Published in final edited form as:

Adm Policy Ment Health. 2011 March ; 38(2): 120–130. doi:10.1007/s10488-011-0337-0.

The Nathan Kline Institute Cultural Competency Assessment Scale: Psychometrics and Implications for Disparity Reduction

Carole E. Siegel,

Nathan S. Kline Institute of Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Gary Haugland,

Nathan S. Kline Institute of Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Eugene M. Laska,

Nathan S. Kline Institute of Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Lenora M. Reid-Rose,

Coordinated Care Services, Inc., 1099 Jay Street, Bldg J, Rochester, NY 14611, USA

Dei-In Tang,

Nathan S. Kline Institute of Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Joseph A. Wanderling,

Nathan S. Kline Institute of Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Ethel D. Chambers, and

PO Box 26, Ramble PO, Hanover, Jamaica, West Indies

Brady G. Case

Nathan S. Kline Institute of Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Carole E. Siegel: siegel@nki.rfmh.org

Abstract

The NKI Cultural Competency Assessment Scale measures organizational CC in mental health out-patient settings. We describe its development and results of tests of its psychometric properties. When tested in 27 public mental health settings, factor analysis discerned three factors explaining 65% of the variance; each factor related to a stage of implementation of CC. Construct validity and inter-rater reliability were satisfactory. In tests of predictive validity, higher scores on items related to linguistic and service accommodations predicted a reduction in service disparities for engagement and retention outcomes for Hispanics. Disparities for Blacks essentially persisted independent of CC scores.

Keywords

Cultural competency; Organizational level measurement; Mental health; Disparities; Cultural groups

Introduction

Substantial inequities in mental health care to the cultural groups of African Americans, Latinos, Asian Americans, Pacific Islanders and Native Americans were highlighted in the Supplement to Mental Health of the 2001 Surgeon General's report (USPHS 2001). Cultural groups at risk for differential mental health treatment more broadly include those whose societal standing, world view or values might impede their access to and receipt of services, and are the groups of concern in this report. While there is no universally endorsed definition of cultural competency (CC), one that is broad and commonly cited describes it as the set of congruent behaviors, attitudes and skills, policies and procedures that come together in a system, agency or individual, to enable mental health caregivers to work effectively and efficiently with cultural groups (New York State Office of Mental Health 1998, adapted from Cross et al. 1989). Underlying most definitions is the premise that as CC increases in a health care environment, members of cultural groups will experience better outcomes (DHHS 2003), and service inequities will decline.

Over the past 15 years, efforts have been made by federal, state and various cultural group organizations to specify the personal attributes of individual providers and the policies and procedures of mental health organizations that might improve the delivery of mental health care to members of cultural groups. Several theoretical frameworks (Sue et al. 1992) along with measurement scales (Campinha-Bacote 1998; Mason 1995; Goode and Tawara 2002) for examining personal attitudes, behaviors and skills of individual care providers have been proposed. There are also scales that measure CC on an organizational level. A recent review (Harper et al. 2006) includes 17 such scales, two of which are reported to have been psychometrically tested. However, these results are not yet published. In general health care, there are only two peer-reviewed studies on the psychometric properties of CC scales, one for general health care providers (Doorenbos et al. 2005) and the other for a patient report scale (Lucas et al. 2008). Thus, in both general and mental health, there remains a dearth of scientifically validated tools to measure organizational CC (Roper and Mays 2000).

In this paper, we describe the development of the Nathan Kline Institute (NKI) Cultural Competency Assessment Scale (CCAS) that was designed to examine organizational level CC in mental health outpatient settings. Its items were selected to provide a roadmap for introducing CC into an agency. The results of tests of the scale's psychometric properties, based on data on mental health care in public clinics that serve substantial numbers of Hispanics, Blacks and Whites, are presented.

Methods

The development of the CCAS was guided by a panel of psychiatrists, social workers, nurses, agency administrators, consumers and family members. Many panel members had previously served on national committees and work-groups addressing CC. Members included six African-Americans, four Asian-Americans, three Latinos, two American Indian-Alaska Natives and four Caucasians.

