The New School Psychology Bulletin
2007, Vol. 5, No. 2

Print ISSN: 1931-793X; Online ISSN: 1931-7948

Review Report

Cross-Cultural Awareness in Quality of Life Assessment

Scott M. Debb
Evanston Northwestern Healthcare

Since many tools for clinical assessment are available to practitioners of all medical and psychiatric fields, when a healthcare professional assesses the mental health of an individual, the use of culturally appropriate assessment measures used becomes increasingly important. Since many quality of life (QOL) assessment measures utilize Likert scaled questions, awareness and understanding of potential bias can greatly aid in clinical interpretation of scores, case conceptualization and treatment planning. Mental health professionals need to be cognizant of cultural issues and sensitive to the idea that the categorization of symptoms that our culture has clustered into disorders is dependent on specific cultural values (Draguns, 1987). Ultimately, cultural bias can greatly affect the validity and interpretability of assessment data. Modifications to assessment measures must also be considered in light of cross-cultural concerns, primarily as related to content validity.

Keywords: quality of life (QOL), cross-cultural assessment, item response theory (IRT), differential item functioning (DIF), construct validity

While the exact definition of what comprises culture is debatable (Carter & Forsyth, 2007), culture is generally regarded as the intangible systems of learned and shared beliefs common to a group or groups (Whaley & Davis, 2007; Sattler, 2001; Carter & Qureshi, 1995). In our culture, cultural bias tends to be a fairly well-ingrained phenomenon and bridging the gap between what is observed and its clinical relevance is often assumed to be achieved through the skill of trained practitioners. The concept of bias deals with interpreting and judging the content of what is particular to a culture other than our own based on the norms, assumptions and conventions of our own culture.

This includes beliefs about dress, hygiene, family structure, rules of language, logic, laws or justice. To this end, what we consider to be normal beliefs are governed by those that are normative in our culture (Dana, 1993). In many instances, these beliefs and conventions are useful in the sense that they form the basis for how a culture can generalize the more common elements present in society, such as rules of criminal conduct or proper manners. The major problems arise, however, when people from different cultures are grouped together, such as in the United States, where the multitude of different cultures makes cross-cultural awareness across every subculture nearly impossible.

Healthcare professionals have a number of clinical tools available to assess their patients. This could include a blood test to determine the presence of cancer in an oncology patient, an X-Ray to determine if a bone is broken in an orthopedic patient, a urine test to check the sobriety of a substance abuser, an intelligence test to determine the functional ability of a student needing an individual education plan, or an interview or quesitonnaire to gauge a psychiatric patient's mental health. All methods are diagnostically relevant within their respective fields; however, some tools are more frequently used over others due to their perceived, and oftentimes researched, clincial efficacy.

When specifically using an assessment instrument to clinically assess a patient's mental health, the issue of cultural bias becomes paramount: Interpretation of an assessment measure's results are based on assumptions that do not necessarily take cross-cultural issues into account (Kwate, 2001; Valencia & Suzuki, 2001; Finkelstein, Pickert, Mahoney & Barry, 1998; Suzuki & Valencia 1997; Malgady, 1996). While clinical measures are often used to gauge patient progress through treatment by establishing a baseline at the start of treatment and then administering the same measure multiple times through the course of therapy, the results of the assessments are key components used to determine what services the individual needs and what services they will actually receive. As such, healthcare professionals are responsible for being aware of cultural issues and individual differences between patients.

Cultural Factors In Assessment Measures

Researchers have concluded that assessment measures are ethnocentrically created and that many tests tend to discriminate towards individuals from the culture from which they were developed (Anastasi & Urbina, 1997). According to Anastasi and Urbina (1997), "No single test can be universally applicable or equally fair to all cultures" (p. 342). To address this point, Likert type assessment measures (employing a multiple choice response scale such as

Correspondence concerning this article should be addressed to Scott M. Debb, Evanston Northwestern Healthcare, Center on Outcomes Research and Education, Evanston, IL. Email: sdebb@enh.org

using numbers 1-5 to represent severity of symptoms) try to utilize large samples to be representative of the target population being studied to accurately gauge the test's validity. Given this, assessment measures are still bound by the unique characteristics of the culture from which they were developed (Reynolds, 1982). This is known as cultural loading, and is different from cultural bias in the sense that cultural bias in assessment refers to a substantial difference between predicted and actual test results for specific groups of individuals (Suzuki & Valencia, 1997; Reynolds, 1982). Whether or not a measure is culturally loaded or biased, the end result is similar in that groups of people can potentially be misrepresented by the data.

