A RESEARCH REPORT OF
THE RESEARCH CONSORTIUM OF
COUNSELING & PSYCHOLOGICAL SERVICES IN
HIGHER EDUCATION

A Shortened Form of the Outcome Questionnaire:
A Validation of Scores Across Ethnic Groups

by S. Natasha Beretvas and Lisa K. Kearney
Department of Educational Psychology
Augustine Barón
Counseling & Mental Health Center
The University of Texas at Austin

A Shortened Form of the Outcome Questionnaire:

A Validation of Scores Across Ethnic Groups

In the last ten years, a rise in multicultural awareness has spurred examinations of the validity and reliability of scores on psychological scales used to measure personality and intellectual factors for populations of varying ethnicities (e.g., Butcher, 1996).

In an effort to consider linguistic and cultural differences that might affect the validity of these measures with different populations, the use of these instruments with non-Caucasian clients must be evaluated (Butcher, Nezami, & Exner, 1998). This is of the utmost concern when using scores on these measures for clinical decisions, such as diagnosis, treatment planning, or planning of services.

Many of the standard psychological assessment instruments have been normed using primarily European-American clients, resulting in scores on these tests being culturally appropriate primarily for Caucasian, English-speaking individuals, with a Western world view (Dana, 1995). In the last 20 years, many tests, such as the MMPI, Millon Clinical Multi-axial Inventory, and the Rorschach, have been re-examined to assess for validity in cross-cultural and ethnically diverse settings (e.g., Butcher, Braswell, & Raney, 1983; Davis, Greenblatt, & Dahlstrom, 1990; Eysenck & Eysenck, 1983). Such research has often found a pattern of either over- or under-pathologizing of patients from culturally diverse backgrounds. New norms need to be established describing additional populations to insure that such pathologizing does not occur.

One such psychological assessment instrument, The Outcome Questionnaire 45 (OQ45), a symptom and distress inventory developed by Lambert, Lunnen, Umphress, Hansen, & Burlingame (1994), has been found to be useful for examining the effectiveness of psychotherapy over time (Kadera, Lambert, & Andrews, 1996). The instrument was designed to assess "patient functioning" (Mueller, Lambert, & Burlingame, 1998, p. 250) and scores are used to track change in client symptomatology on a session-by-session basis. The OQ45 is widely used in university counseling centers and other mental health settings to evaluate client progress in therapy. Little research has been done that has evaluated the validity and reliability of scores on the OQ45 with non-Caucasian samples. Thus, this study was designed to investigate the functioning of scores on OQ45 items to validate its factor structure across Caucasians and non-Caucasians.

The OQ45

The OQ45 consists of 45 items each associated with the same Likert-type five-point scale. Scores on the whole scale are used as well as scores on each of three subscales, assessing Symptom Distress (SD), Social-Role functioning (SR) and Interpersonal Relationships (IR). The SD, SR and IR subscales consist of 22 items, 9 items, and 11 items, respectively, without the use of items 11, 27, and 40. The SD subscale (items 2, 3, 5, 6, 8, 9, 10, 13, 15, 22, 23, 24, 25, 29, 31, 33, 34, 35, 36, 41, 42, and 45) taps into general emotional and lifestyle stressors such as depression, anxiety, stress, substance abuse, and suicidality. Scores on the SD scale can range from 0 to 88. The SR subscale (items 4, 12, 14, 21, 28, 32, 38, 39, and 44) measures clients' work relations and leisure activities, with scores ranging from 0 to 36. The IR subscale (items 1, 7, 16, 17, 18, 19, 20, 26, 30, 37, 43) assesses clients' satisfaction with interpersonal relationships, especially marital and family relationships and friendships (Kadera, et.al., 1996; Lambert, et. al., 1994). Scores on the IR subscale can range from 0 to 44. Scores on the complete OQ45 scale, with a range of 0 to 180, have been reported to be reliable and valid, distinguishing well between clinical and non-clinical subjects (Umphress, Lambert, Smart, & Barlow, 1997).

Despite its popularity, prior research has indicated some problems with the OQ45. While the creation of the OQ45 subscales was theoretically driven, few empirical validations of the factor structure of these subscales have been conducted. Using confirmatory factor analysis (CFA), Mueller et al. (1998) attempted to confirm the three factors hypothesized to explain responses to the three OQ45 subscales. As the authors pointed out, the fit of the one-, two-, and three-factor models was found to be inadequate. The authors suggested that fit of these models might be improved by discarding some of the items with low factor loadings (Mueller, et al., 1998), though the comparative fit index and normed fit index values of lower than .2 strongly indicated the need for more in-depth analyses.