Scale Development

Prior work: Performance Measure Selection and Benchmarking—We briefly summarize highlights of the background work preceding the development of the CCAS, as already reported in two articles (Siegel et al. 2000; Siegel et al. 2003) in this journal. The CCAS is comprised of selected performance measures of CC identified by the panel and benchmarked in a subsequent study. Researchers identified 231 indicators of CC from the published literature and from national reports. The panel selected 85 of these judged to have high importance in a CC process, and for which it was feasible to collect reliable data. From these, the 44 organizational level measures were selected for benchmarking. The focus on the organizational level as a starting point was based on the consensus that achievements of CC at this level would permeate the environment and impact all staff. To benchmark the items, panelists identified 21 exemplary agencies which in their estimation had already introduced organizational level practices related to CC with some success. While selection criteria were not precisely defined, these agencies were known by the panelists to have longevity and to be well accepted in their communities. Today, these features would be included in a list of criteria for programs having community-based evidence of their effectiveness (Isaacs et al. 2005). A structured questionnaire investigating the extent to which the 44 items had been adopted by these agencies was administered by telephone. Survey results informed the categories of the rating scales for each of the criterion of the CCAS.

Current work: CCAS Criteria Selection and Pilot Testing of the Scale

Criteria: Study investigators and panelists grouped performance measures into eleven criteria intended to provide a roadmap for introducing cultural competency into an agency. To illustrate, the first criterion of the scale outlines initial preparation: the agency adopts a CC mission statement, provides a dedicated CC budget, and assigns a person or group to be responsible for CC activities. The last criterion in the scale indicates a high level of awareness and resources for CC activities and would indicate that the agency is far along in the process of introducing CC. The CCAS queries whether existing services are reviewed and adapted for the cultural groups that an agency serves, and if new services are introduced to fit their needs. For each criterion, there is a scale based on four levels of desirable activities ranked by increasing difficulty of achievement. Each level requires a specific activity to occur as well as all activities identified in the lower levels. Because the CCAS was used in the SAMSHA project, The Implementation of Evidence Based Practices Toolkits (Siegel et al. 2005), the format of the criteria ratings are consistent with the format of the scales of the fidelity instruments developed by Bond et al. (2000). Table 1 lists the 11 CCAS criteria and the rationale for their selection. The complete instrument with its instructions for use is available at <http://cecc.rfmh.org>.

Pilot Testing: Pilot tests were conducted to ensure the clarity of wording and the feasibility of administration. Reviewers were also asked to examine the logical sequencing of the events in the rating of each criterion. Seven agencies in Monroe County, New York participated in the first pilot. The Monroe County Department of Mental Health contracts with agencies for the delivery of behavioral health care services. The county is substantially middle class with pockets of poverty. In 2000, there were 735,343 residents of whom 79% were white, 14.7% were Black and 5.3% were Hispanic of any race. CC assessments have been conducted in county agencies since 1999 as part of contractual obligations, so agency personnel are familiar with its concepts. In a second pilot, the scale was reviewed by staff of the Charles B. Wang Community Health Center, an Asian-American agency in lower Manhattan. Center staff recommended a revision to make the scale more applicable to monocultural sites. Subsequently, a modified CCAS form was developed for sites serving Chinese clients. The third pilot was conducted with researchers from the Center for

Multicultural Mental Health Research, in collaboration with the Cambridge Health Alliance (CHA), a multi-site/setting community agency affiliated with Harvard Medical School. Scales were administered at five outpatient clinics that serve either multiple cultural groups comprised of Latinos from Central and South America, Portuguese and Asians, or Latinos exclusively. Changes that resulted from the pilots included a reordering of some scale criterion categories, the introduction of the concept of a “parent organization” (PO) of the agency where some CC activities may take place, alteration of the definition of a prevalent cultural group from one based purely on a population size threshold to one determined by the agency, and addition of more complete instructions to the scale.