The presence of some cultural bias in clinically validated instruments has extreme implications for all types of minorities. There has been much research dedicated to increasing awareness of the dangers of misusing common testing tools in cross-cultural contexts (Urani, 2005; Lievens, Harris, Van Keer & Bisqueret, 2003; Malgady, 1996; Miller-Jones, 1989). According to Hinkle (1994), the major difficulties with establishing precedence within cross-cultural assessment relate to: the determination of what constitutes culturally appropriate equivalence criterion from the norm culture to the subgroup populations, the lack of specific subcultural norms for nationally standardized tests, the acknowledgement and understanding of differences in culture-specific response styles, and the realization of the differing attitudes various cultures have toward the act of being tested.

What the majority defines as normal can be defined as behavior that does not violate cultural norms as it pertains to what is considered appropriate and acceptable, which does not deviate significantly from what is assumed to be appropriate for an individual's age group (Kaufman, 1989). This is to say that the normal behaviors observed in one subculture may be considered abnormal in comparison to the mainstream culture or to other subcultures (Frazier & DeBlassie, 1984). These differences affect assessment in the sense that in a society such as the United States where the proportions of minority subcultures are rapidly increasing, the normative data for minorities becomes skewed and has the potential to under- or misrepresent minority groups in many ways.

Quality of Life Assessment

A specific definition of what constitutes a person's quality of life (QOL) typically depends on the clinical context of the healthcare setting (Guyatt, Feeny, & Patrick, 1993) QOL can be defined more generically as an all-encompassing concept that summarizes the "broad-based assessment of the combined impact of disease and treatment and the trade-off between the two" (Cella, Tulsky, Gray, Sarafin, Linn, Bonomi, et al., 1993, p. 570). Cella (1994) noted that understanding QOL was becoming increasingly important because of its centrality to the area of supportive care. That trend has continued today, and QOL has become a primary measurement index in medical settings and secondarily used by mental health professionals.

QOL is a subjective concept that is measured from the patient's perspective (Cella, Lloyd, & Wright, 1996). To accurately assess a patient's QOL from a medical or psychological perspective, many different areas of his or her life must be examined, including areas of functional, physical, social, and emotional well-being. As such, it makes interpretation of assessment measures and comprehensive treatment planning more difficult because of the multi-dimensional nature of QOL and the variability in assessing the underlying constructs of well-being.

What is normal?

It has been argued that normatively referenced test and assessment results typically assume generalizability between different subgroups, stemming from cultural concerns, it has also been argued that the methods and assumptions behind the development of these norms be replaced with systems that allow for more cultural freedom (Reilly, 1991; Galagan, 1985; Messick, 1984; Ysseldyke & Thurlow, 1984). This implies that tests of generalized functioning can conceivably provide a less than accurate description of a person's actual functioning. In terms of clinical conceptualization of a mental health patient, the range of differences related to demographics in the United States (such as socioeconomic status, religion, ethnicity and primary language), can impede a clinician's ability to accurately understand a patient from within the boundaries of their cultural framework.

To this end, Dana (1993) points out that test misinterpretation is also an important factor to consider, emphasizing assessor bias related to reliability, in addition to any pitfalls in validity embedded in the actual assessment measure. For example, Guillemin, Bombadier, and Beaton (1993) note that while many questionnaires aimed at measuring QOL are multicultural in scope, the majority of the clinical measures used have been developed in English-speaking countries. By utilizing an arbitrary benchmark of normality based on a subjective standard, the possibility of inaccurately assessing or conceptualizing, and subsequent incorrect or inadequate treatment planning for a patient increases.

There are many forms of assessment that measure QOL for different populations and subpopulations, and it can be argued that each contains a certain amount of bias on some level. This presence has the potential to misrepresent any given patient, especially if they are from a minority group,

whether the minority is an ethnic group such as African Americans, or a subpopulation of African American mental health patients (Rao, Debb, Blitz, Choi, & Cella, in press; Debb, Blitz, & Choi, 2007; Kim, Pilkonis, Frank, These, & Reynolds, 2002; Bedi, Maraun, Chrisjohn, 2001). This can be due to bias being indirectly designed into the measure itself (such as by creating test questions that are culturally specific to the majority but are culturally inappropriate and not generalizable to other subgroups) or because of varying styles of data interpretation by the clinician. Even with tests being carefully designed, there can never be one measure capable of withstanding cross-cultural scrutiny across every subculture.