A second study by Vermeersch, Lambert, & Burlingame (2000) examined the OQ45 for item sensitivity to change. One of the primary purposes of scores on the OQ45 is to track patient change over time in therapy; thus, the researchers wished to establish the sensitivity of each item over time using hierarchical linear modeling. Of the 45 items, 37 items met the authors' criteria for sensitivity to change. The authors suggested that the remaining eight items might not have been sensitive to change because they were more static in nature, tapping into physiological complaints and interpersonal relationships.

Both of these studies indicated the need to re-evaluate the factor structure of the responses to OQ45 items. In addition, neither of these studies evaluated the use of OQ45 scores with ethnically diverse populations for which the structure might also differ.

One study that examined ethnic differences on OQ45 scores found no significant differences in overall scores between Caucasian and African-American clients (Nebeker, Lambert, & Huefner, 1995). However, post-hoc analyses revealed differences for a number of individual items, suggesting that some items may possibly be functioning differently for different populations. This, in combination with the aforementioned questionable factor structure of OQ item scores, led to the design of the current study. This research was designed to identify a factor structure for OQ item scores that could be sustained across both Caucasian and non-Caucasian populations.

Method

The Research Consortium of Counseling and Psychological Services in Higher Education (hereafter referred to as the Research Consortium) was founded in 1990 to further research efforts assessing counseling processes in university settings. The Research Consortium is comprised of 40 to 50 primarily state-supported universities. Enrollment ranges from 2,000 to 50,000 with most schools ranging from 15,000 to 25,000. During the time period of the reported study (1997-1998), 4,679 clients were involved in a therapy outcome study conducted by member centers and agreed to participate by filling out various psychological measures. This study focused solely on participants' responses to the 45 items on the OQ45 scale.

Participants

Participants in the current study were selected from the 4,679 clients who had sought counseling from one of the consortium counseling centers during the 1997 - 1998 school year. Criteria for inclusion in this study were: (1) that clients had consented to participate in the study and (2) that they had filled out the intake questionnaire that requested demographic information. The ethnic breakdown of this sample was: 4.1% African American (n=135), 5.7% Asian American (n=187), 10.4% Hispanic (n=344), .4% Native American (n=13), 73.3% Caucasian (n=2416), and 6.0% International students (n=199). The mean age of students in the sample was 23.38 (SD=5.78). Males comprised 33.4% of the sample while females comprised 66.6%.

Measures

Participants' responses to the OQ45 and demographic information describing the participants were analyzed in this study. The OQ45 consists of 45 items each associated with a 5-point Likert scale including responses ranging from "never" to "almost always". Once responses to the nine negatively worded items have been reverse coded, scores on the overall scale can be tallied such that higher scores represent increasing levels of psychopathology. Previous psychometric evaluations have revealed internal consistency levels of .93 and test-retest reliability of .84 (Kadera, et. al., 1996; Umphress, et. al., 1997). Test-retest reliability for subscale scores have been estimated to range from .78 to .82 with internal consistency estimates from .71 to .92 (Lambert et al., 1994).

Procedure

Participants were asked to complete the OQ45 at their initial intake session and before each subsequent session of therapy. Participants filled out the OQ forms in a manner consistent with the instructions, which asked them to answer the questions "Looking back on the past week including today," (Lambert, et. al., 1994). Participants completed their questionnaires in the waiting room before they were called in to sessions.

Analyses

To investigate the factor structure of scores on OQ45 items, linear exploratory and confirmatory factor analyses were conducted with maximum likelihood estimation using the programs SPSS (2000) and EQS (Bentler, 1993). Initially, the fit of the one- and three-factor models (matching the SD, SR and IR subscales) for responses of all participants was investigated using CFA. Fit indices used included the normed fit index (NFI), the non-normed fit index (NNFI), and the comparative fit index (CFI). Values of .90 or higher were used to define adequate model fit (Tanaka, 1993) for these indices. The root mean square error of approximation (RMSEA) was also used for which a value of less than .06 (Hu & Bentler, 1999) provided evidence of sufficient model fit.