Psychometric Study

Agency Sample: 27 contract mental health service agencies of the Monroe County Office of Mental Health participated. 11 provide clinic care to adults and six to children and youth. Seven provide day treatment care and three provide intensive psychiatric rehabilitation. Coordinated Care Services, Inc. (CCSI) manages the contracts of these agencies and maintains a mental health information system on their outpatient service contacts which was made available for the study. The study protocol was reviewed and approved by the IRB of the Nathan S. Kline Institute for Psychiatric Research who also served as the IRB for the Monroe County sites.

Psychometric Procedures: We examined the underlying dimensionality of the CCAS, construct validity, inter-rater reliability and predictive validity using standard statistical approaches described below. Other statistical test procedures that we considered but rejected include Cronbach’s alpha, which assumes a one-dimensional latent structure, and Item Response Theory, which requires a much larger sample size than we had available.

Factor Analysis: A principal components factor analysis of the CCAS scores of the 27 sites followed by a varimax rotation were used to identify the number of factors that account for a reasonable amount of the variance. Items that had the highest loading scores on the principal factors were identified followed by confirmatory factor analysis to substantiate the fit of the factor structure.

Construct Validity: A 97 item checklist developed by CCSI and in use for four years at the participating sites was chosen as the ‘gold-standard’ of measurement of the CC construct. It was chosen because it contained most of the items cited in other performance measure lists. The items in the checklist were independently mapped by two researchers onto CCAS criteria and through discussions, disagreements were reconciled. Checklist items were considered primary if their wording was similar to the CCAS criterion and secondary if the intent was similar but the wording differed. For example the CCSI item; “Evidence that cultural competence is an integral part of employee evaluation and performance appraisal” was considered secondary to the CCAS criterion: “Recruitment, hiring and retention of staff from/or experienced with the most prevalent cultural groups of service users.”

The score for the Checklist items associated with a particular CCAS item is the weighted average of the primary and secondary items (2/3, 1/3 weights). If no secondary items were identified, the score was the average of the associated primary items. The total score for the CCSI Checklist is the average of the checklist scores for the 11 CCAS category items, averaged in with the scores on the remaining items that had not been associated with any CCAS item.

A “local rater,” who was a senior quality assurance employee or senior administrator of the agency, and two “expert raters,” who were members of the research project staff, rated the sites utilizing both the Checklist and the CCAS. The local raters assessed their own sites

with the CCAS. At approximately half of the agencies, one of the expert raters administered the CCAS and the other rater administered the Checklist. The raters switched instruments in evaluating the remaining sites. Each rater was blind to the scoring of the other. The ratings of the local and expert raters were combined to create a “consensus” score. If the two scores were only one unit apart the consensus score is the average of local and expert rater scores. If they were more than one unit apart, the item was re-scored by a third researcher blinded to the original scores based on auxiliary written information obtained from the agency. Construct validity was estimated by Pearson correlation coefficients (Kraemer 1976) of Checklist ratings and the CCAS consensus scores for both criterion specific and Total scores.

Inter-Rater Reliability: Since conducting independent rater assessments by many raters at the 27 agency sites would have been disruptive and logistically impractical, simulated site scenarios were developed to be rated in a process analogous to using video tapes in assessment of clinical instrument reliability (Siegel et al. 2007). Expert site ratings along with supplementary agency materials were used by study researchers to develop the scenarios. To increase the expected variability in ratings among sites, one additional scenario was added describing a vastly different situation from the others. These were independently reviewed and revised by another Panel member. Training of local raters involved rating and discussion of five distinct training scenarios. The first one was rated by the group as a whole, and the remaining scenarios rated individually and discussed by the group. Raters felt the training scenarios did not always provide sufficient detail to evaluate each criterion, but did project a view of the overall level of CC as would be reflected by the Total Score. Over a four month period after their training, each local rater scored each criterion for each of the 27 test scenarios. These were averaged to obtain the Total score that was used to estimate the intra-class correlation (ICC), a measure of inter-rater reliability.