Item Response Theory

To minimize the effect of any bias that might be present in an assessment measure, one technique is to not utilize the entire test or subscale in its entirety, in terms of an overall score comparative to other test takers, but rather to look at each item in the measure individually. In a manner of speaking, this method serves as a control for individual differences that may not show up via a combined score. Effectively, each question becomes its own test.

Item Response Theory (IRT) is one of the common methods utilized in order to study item-level differences in how respondents answer QOL Likert scaled questionnaire data. IRT is based on the assumption that the performance of an individual on a test, directly pertaining to a specific item, can be predicted by the person's traits or abilities and that the relationship between the individual's performance on the item and their ability is monotonically consistent (Hambleton, Swaminathan, & Rogers, 1991). This is to say that the probability of endorsing an item increases as their underlying ability increases. This method of analyzing item and overall test appropriateness is in contrast to classical measurement systems that are generally limited by their inability to differentiate tester from testee characteristics. In addition, classical theories generally pertain only to the entire test measure, rather than focusing on individual items.

One method of utilizing IRT is to assess differential item functioning (DIF). According to Holland and Wainer (1993), DIF refers to the observation that an item exhibits different statistical properties for different groups of people, assuming that their inherent ability level is the same or in some other way controlled for. DIF analysis tries to determine if test questions are fair and appropriate for assessing the knowledge of various groups of people and if an item in a test measure is performing differently for one group of examinees relative to another group. The analysis is based on the assumption that test-takers who have similar knowledge (for example, based on an overall total test scale or subscale score) should perform comparably on each test item regardless of what their specific group's characteristics are. For example, Rao, Debb, Blitz, Choi & Cella (in press) implemented a one-parameter logistic Rasch model to detect item level DIF in a sample of African American and Caucasian chronic illness patients. The researchers used a QOL questionnaire with four subscales and ran separate multivariable linear regression models with each subscale as the dependent variable and race as the primary independent variable. Their results showed that 6 of the 26 test items exhibited DIF. This indicates that although African Americans and Caucasians experienced a similar QOL status, what they identified as being better or worse in their lives differed in some way for six items.

Utilizing the Data

Clay (2006) states that all assessments have their use, whether this is accomplished by interpretation of scores from an entire measure or utilization of the smaller pieces of the measure. Examining the available data on this level allows for a microscopic and more personal view of available information, rather than looking at assessment data from a more global, comparatively based perspective. This ultimately allows the clinician to develop a psychological conceptualization about their patient based on smaller pieces of a puzzle rather than the larger, more generic clusters.

According to Slife, Williams, and Barlow (2001), "It is essential to integrate traditional theories with increased cultural awareness [because] most theories are rooted in European-American middle-class values with a style and technique not appropriate to different culture groups" (p.282). Dana (1993) also points out that assessment measures that have been developed in the United States have been successfully adapted and employed in other countries; however these same tests have not been adapted for use with multicultural populations within the United States (p.92). For example, the FACT-G (Cella et al., 1993), is a questionnaire that has been successfully adapted and validated for a number of languages in a number of countries, such as Korean in Korea (Yoo, Suh, Kim, Eremenco, Kim, & Kim, 2006); Malayalam in India (Pandey, Thomas, Ramdas, Eremenco, & Nair, 2004), Pedi, Tswani, and Zulu in South Africa (Mullin, Cella, Change Eremenco, Mertz, et al., 2000), etc.; however, within the United States' borders, only English (Cella, 1993) and Spanish-speaking patients (Cella, Hernandez, Bonomi, Corona, Vaquero, et al., 1998) have been used to validate conceptually equivalent translated measures.

That is, while the need to adapt a diagnostic QOL questionnaire for use with multiple languages is important, specific cultural cautions also need to be addressed when

the measure is being used with individuals that speak the same language but are from different geographic locations (e.g., a Dutch adaptation for the Netherlands or Belgium; an English adaptation for England or South Africa; a French adaptation for France, Switzerland or Tunisia; Italian for Italy or Switzerland; Portuguese for Brazil or Portugal; Russian for Latvia, Russia or Ukraine; Spanish for Chile, Mexico, Spain or the United States, etc.). Geisinger (1994) notes that areas of caution in test content include the style of language, the level of sophistication, vocabulary, familiarity with cultural associations to societal norms (e.g., occupations, social relationships, familial relationships, etc.), and subsequent to the actual test administration, the interpretability of what has been measured. In other words, the construct validity of the adapted or modified measure may not be equivalent to the source it is based from (Hambleton & Patsula, 1998).