Inadequate model fit was identified for both the one- and three-factor models. It was hypothesized that perhaps the factor structure across the ethnicities might differ. Thus, it was decided to evaluate the fit of the same pair of models using a solely Caucasian sample. Unfortunately, neither model fit the data for this sample.

Instead of resorting immediately to the use of CFA modification indices to assist in an exploration of the factor structure, the 45 items were then content-analyzed and categorized to determine the number of potential factors. Seven factors were identified. Following this, an exploratory factor analysis (EFA) was conducted to investigate the pattern of items' loadings on seven correlated factors. The use of EFA as a precursor to further investigation of factor structure in CFA was advocated by Gerbing and Hamilton (1996). They noted that "EFA is a useful tool to aid the researcher in recovering an underlying measurement model that can then be evaluated with CFA" (1996, p. 71). The model identified through the EFA results was then explored further using CFA to establish a final model with acceptable model fit values. This final model was cross-validated using responses of another Caucasian sample and then assessed for fit to the responses of Hispanic, Asian-American and African American samples.

<div align="center">Results</div>

One and Three-Factor Models

When scores on the OQ45 are used to assess the impact of therapy, they are typically summed together either 1) across the 45 items to represent the scale score or 2) summed together within each of the SD, SR and IR subscales. Underlying these summations is the assumption that the responses are assessing either 1) a single dimension of patient functioning or 2) three correlated dimensions of patient functioning. These two assumptions were assessed through a CFA of 3,143 participants' responses to the OQ45.

As evidenced by the values of the model fit indices for these two models (see Table 1), neither model fit the data. The same two models were then evaluated for fit to the responses of a solely Caucasian sample. Since these responses also did not fit the data (see Table 1), a content analysis of the 45 items was then conducted to assist in estimating how many dimensions the OQ45 items might be assessing.

Table 1

Model Fit Indices for CFAs

Model	Sample	χ^2	DF	NFI	NNFI	CFI	RMSEA
One-Factor, all items	All	20483.1	945	.61	.60	.62	.08
	(*n* = 3,143)						
Three Factors, all items	All	16257.2	816	.66	.66	.68	.08
	(*n* = 3,143)						
One Factor, all items	Caucasian	10788.3	945	.62	.62	.64	.08
	(*n* = 1,634)						
Three Factors, all items	Caucasian	8577.2	816	.67	.68	.69	.08
	(*n* = 1,634)						
Seven Factors, all items	Caucasian	6886.7	925	.76	.77	.78	.06
	(*n* = 1,634)						
Six Factors, 33 items	Caucasian	3567.2	480	.83	.84	.85	.06
	(*n* = 1,634)						
Four Factors, 21 items	Caucasian	2153.0	246	.87	.87	.88	.07
	(*n* = 1,634)						

EFA – Caucasian Sample

As a result of the content analysis, it was hypothesized that there might be seven factors underlying scores on the OQ45. An EFA was then conducted using the Caucasian sample (*n* = 1634) specifying that seven inter-correlated factors should be extracted. A cutoff of .40 was used to identify items that loaded "substantially" on a factor, those items that were then used to define the factor. The seven factors closely matched those that had been hypothesized a priori and included the following: a) Depression/Anxiety (items 5, 8, 10, 15, 23, 25, 33, 36, 40), b) Feelings of Well-being (items 1, 12, 13, 20, 21, 24, 31, 43), c) Impact of Stress (items 2, 3, 4, 22, 28, 38), d) Alcohol Abuse (items 11, 26, 32), e) Conflict Management Skills (items 19, 30, 39, 44), f) Somatization (27, 29, 36, 41, 45), and g) Romantic Relationship Satisfaction (7, 17, 18, 37). Only six items did not have loadings higher than .40 on any of the seven factors. In the CFA that followed, each

of these items were reviewed and assigned to the factor that best matched what the item seemed designed to assess. Items 6, 35 and 42 were added to the list of items assessing Depression/Anxiety. Item 14 was hypothesized to assess the Impact of Stress. Item 16 was considered a measure of Conflict Management Skills, while item 9 appeared to assess Somatization. (See Table 2 for a list of items and their associated item numbers).