The ICC was obtained from a statistical model in which raters were considered random and scenarios were considered fixed effects (Shrout and Fleiss 1979). The data were analyzed using a mixed random effects complete block two-way (raters by scenarios) ANOVA. The Spearman-Brown Prophecy formula (Nunnally and Bernstein 1994) was used to estimate the gain in reliability that would be obtained were a rating based on the average of the scores of multiple raters.

Predictive Validity

Service Use Outcomes: We hypothesized that culturally competent programs will do a better job of engaging and retaining clients from cultural groups, and therefore disparities in service outcomes will decrease as CCAS scores increase. Two binary service use outcome measures were identified. The first is an *engagement outcome*: does the consumer drop out of service in the first month after admission? The second is a *retention outcome*: does a consumer receive at least four services of any type during the first six months after admission. The one month and four services ‘thresholds’ were chosen by evaluation specialists working with Monroe County. In their judgment, the thresholds represent minimally acceptable standards of care.

Service User Sample: Data are drawn from the service encounters of a first admission cohort entering the agencies between 5/1/2005 and 4/30/2006 ($n = 10,879$), a one year period centered at the time of the CCAS assessment of the sites. Two of the 27 sites (both continuous day treatment) were removed from analyses as one had no new admissions and the other recorded services monthly so that study outcome measures could not be obtained. There were 456 individuals in the sample admitted to multiple study site clinics within 182 days of the index admission. These clients were removed as their outcomes were impacted

by exposure to more than one clinic with possibly different levels of CC. Another 210 sample members were both admitted and discharged on the same day and one individual was missing gender specification. These cases were removed resulting in a sample of $N = 10,193$ individuals entered into the calculation of the service engagement outcome. In the analysis of the retention outcome, an additional 156 sample members who received crisis or inpatient visits within the first month after their admission visit were excluded, resulting in a sample of $N = 10,037$. They were excluded because it was expected that their subsequent utilization patterns would tend to be more intense than those who did not require such services, a phenomenon that would be independent of CC of an agency. Table 2 presents some summary statistics for the larger sample. Across the 25 settings, the percent who were Hispanic ranged from 0 to 23% (mean = 10%), and the percent who were Black ranged from 0.09–51% (mean = 28%). The large majority of subjects in the sample received services in adult clinics.

Statistical Model: A logistic hierarchical linear regression model was employed to measure whether the CCAS score is a predictor of a service outcome. A random effects model was utilized since service encounter observations are likely to be correlated within an agency. In the model, the CCAS score is an independent variable along with other covariates related to client characteristics and to service environments. The level-1 model includes client covariates: age, gender, cultural group and diagnosis. Patients' principal diagnosis were classified into four groups: Mood (International Classification of Disease-9 codes 296.00–296.99, 300.40–300.49, 311.00–311.99), Psychotic (codes 295.00–295.99, 297.00–297.99, 298.00–298.99), Anxiety (codes 300.00–300.02, 300.21–300.23, 300.29–300.30, 308.30, 309.81) and Other (all other codes). While variables related to the socioeconomic status of a client were in the data set, these could not be included because of the extent of missing data. The level-2 model included agency covariates: CCAS score, the number of clients served, percent Black, percent Hispanic and type of setting. Interaction terms were included for CCAS score by cultural group. Mutually exclusive self-reported cultural group categories were: Black, Hispanic, White and Other. The models were fit for each of the 11 CCAS criterion and for the total score.

The odds ratio of the outcome measure is defined as the odds of a positive outcome on the binary outcome measure for Whites divided by the odds for a cultural group by diagnosis. The odds ratio was calculated from the model for each group. It is a measure that reflects a divergence in positive service outcome rates of whites compared to each of the groups. (An important caveat which we discuss later is that a divergence may not be a disparity.) Values of the odds ratio greater than one indicate Whites have a higher likelihood of a positive service outcome than the comparison group. If the model indicates that an increase in the CCAS score corresponds to a decrease in the odds ratio, then predictive validity has been demonstrated. We report below only those tests that were statistically significant at the $P < 0.05$ level. No adjustments were made for multiplicity.

