular group difference is driven by socioeconomic challenges or cultural traditions, or if those can be or should be teased apart.

Tanzer and Sim (1999) describe four levels of modifications in test adaptations, distinguished by the complexity and number of changes needed for a culturally competent adaptation: *application*, *translation*, *adaptation* and *assembly*. *Application* would be making no changes to the instrument, *Translation* would require a simple literal translation, and *Adaptation* is when content and expressions from the U.S. must be replaced with expressions from the target cultures that have the same meaning. *Assembly* modifications are where enough modifications have to be made to the instrument that, practically speaking, a new test is created. For example, tests with incomplete overlap of the definition and/or structure of the construct across populations or behaviors or symptoms associated with the construct (Cheung et al., 1996; Yang and Bond, 1990; Zhang and Bond, 2011). Based on our findings from the current study as well as results of the quantitative study (Velligan et al., 2010) the ILS, TABS and UPSA would need Assembly-level adaptation for use in other cultures. Another approach would be to adapt only those subtests with the fewest adaptation challenges. Our research group is currently conducting a psychometric study of a combination of UPSA and TABS subscales that were rated as most adaptable across countries.

Cross cultural test adaptation should follow the well-established ITC guidelines described elsewhere in detail (Van de Vijver and Hambleton, 1996; Hambleton, 2001). However, we briefly review relevant recommendations of test development and adaptation experts.

Jeanrie and Bertrand (1999) focused on certain aspects of the ITC Guidelines for test adaptation, specifically emphasizing the use of translators and judges fluent in both languages, with the content of the scale and with the item writing process. They also support the use of systematic judgmental scales when conducting translations, where raters rate for content, conceptual and linguistic equivalence of a new item. Sireci et al. (2006) point out the importance of using expanded, comprehensive translation processes in a series of steps, such as (a) initial translation, (b) review and discussion among the translators and key stakeholders, (c) back translation by a set of independent translators, (d) review and discussion of original and back translations, (e) revision of original translations (if necessary), and (f) final approval. Analyzing data from a translated instrument, the authors conclude that quantitative assessments of an adapted measure are as necessary as qualitative assessment—for example differential item functioning analyses. Tanzer and Sim (1999) note for cross cultural test adaptations, additional information included in test manuals should include (1) the identification of strengths and weakness of the original measures, the resulting changes and their justifications; (2) evidence for construct, content, and ecological equivalence; (3) the evidence for the validity of comparisons across populations; and (4) the specific qualifications needed for using the test in a multicultural/multilingual context. A more novel approach to test development and adaptation could involve use of the internet for culturally relevant information such as what are considered common foods in varying cultures or how emergency systems may work in various countries. In a globalized society, the internet and its enhanced use has the potential to provide a wealth of information for U.S. and non-U.S. test developers in assessment of culturally appropriate content for assessing functional outcome.

Of course, an alternate option for schizophrenia researchers is the development of a new scale as modifying an existing scale too much may render the adapted scale invalid and no longer measuring functional outcome. However, given the four scales we studied are commonly used and well-accepted in schizophrenia research due to their reliability and validity, we recommend researchers consider adaptation, or close mirroring of the principles (e.g., purpose of subscales, test format, etc.) in each of the measures.

Limitations. This is a qualitative study and results are not representative of all viewpoints within the countries studied. We did not conduct ethnographic interviews with the participants in this study, which may have yielded different feedback than asking respondents to write answers. The data were provided by experienced clinical trial researchers; however, the scales were not administered to clinical trial participants. It is important to note that although we report better adaptability for countries that are similar to the U.S. (i.e. Germany and Spain), there were also fewer raters in these countries and thus less qualitative feedback. While these four instruments are now considered the most reliable and valid instruments in the U.S. for assessing functional outcome in schizophrenia, there may be cultural limitations of these instruments in the U.S., which we do not address in this study. For example, the U.S. “white” population consists of a range of cultural backgrounds (i.e., patient of Swedish descent, patient of Greek descent, Mexican immigrant) who may differ in some of the cultural differences noted in other countries (i.e. eye contact, gender roles, SES). Another important topic not addressed in this study is cross cultural differences in patient recruitment and

retention and motivation to participate in clinical schizophrenia research.

These results provide valuable information for testing experts in structuring the adaptation process. For example, test adapters for functional outcome measures now know there is rarely content equivalence across cultures, and concepts may also differ, such as what is considered to be functional use of transportation, the health care system or acceptable daily hygiene. Test adapters can improve their knowledge of the rural makeup of the country, typical gender roles, and differences in social, financial and transportation systems.

Funding

This work was supported by an extension to contract HHSN 278 2004 41003C from the National Institute of Mental Health (Stephen Marder, Principal Investigator).

Acknowledgments

The authors wish to thank other members of the CIM committee: Keith Neuchterlein, Nina Schooler, Judith Jaeger, Nancy Peters, Raimund Buller, Stephen Marder and Sanjay Dube. We also thank members of the MATRICS-CT (Co-primary and Translation) Scientific Board and the Cross-cultural Subcommittee. We also wish to thank the investigators and research assistants across all sites for their time and effort on this project.