Table 2

Complete Set of OQ45 Items

---

1. I get along with others.

2. I tire quickly.

3. I feel no interest in things.

4. I feel stressed at work/school.

5. I blame myself for things.

6. I feel irritated.

7. I feel unhappy in my marriage/significant relationship.

8. I have thoughts of ending my life.

9. I feel weak.

10. I feel fearful.

11. After heavy drinking, I need a drink the next morning to get going. (If you do not drink mark "never").

12. I find my work/school satisfying.

13. I am a happy person.

14. I work/study too much.

15. I feel worthless.

16. I am concerned about family troubles.

17. I have an unfulfilling sex life.

18. I feel lonely.

19. I have frequent arguments.

20. I feel loved and wanted.

---

21. I enjoy my spare time.

22. I have difficulty concentrating.

23. I feel hopeless about the future.

24. I like myself.

25. Disturbing thoughts come into my mind that I can't get rid of.

26. I feel annoyed by people who criticize my drinking (or drug use). (If not applicable mark "never").

27. I have an upset stomach.

28. I am working/studying less well than I used to.

29. My heart pounds too much.

30. I have trouble getting along with friends and close acquaintances.

31. I am satisfied with my life.

32. I have trouble at work/school because of drinking or drug use. (If not applicable mark "never").

33. I feel that something bad is going to happen.

34. I have sore muscles.

35. I feel afraid of open spaces, or of driving, or being on buses, subways, etc.

36. I feel nervous.

37. I feel my love relationships are full and complete.

38. I feel that I am not doing well at work/school.

39. I have too many disagreements at work/school.

40. I feel something is wrong with my mind.

41. I have trouble falling asleep or staying asleep.

42. I feel blue.

43. I am satisfied with my relationships with others.

44. I feel angry enough at work/school to do something I might regret.

45. I have headaches.

CFA – Caucasian Sample

A CFA was conducted using the Caucasian sample's responses assuming the seven inter-correlated factor model described above. The model did not fit the data (see Table 1). All items that had loadings lower than .50 were deleted from the model. This meant that only two items (18 and 37) remained to define the Romantic Relationships Satisfaction factor. It was felt that the factor would not be reliably measured with only two items. Therefore the two remaining Romantic Relationship items were also deleted from the model. The resulting six-factor model did not fit the data (see Table 1).

The Conflict Management Skills factor was re-evaluated. It appeared that the content of item 44 ("I feel angry enough at work/school to do something I might regret) did not quite fit with the other three items (for example, "I have too many disagreements at work/school"). In addition, the Somatization factor was carefully re-assessed. Items 27 ("I have an upset stomach") and 34 ("I have sore muscles") seemed only to provide weak reference to potential somatization of psychological dysfunction. While the Somatization factor would have provided a useful contribution to the assessment of patient functioning with the omission of items 27 and 34, only three items would remain to assess the factor. It was decided that these items under-represented the construct of Somatization and so the factor was deleted.

The remaining four factors included Depression/Anxiety, Feelings of Well-Being, Impact of Stress, and Alcohol Abuse assessed by 6, 6, 6 and 3 items, respectively. Two more items were deleted from the Depression/Anxiety factor to equate the number of item assessing each of the first three dimensions. This final model was found to fit the data marginally well (see Table 1). Table 3 lists the items selected for this final model and their associated factor loadings. The values of the factor loadings ranged from .546 to .819 and each was significantly different from zero ($p < .001$). The correlations between the four factors ranged from .223 to .783. The strongest correlation was that between the Depression/Anxiety and the Feelings of Well-Being factors ($r = .783$, $p < .001$). (It should be remembered that the Well-Being items were reverse-coded implying that the stronger the feelings of Depression, the less the feelings of well-being). Also, as expected, there were moderately strong correlations between the Impact of Stress factor and the Depression/Anxiety factor ($r = .697$, $p < .001$) and with Feelings of Well-Being ($r = .551$, $p < .001$). The correlations between the Alcohol Abuse Effects factor and each of the other factors were low (ranging from .223 to .260).