Results

Agency CCAS Scores

Figure 1 presents the frequency distribution of the total CCAS scores across the 27 sites along with the mean, standard deviation and quartiles of the total scores. The total scores ranged from 23 to 47 (the highest achievable score is 55). Both the mean and median are approximately 37. Table 3 reports the means and standard deviations across the 27 sites for each CCAS item. The mean of criterion scores ranged from 2.09 to 4.56. The highest scoring item was 'having data' and the lowest scoring item was 'having translated/formatted service and educational material'.

Factor Analysis

Three factors accounted for 65% of the variance. The items with the highest loadings for factor 1 were: ‘agency commitment’, ‘having a CC committee’, ‘integration of CC committee in agency’, and ‘conducting staff training’; for factor 2 were: ‘having data’, ‘recruiting/hiring/retention’ and ‘having translated/formatted service and educational materials’; and for factor 3: ‘having interpreters’, ‘having bilingual staff’, ‘having translated key forms’, and ‘assessing/adapting/having new services’. A confirmatory factor analysis on these three factors confirmed the fit (Chi square ns, $P = 0.24$, root mean square estimate = 0.08).

Construct Validity

Table 3 also reports the Pearson correlation coefficient between CCAS and Checklist scores. The criteria ‘agency commitment’, ‘integration of CC committee in agency’, ‘conducting staff training’, ‘having bilingual staff’ and ‘having translated key forms’ showed substantial significant correlation (0.44–0.58), but other items did not. The correlation for total scores was 0.37 which significantly differed from zero.

Inter-Rater Reliability

The overall estimate of the ICC is 0.47 with a 95% confidence interval of (0.36, 0.63). Using the Spearman-Brown Prophecy formula, if the average of the total score of two raters is used, the ICC is predicted to increase 36% to 0.64.

Predictive Validity

Model based estimates of service use outcomes indicate that for all diagnostic groups, Hispanics did not significantly differ from Whites. For all diagnostic groups except psychoses, Blacks were significantly less likely than Whites to have either of the two positive service events occur (See Figs. 2, 3).

For the engagement variable, age and gender were the only client covariates that were significant, with older persons and females more likely to have a second visit in one month. For the retention variable, the only client covariate that was consistently statistically significant across models for the CCAS criteria was age, with the likelihood of a successful outcome increasing with advancing age. No site descriptor covariates were consistently statistically significant across models for either service outcome variables.

For the engagement variable, odds ratios for Hispanics with mood disorders significantly decreased with increasing CCAS scores for the criteria ‘having interpreters’ and ‘assessing/adapting/having new services’. Odds ratios for Blacks with mood disorders decreased for ‘having interpreters’ and increased for ‘having hiring and retention policies’. Odds ratios for persons with psychotic and other diagnoses were not impacted by increases in CC scores.

For the retention variable, odds ratios decreased for increasing CCAS scores for Hispanics with mood disorders for the criteria ‘having interpreters’, ‘having translated key forms’, ‘having translated/formatted educational/service materials’, and for ‘assessing/adapting/having new services’; for Hispanics with anxiety disorders for ‘having bilingual staff’; and for Hispanics with other diagnoses for the total score, ‘recruiting/hiring/retention’ ‘having translated key forms’, ‘having translated/formatted educational/service materials’, and for ‘assessing/adapting/having new services’. Odds ratios for Blacks were unaffected by increases in CC scores except for the criterion ‘having interpreters’ for those with a mood disorder. Odds ratios for persons with psychotic disorders were not impacted by increases in CC scores.