Building on this, Lezak, Howieson, and Loring (2004) state that when assessing someone, the examination should be catered to the patient's needs, abilities and limitations. Especially in mental health settings, this minimizes obstacles imposed by the testing situation, which could include cultural context, language, patient's attitude or cultural orientation toward psychological testing, or ethnic background of the clinician. Lezak also pointed out that by adapting the assessment to the particular individual, the clinician is able to better serve their patients because they receive the richest data from the assessment measure and process, allowing more specific information to be inferred from the results. This is especially true when assessing specific subpopulations, where the instruments employed may be bound by cultural or physical disabilities. This is important in QOL assessment because communication difficulties (verbal or non-verbal), can lead to a false conclusion regarding the severity of a patient's symptoms (Tchen et al., 2003).

A clinician's awareness of cross-cultural issues in psychological assessment is an important factor to consider regarding obtaining accurate results from the data. Clinicians using assessment data must pay attention not only to the test's developmental context, but also to their own cultural developmental context. This refers directly to the examiner's background and how it might differ from that of their patients. Dana (1993) notes that test score interpretation can be significantly influenced by the clinicians' culture, including their home country, current place of residence, race, religion, or gender. To counteract these potential biases, one possible remedy is to increase or improve the fairness of test development standards, which aim to minimize test bias that is the result of interpretability flaws (Anastasi, 1988). This also relates to increasing awareness regarding the implications of misinterpretation, adequately training clinicians to be sensitive and open about culture specific issues, as well as promoting cross-cultural awareness in the areas of test research and development (Hinkle, 1994). This is increasingly important when the examiner and examinee are from different cultures, or if the examinee's culture conflicts on some level with the culture for which the assessment measure was intended to assess, even if it has been adapted for a specific context.

Conclusion

Although QOL assessment conducted via self-report questionnaire can be a rich source of data regarding a patient's ability to function on a daily basis, cross-cultural issues pertaining to the interpretability of the assessment results can cause concern regarding the validity of a clinician's conceptualization. Awareness of these cultural issues must be maintained on a number of levels, beginning with the research phase of test development and extending through the time of its clinical application. Successful validations of cultural and illness-specific adaptations, modifications of measures, and creative interpretation techniques can help control some interpretability concerns; however, bias can never be fully eliminated from any given measure for every situation.

References

Anastasi, A. (1988). *Psychological testing* (4th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Bedi, R. P., Maraun, M. D., & Chrisjohn, R. D. (2001). A multisample item response theory analysis of the Beck Depression Inventory-1A. *Canadian Journal of Behavioural Science, 33*(3), 176-185.

Carter, R. T., & Forsyth, J. M. (2007). Examining race and culture in psychology journals: The case of forensic psychology. Professional Psychology: *Research and Practice, 3*(2), 133-142.

Carter, R. T., & Qureshi, A. (1995). A typology of philosophical assumptions in multicultural counseling and training. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 239–262). Thousand Oaks, CA: Sage.

Cella, D. F. (1994). Quality of life: Concepts and definitions. *Journal of Pain Symptom Management, 9*(3), 186-192.

Cella, D., Hernandez, L., Bonomi, A. E., Corona, M., Va quero, M., Shiomoto, G., & Baez, L. (1998). Spanish language translation and initial validation of the Functional Assessment of Cancer Therapy quality of

life instrument. *Medical Care, 36(*9), 1407-1418.

Cella, D. F., Lloyd, S. R., & Wright, B. (1996). Cross-cultural instrument equating: Current research and future directions. Quality of Life and *Pharmacoeconomics in Clinical Trials, 73*, 707-715.

Cella, D. F., Tulsky, D. S., Gray, G., Sarafin, B., Linn, E., Bonomi, A., et al. (1993). The Functional Assessment of Cancer Therapy Scale: Development and validation of the general measure. *Journal of Clinical Oncology, 11*(3), 570-579.

Clay, R. A. (2006). Assessing assessment: Psychological assessment enjoys new respect, applications and approaches. *Monitor On Psychology, 37*(1), 44-46.

Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology.* Needham Heights, MA: Simon & Schuster.