Table 3

Factor Loadings for the Four-Factor Model, Using the Caucasian Sample's Responses

Depression/Anxiety Factor		Loading
5	I blame myself for things.	.603
8	I have thoughts of ending my life.	.573
10	I feel fearful	.561
15	I feel worthless.	.819
23	I feel hopeless about the future.	.735
40	I feel something is wrong with my mind.	.640
Feelings of Well-Being Factor		
13	I am a happy person.	.757
20	I feel loved and wanted.	.678
21	I enjoy my spare time.	.596
24	I like myself.	.779
31	I am satisfied with my life.	.817
43	I am satisfied with my relationships with others.	.625
Impact of Stress Factor		
2	I tire quickly.	.546
4	I feel stressed at work/school.	.563
12	I find my work/school satisfying.	.568
22	I have difficulty concentrating.	.677
28	I am working/studying less well than I used to.	.626
38	I feel that I am not doing well at work/school.	.620

Alcohol Abuse Factor		
11[a]	After heavy drinking, I need a drink the next morning to get going.	.589
26	I feel annoyed by people who criticize my drinking (or drug use).	.745
32	I have trouble at work/school because of drinking or drug use.	.708

Note. Respondents chose responses to each item from a 5-point Likert scale ranging from "never" to "almost always".

[a]*For the three Alcohol Use items, respondents were prompted to select "never" as their response if they did not drink nor use drugs.*

The following internal consistency reliability (Cronbach's alphas) for Caucasian participants' scores on the Depression/Anxiety, Feelings of Well Being, Impact of Stress, and Alcohol Abuse subscales were obtained: .81, .86, .76 and .70, respectively. Test-retest reliability coefficients for the Caucasian sample were as follows: Depression/Anxiety – .60, Feelings of Well-Being – .63, Impact of Stress – .55, and Alcohol Abuse – 1.00.

Cross-Validation of Final Model

The final model was cross-validated using the final, reported OQ45 scores of Caucasians. The model fit this data (χ^2 (183, n = 1,274) = 1052.5, $p < .001$; NFI = .93; NNFI = .93; CFI = .94; RMSEA = .06) with very similar patterns of loadings and inter-correlations.

The four-factor, 21-item model was also assessed for fit to three separate samples: a sample of Hispanic (n = 218), Asian-American (n = 119), and African-American (n = 93) participants. (It should be emphasized that the Asian- and African-American sample sizes were very small for the number of parameters being estimated in this model. The results for those two samples should be considered very preliminary).

The results supported adequate fit of the model to the responses of the Hispanic participants (χ^2 (183, n = 218) = 346.9, p < .001; NFI = .83; NNFI = .90; CFI = .91; RMSEA = .06). The models' fit for the small samples of Asian- and African-Americans did not quite indicate adequate fit (χ^2 (183, n = 119) = 309.8, p < .001; NFI = .70; NNFI = .82; CFI = .85; RMSEA = .08 for the Asian-American sample; and χ^2 (183, n = 93) = 261.7, p < .001; NFI = .73; NNFI = .88; CFI = .90; RMSEA = .07 for the African-American sample).

A multi-group CFA was conducted constraining factor loadings and inter-correlations to be equal for the Hispanic and Caucasian samples. The model fit the data sufficiently well (χ^2 (393, Caucasian n = 1,274; Hispanic n = 218) = 1531.3, p < .001; NFI = .91; NNFI = .92; CFI = .93; RMSEA = .04). It was hypothesized that the uneven sample sizes might have weighed the fit of the model more strongly in favor of the Caucasians. Therefore another multi-group analysis was conducted using a more equivalently sized, random sample of Caucasians. The model constraining the loadings and inter-correlations between factors still fit the data adequately (χ^2 (393, Caucasian n = 306; Hispanic n = 218) = 871.7, p < .001; NFI = .85; NNFI = .91; CFI = .91; RMSEA = .05).

Means and standard deviations for Hispanic students' responses to each of the subscales were calculated: Depression/Anxiety (M = 9.85, SD = 5.19), Feelings of Well Being (M = 10.70, SD = 4.81), Impact of Stress (M = 14.07, SD = 4.76), and Alcohol Abuse (M = .76, SD = 1.73). Statistics describing the responses of Caucasian students included the following: Depression/Anxiety (M = 9.69, SD = 5.16), Feelings of Well Being (M = 10.82, SD = 4.93), Impact of Stress (M = 13.10, SD = 4.80), and Alcohol Abuse (M = .64, SD = 1.42).

A MANOVA was conducted to investigate differences between Hispanics and Caucasians across the four subscales (Wilk's Lambda = .993, F = 4.88, p < .001). Post-hoc analyses indicated that the only significant difference detected between the two groups was for scores on the Impact of Stress scale (F = 13.26, p < .001), with Hispanic students reporting significantly higher on the Impact of Stress factor in comparison to Caucasian students.