Discussion

The CCAS criteria and the activities they define serve as a roadmap for the introduction of CC activities into an agency, as well as a marker of progress. The range of scores observed in the study is indicative of the differential progress of the participating agencies. Factor analysis suggests that there are three stages in the process of becoming increasingly culturally competent. In the first stage, corresponding to the first factor, the emphasis is on administrative elements (agency has a commitment, receives community input from a CC committee and conducts staff training). In the second stage, the agency engages in activities designed to better understand and serve their communities (collects data, institutes recruiting/hiring/retention policies and creates translated and easy to read service descriptions and educational materials). The third stage is comprised of activities that are directly associated with clinical care (has interpreters, bilingual, bicultural staff and reviews, adapts and institutes new services). Notably, none of the administrative items, including conducting staff training, contributed to the prediction of reduced service outcomes disparities. In contrast, all of the clinically related items had an impact. Administrative activities, such as showing commitment by having a mission statement, are essential elements of the CC process, but were insufficient in and of themselves to reduce odds ratios. However, it seems self evident that these activities must be in place before activities more closely related to direct care, such as adapting a service to a cultural group, can commence. Surprisingly, training activities were not predictive of disparity reduction. This may be a consequence of the often heard criticisms that existing training curricula are of mixed quality, overly broad and too distant from the clinical process.

Construct validity was demonstrated for several CCAS criteria. Low correlation of the CCAS Total Score and the CCSI total score was expected as some features of CC contained in the CCSI checklist are not reflected in the CCAS. If there is future modification of the scale, some of these items might be usefully introduced as discussed below. The inter-rater reliability of the CCAS is satisfactory, but is substantially improved if two or more raters perform the evaluation. Field experience in the use of the scale suggests that a rater needs considerable interpersonal skills and a determined attitude to obtain the necessary information to complete the CCAS. Multiple raters with varying agency connections and sufficient levels of assertiveness will likely increase the chance of obtaining accurate and comprehensive coverage of the scale items.

Although Hispanics and Whites did not statistically differ in their average outcomes for both service measures, Blacks and Whites did. At the same time, CCAS items demonstrated predictive validity for Hispanics versus Whites, but not for Blacks versus Whites. Though these findings may appear to be counter-intuitive, they are not contradictory. The fact that there are no differences between Hispanics and Whites over all facilities may indicate that in this sample CC is working effectively to eliminate Hispanic disparities. But this does not preclude higher CCAS scores predicting smaller odds ratios. The former is the difference between the average outcome (of a univariate measure) of the Whites and Hispanics, while the latter quantifies the (bivariate) relationship between the odds ratio and the CCAS score across the 27 facilities in terms of a slope in the logistic regression.

The meaning of a health care disparity has been widely discussed (e.g. IOM 2002; McGuire et al. 2006; Cook et al. 2009; Duan et al. 2008). In these definitions, differences or divergences in service outcome rates are viewed as disparities when they are estimated within stratum in which the groups are comparable with respect to conditions, other than the service itself, that can impact outcome, e.g., health status and access. We included covariates for diagnosis, age and gender in the logistic hierarchical models as proxies for health status. It would have been desirable also to include direct measures of socioeconomic status as they

are contributors to access, but there were no relevant variables for which data were systematically available. However, the clinics in the study are part of a widely available public mental health system of the county, so it is likely that the socioeconomic status of clients is fairly homogeneous and that access is not an issue. We obtained odds ratios separately for comparable diagnostic groups and consider these as measures of disparity.

The predictive validity finding that CC scores on criteria related to linguistic capacity of the agency resulted in disparity reduction for Hispanics but not for Blacks is not unexpected. However, linguistic accommodations of a different sort than are required by Hispanics may be needed to enhance the care experience for Blacks. Staff familiar with the language and phrases of the daily lives of Blacks, even to the extent that the vernacular is used in the clinical setting when appropriate, might improve the clinical experience. It is disconcerting, however, that the service related criterion predicted disparity reductions for Hispanics but did not for Blacks. New or adapted services for Blacks may not have been in place in the studied agencies. Identification of the cultural elements that more effectively engage and retain Black clients is currently the subject of considerable research and development. The reversal of odds ratios for Blacks for the engagement outcome with increasing CCAS scores on hiring and retention policies might be spurious or speculatively attributable to hiring practices that increased the availability of Hispanic but not Black staff.

There may be organizational features of an agency that are not assessed in the CCAS that may be of greater relevance to Blacks than the ones now included. It might be valuable to measure the degree to which services are flexibly delivered, appointment times are available and that transportation, child care, and food are provided as needed. On the other hand, there may be no organizational level items that are able to predict disparity reduction for Blacks because the therapeutic alliance between clients and providers is most critical and dominates all other considerations (Whaley 2001; Cruz and Pincus 2002).