Debb, S. M., Blitz, D. L., & Choi, S. W. (2007, August). *Ethnicity and differential item functioning in quality of life assessment.* Poster session presented at the 115th Annual Convention of the American Psychological Association, San Francisco, CA.

Draguns, J. G. (1987). Psychological disorders across cultures. In P. Pedersen (Ed.), *Handbook of cross-cultural counseling and therapy* (pp.55-62). New York: Praeger.

Finkelstein, B., Pickert, S., Mahoney, T., & Barry, D. (1998). *Discovering culture in education: An approach to cultural education program evaluation.* Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED436538).

Frazier, D., & DeBlassie, R. (1984). Diagnosing behavior disordered early adolescents as a function of cultural differences. *Adolescence, 19*, 387-390.

Galagan, J. (1985). Psychoeducational testing: Turn out the lights, the party's over. *Exceptional Children, 52*, 288-289.

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

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

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine, 118*(8), 672-679.

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

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research, 45*, 153-172.

Hinkle, J. S. (1994). Practitioners and cross-cultural assessment: A practical guide to information and training. *Measurement & Evaluation in Counseling and Development, 27*(2), 103-116.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kaufman, J. (1989). *Characteristics of children's behavior disorders* (4th ed.). Columbus, OH: Merrill.

Kim, Y., Pilkonis, P. A., Frank, E., These, M. E., & Reynolds, C. F. (2002). Differential item functioning of the Beck Depression Inventory in late life patients: Use of Item Response Theory. *Psychology and Aging, 17*(3), 379-391.

Kwate, N. A. (2001). Intelligence or misorientation? Eurocentrism in the WISC-III. *Journal of Black Psychology, 27*(2), 221-238.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*(3), 476-489.

Malgady, R. G. (1996). The question of cultural bias in assessment and diagnosis of ethnic minority clients: Let's reject the null hypothesis. *Professional Psychology: Research and Practice, 27*(1), 73-77.

Messick, S. (1984). Assessment in context: Appraising students' performance in relation to instructional quality. *Educational Research, 13*, 3-8.

Miller-Jones, D. (1989). Culture and testing. *American Psychologist, 44*(2), 360-366.

Mullin, V., Cella, D., Chang, C.-H., Eremenco, S., Mertz, M., Lent, L., et al. (2000). Development of three African language translations of the FACT-G. *Quality of Life Research, 9*, 139-149.

Pandey, M., Thomas, B. C., Ramdas, K., Eremenco, S., & Nair, M. K. (2004). Reliability & validity of the Malayalam Functional Assessment of Cancer Therapy for head & neck cancer. *Indian Journal of Medical Research, 120*, 51-55.

Rao, D., Debb, S., Blitz, D., Choi, S., Cella, D. (in press). Ethnic/racial differences in the health-related quality of life of cancer patients. *Journal of Pain and Symptom Management.*

Reilly, T. F. (1991). Cultural bias: The albatross of

assessing behavior-disordered children and youth. *Preventing School Failure, 36*(1), 50-53.

Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-209). New York: Wiley.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Jerome M. Sattler Publisher.

Slife, B. D., Williams, R. N., & Barlow, S. H. (2001). *Critical issues in psychotherapy.* Thousand Oaks, CA: Sage.

Suzuki, L. A., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence: *Educational implications. American Psychologist, 52*(10), 1103-1114.

Tchen, N., Bedard, P., Yi, Q.-L., Klein, M., Cella, D., Eremenco, S., & Tannock, I. F. (2003). Quality of life and understanding of disease status among cancer patients of different ethnic origin. *British Journal of Cancer, 89*, 641-647.

Urani, M. A. (2005). An examination of the cultural variables affecting WISC-IV scores in an orthodox Jewish population of school-aged children. Unpublished master's thesis, Midwestern University, Downers Grove, Illinois.

Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students.* Thousand Oaks, CA: Sage.

Whaley, A. L., & Davis, K. E. (2007). Cultural competence and evidence-based practice in mental health services: A complementary perspective. *American Psychologist, 62*(6), 563-574.

Yoo, H., Suh, C., Kim, S., Eremenco, S., Kim, H., & Kim, S. (2006). Korean translation and validation of the Functional Assessment of Cancer Therapy-lung (FACT-L) version 4. *Quality Life Research, 15*, 161-166.

Ysseldyke, J., & Thurlow, M. (1984). Assessment practices in special education: Adequacy and appropriateness. *Educational Psychologist, 9*, 123-126.