Discussion

This study was designed to contribute to the validation of scores on the OQ45. As has been intimated in past research, it was found that the one- and three-factor models conventionally associated with scores on the OQ45 did not fit the data. In the original sample used to norm scores on the OQ45, 93.8% of the respondents were Caucasian (Lambert et al., 1994). Yet, responses of an all-Caucasian sample in our study did not fit either model. This led to a rather rigorous shortening of the scale resulting in the ultimate deletion of 24 of the 45 items. Four related dimensions appeared to explain responses to the remaining 21 items. These factors were named Depression/Anxiety, Feelings of Well-Being, Impact of Stress, and Alcohol Abuse. While the set of items comprising the original OQ45 appear to assess a wider range of psychological distress, it appears that they might be too lacking in focus when considered as a whole. This study attempted to pool together the items that seemed to assess more specific dimensions of patient functioning.

It seemed a pity to omit the items assessing the Somatization factor. However, as noted, the scope of the symptoms that were being assessed seemed to extend beyond a single dimension, measuring such variables as upset stomach, pounding heart, nervousness, sleeping difficulties, and headaches. There appeared to be an intermingling of a number of possible psychiatric symptoms that could range from depression to anxiety disorders. Such a range in symptomatology makes somatization items difficult to identify as a single dimension. In addition, since some of these symptoms may be representative of more than one disorder, it would not be possible to place the items within the factor describing the relevant, associated disorder. Unfortunately, by omitting such items, we may be losing important data for clinicians who may specifically work with populations who may more readily report somatic rather than psychological symptomatology. For example, Asian Americans are more likely to report physical distress because mental health illnesses carry a greater stigma within their cultural context (D. Sue, 1997; Toarmino & Chun, 1997).

As Messick (1989) emphasized, when evaluating the validity of scores on a test, the *use* of the test's scores must also be considered. The OQ45 provides an efficient assessment of a patient's *general* psychological functioning. While many of the items from the original scale were deleted, this shortened form could prove more practical for a counselor who might be interested in repeated measures of a patient's functioning, eliminating the frustration of completing longer forms at each and every session. In

addition, the more specific focus of the subscales could be especially useful for practitioners working within college populations. The four subscales measuring depression/anxiety, well-being, impact of stress, and alcohol abuse might prove more appropriate with this population since these four categories represent some of the most common reasons why college students seek counseling (Miller & Rice, 1993; Surtees, Pharoah, & Wainwright, 1998). Additionally, because university counseling centers typically serve clientele with greater psychological functioning than other outpatient counseling centers (Erickson-Cornish, Riva, Cox-Henderson, Kominars, & McIntosh, 2000; Sharkin, 1997), the shortened measure may prove most effective as it is a general measure of psychological distress and well-being. For example, many of the students seen in university counseling centers may not have diagnoses on Axis I or II, but rather are reporting common stressors related to college life and diminished well being (i.e. V codes). Further, because university counseling centers often employ short-term models of therapy (Tyron, 1999), a shortened scale which takes less time to complete may also be of benefit to practitioners, whose time with clients is especially precious.

Limitations and Directions for Future Research

The four-factor model was found to hold adequately using the responses of Hispanic participants. However, this finding should be considered preliminary due to the moderate sample size used. Yet our results hold hope for the invariance of the four-factor structure across at least Caucasian and Hispanic college students. The sample sizes for the Asian- and African-Americans also restrict the generalizability of the findings that had indicated a lack of model fit. The values for the test-retest score reliabilities for scores on the four subscales were low, though there are several explanations for this. The types or foci of the therapy employed were not consistent across counseling centers included in this study. Some of the therapies might have impacted the constructs being assessed by the four subscales, while other forms of therapy might have been designed to affect other outcomes. In addition, the four subscales do not represent the full gamut of possible psychological dysfunctions for which students had sought help. Lastly, the re-test scores were those of the last session attended in 1998. For some students, this might have been their last session, for others, only a session in the middle or end of their full complement of sessions across episodes of treatment. None of these explanations could be further investigated with the data available in this study. Future research should focus on further evaluation of the test-retest reliability of scores on these subscales.