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

- Bilder, R., Ventura, J., Cienfuegos, A., 2003. Clinical Global Impression of Cognition in Schizophrenia, Version 3.2: Interviewer's Manual: Definitions and Guidelines. University of California, Los Angeles.
- Bilder, R., Ventura, J., Reise, S., Keefe, R., 2008. Cognitive Assessment Interview (CAI). Interviewer's Manual: Definition and Rating Guidelines. Neuropsychiatric Institute, UCLA, California.
- Buchanan, R., Davis, M., Godd, D., Green, M.F., Keefe, R.S., Leon, A.C., Nuechterlein, K.H., Laughren, T., Levin, R., Stover, E., Fenton, W., Marder, S.R., 2005. A summary of the FDA-NIMH-MATRICES workshop on clinical trial design for neurocognitive drugs for schizophrenia. *Schizophrenia Bulletin* 31, 5–19.
- Charmaz, K., 2006. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Sage, London.
- Cheung, F.M., Leung, K., Fan, R.M., Song, W.Z., Zhang, J.X., Chang, J.P., 1996. Development of the Chinese personality assessment inventory. *Journal of Cross Cultural Psychology* 27, 181–199.
- Cook, L., Schmitt-Cascalliar, A.P., Brown, C., 2005. Adapting achievement and aptitude tests: a review of methodological issues. In: *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates, London, pp. 171–192.
- Gold, J.M., Harvey, P.D., 1993. Cognitive deficits in schizophrenia. *Psychiatric Clinics of North America* 16, 295–312.
- Green, M.F., 1996. What are the functional consequences of neurocognitive deficits in schizophrenia. *American Journal of Psychiatry* 153, 321–330.
- Green, M.F., Nuechterlein, K.H., Kern, R.S., Baade, L.E., Fenton, W.S., Gold, J.M., Keefe, R.S., Mesholam-Gately, R., Seidman, L.J., Stover, E., Marder, S.R., 2008. Functional coprimary measures for clinical trials in schizophrenia: results from the MATRICS psychometrics and standardization study. *American Journal of Psychiatry* 165, 221–228.
- Green, M., Schooler, N., Kern, R.S., Frese, F.J., Granberry, W., Harvey, P.D., Karson, C.N., Peters, N., Stewart, M., Seidman, L.J., Sonnenberg, J., Stone, W.S., Walling, D., Stover, E., Marder, S.R., 2011. Evaluation of functionally meaningful measures for clinical trials of cognition enhancement in schizophrenia. *American Journal of Psychiatry* 168, 400–407.
- Hambleton, R.K., 2001. The next generation of the ITC Test Translation and Adaptation Guidelines. *European Journal of Psychological Assessment* 17 (3), 164–172.
- Jeanrie, C., Bertrand, R., 1999. Translating tests with the International Test Commission's Guidelines: keeping validity in mind. *European Journal of Psychological Assessment* 15, 277–283.
- Keefe, R.S.E., Poe, M., Walker, T.M., Kang, J.W., Harvey, P.D., 2006. The schizophrenia cognition rating scale: an interview-based assessment and its

- relationship to cognition, real-world functioning and functional capacity. *American Journal of Psychiatry* 163, 426–432.
- Loeb, P.A., 1996. *Independent Living Scales Manual*. Psychological Corporation, San Antonio.
- Nampijja, M., Apule, B., Lule, S., Akurut, H., Muhangi, L., Elliot, A.M., Alcock, K.J., 2010. Adaptation of Western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *British Journal of Educational Psychology* 80, 15–30.
- Nuechterlein, K.H., Robbins, T.W., Einat, H., 2005. Distinguishing separable domains of cognition in human and animal studies: what separations are optimal for targeting interventions? A summary of recommendations from breakout group 2 at the measurement and treatment research to improve cognition in schizophrenia new approaches conference. *Schizophrenia Bulletin* 31, 870–874.
- Patterson, T.L., Goldman, S., McKibbin, C.L., Hughs, R., Jeste, D.V., 2001. UCSSD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophrenia Bulletin* 27, 235–245.
- Sireci, S.G., Yang, Y., Harter, J., Ehrlich, E.J., 2006. Evaluating guidelines for test adaptations: a methodological analysis of translation quality. *Journal of Cross-Cultural Psychology* 37 (5), 557–567.
- Tanzer, N.K., Sim, C.Q.E., 1999. Adapting instruments for use in multiple languages and cultures: a review of the ITC Guidelines for Test Adaptations. *European Journal of Psychological Assessment* 15, 258–269.
- Van de Vijver, F.J.R., Hambleton, R.K., 1996. Translating tests: some practical guidelines. *European Psychologist* 1, 89–99.
- Velligan, D.I., Mahurin, R.K., Diamond, P., Hazleton, B.C., Eckert, S.L., Miller, A.L., 1997. The functional significance of symptomatology and cognitive function in schizophrenia. *Schizophrenia Research* 25, 21–31.
- Velligan, D.I., Diamond, P., Glahn, D.C., Ritch, J., Maples, N., Castillo, D., Miller, A.L., 2007. The reliability and validity of the test of adaptive behavior in schizophrenia (TABS). *Psychiatry Research* 151, 55–66.
- Velligan, D.I., Rubin, M., Fredrick, M.M., Mintz, J., Nuechterlein, K.H., Schooler, N.R., Jaeger, J., Peters, N.M., Buller, R., Marder, S.R., Dube, S., 2010. The cultural adaptability of intermediate measures of functional outcome in schizophrenia. *Schizophrenia Bulletin*. (epub ahead of print).
- Yang, K.S., Bond, M.H., 1990. Exploring implicit personality theories with indigenous or imported constructs: the Chinese case. *Journal of Personality and Social Psychology* 58, 1087–1095.
- Zhang, J.X., Bond, M.H., 2011. Personality and filial piety among college students in two Chinese societies. The added value of indigenous constructs. *Journal of Cross Cultural Psychology* 29, 402–417.