Clearly more research needs to be conducted assessing the factor structure of scores on OQ45 items. In particular, further research should evaluate the invariance of the factor structure with populations more diverse in terms of ethnicity and beyond the sample of college students, who frequently report higher levels of psychological functioning.

References

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Butcher, J.N., Braswell, L., & Raney, D. (1983). A cross-cultural comparison of American Indian, Black, and White inpatients on the MMPI and presenting symptoms. *Journal of Consulting and Clinical Psychology, 51*, 587-594.

Butcher, J.N. (1996). *Handbook of international MMPI-II resear*ch. Minneapolis, MN: University of Minnesota Press.

Butcher, J.N., Nezami, E., & Exner, J. (1998). Psychological assessment of people in diverse cultures. In S.S. Kazarian & D.R. Evans (Eds.), *Cultural clinical psychology: Theory, research, and practice* (pp. 61-105). New York, NY: Oxford University Press.

Dana, R.H. (1995). Impact on the use of standard psychological assessment on the diagnosis and treatment of ethnic minorities. In J.F. Aponte, R. Y. Rivers, & J. Wohl (Eds.), *Psychological interventions and cultural diversity* (pp. 57-73). Boston, MA: Allyn and Bacon.

Davis, W.E., Greenblatt, R.L., & Pochyly, J.M. (1990). Test of MCMI black norms for five scales. *Journal of Clinical Psychology, 46*, 175-178.

Erickson-Cornish, J.A., Riva, M.T., Cox-Henderson, M., Kominars, K.D., & McIntosh,S. Perceived distress in university counseling center clients across a six-year period.Journal of College Student Development, 41(1), 104-109.

Eysenck, H.J., & Eysenck, S.B.G. (1983). Recent advances in the cross-cultural study of personality. In J.N. Butcher & C.D. Spielberger (Eds.), *Advances in personality assessment* (pp. 41-69). Hillsdale, NJ: Lawrence Erlbaum.

Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling, 3*, 62-72.

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

Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1996). How much therapy is really enough: A session-by-session analysis of the psychotherapy dose-effect relationship. *Journal of Psychotherapy Practice and Research, 5*, 132-151.

Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N. & Burlingame, G. M. (1994). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.1).* Salt Lake City: IHC Center for Behavioral Healthcare Efficacy.

Messick, S. (1989). Validity. In R. L. Linn (Ed.) *Educational Measurement*: (3rd ed., pp. 13-104). New York: American Council on Education and Macmillan Publishing Company.

Miller, G.A., & Rice, K.G. (1993). A factor analysis of a university counseling center problem checklist. *Journal of College Student Development, 34(2)*, 98-102.

Mueller, R.M., Lambert, M.J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of PersonalityAssessment, 70(2),* 248-262.

Nebeker, R.S., Lambert, M.J., & Huefner, J.C. (1995). Ethnic differences on the Outcome Questionnaire. *Psychological Reports, 77*, 875-879.

Sharkin, B.S. (1997). Increased severity of presenting problems in college counseling centers: A closer look. *Journal of Counseling & Development, 75(4)*, 275-281.

SPSS. (2000). *SPSS base 10.0 users' guide*. Chicago: Author.

Sue, D. (1997). Counseling strategies for Chinese Americans. In C.C. Lee (Ed.), Multicultural issues in counseling (2nd ed., pp 173-187) Alexandria, VA: American Counseling Association. (PUT IN APA FORMAT!)

Surtees, P.G., Pharoah, P.D., & Wainwright, N.W. (1998). A follow-up study of new users of a university counselling service. *British Journal of Guidance & Counselling, 26(2),* 255-272.

Tanaka, (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

Toarmino, D., & Chun, C.A. (1997). Issues and strategies in counseling Korean Americans. In C.C. Lee (Ed.), Multicultural issues in counseling, pp. 233-254.

Tyron, G. S. (1999). Counseling dropout relative to client attractiveness, disturbance, and expected number of sessions. *Counseling Psychology Quarterly, 12(3),* 285 – 291.

Umphress, V. J., Lambert, M. J., Smart, D. W. & Barlow, S. H. (1997). Concurrent and construct validity of the outcome questionnaire. *Journal of Psychoeducational Assessment, 15,* 40-55.

Vermeesch, D.A., Lambert, M.J., & Burlingame, G.M. (2000). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74(2),* 242-261.