For the subset of the patients who had psychotic disorders, no relationship between CCAS scores and disparity reduction was found. This is consistent with the expectation that for this diagnostic group the two outcome measures, engagement and retention, are likely to be positive by nature of the illness. This would tend to substantially dominate CC effects, if any are present. These findings may be viewed as supporting the instrument's predictive validity.

In interpreting the results of this study, several caveats must be kept in mind. Because the data are naturalistic, diagnoses are not research-based and therefore may have been recorded unsystematically. Indeed, the diagnoses may have been biased as evidenced by the rate of psychotic disorders among Blacks which was twice the rate among Whites and Hispanics. Similar biases have recently been reported (Hampton 2007). In addition, limitations of the information in the database did not allow refined specification of cultural groups. Lumping clearly distinct cultural groups (e.g., African-American and Afro-Caribbean) into one category (Blacks) could have masked an effect or produced a misleading result because the groups may have heterogeneous service use distributions. Additionally, the outcome measures were based exclusively on patterns of service use rather than clinical or functional outcomes because these were not reported. Finally, the reported results focus on parameters in the models that were statistically different. Each result of a hypothesis test stands as a valid statistical statement, with a statement specific 5% probability of erroneously rejecting the null. However, in aggregate many tests were performed and therefore the chance that at least one of them was falsely positive is greater than 5%.

There is still much to be learned about organizational level practices that successfully promote the delivery of culturally competent mental health services. Future research might

profitably focus on identifying scale components that successfully predict disparity reductions for Blacks and for specific diagnostic groups. These in turn would revise and improve the roadmap that describes the path to CC. Study of diverse organizational types, such as child and adolescent clinics, may enhance understanding of differential effects of CC across outpatient settings. Finally, longitudinal studies of culturally competent administrative processes that facilitate changes in the delivery of clinical services with direct positive effects on treatment outcomes are highly desirable. These may guide the process and promote organizational change that would reduce cultural disparities in mental health care.

Acknowledgments

This work was supported by NIMH Grant R34 MH071489-01A1 and the New York State Office of Mental Health Nathan S. Kline Institute Center of Excellence in Culturally Competent Mental Health.

References

- Bond GR, Evans L, Salyers MP, Williams J, Kim HW. Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*. 2000; 2(2):75–87. [PubMed: 11256719]
- Campinha-Bacote, J. Inventory for assessing the process of cultural competence (IAPCC) among health care professionals. Cincinnati, OH: Transcultural C.A.R.E. Associates; 1998.
- Cook BL, McGuire TG, Meara E, Zaslavsky AM. Adjusting for health status in non-linear models of health care disparities. *Health Services Outcomes Research Method*. 2009; 9:1–21.
- Cross, T.; Bazron, B.; Dennis, K.; Isaacs, M. Towards a culturally competent system of care. Vol. I. Washington, DC: Georgetown University Child Development Center, CASSP Technical Assistance Center; 1989.
- Cruz M, Pincus HA. Research on the influence that communication in psychiatric encounters has on treatment. *Psychiatric Services*. 2002; 53(10):1253–1265. [PubMed: 12364673]
- DHHS. President's new freedom commission on mental health. Achieving the promise: Transforming mental health care in America (DHHS Publication No. SMA-03-3832). MD: Rockville; 2003.
- Doorenbos AZ, Schim SM, Benkert R, Borse NN. Psychometric evaluation of the cultural competency assessment instrument among healthcare providers. *Nursing Research*. 2005; 54(5):324–331. [PubMed: 16224318]
- Duan N, Meng XL, Lin JY, Chen CN, Alegria M. Disparities in defining disparities: Statistical conceptual frameworks. *Statistics in Medicine*. 2008; 27:3941–3956. [PubMed: 18626925]
- Goode, TD.; Tawara, D. Self-assessment checklist for personnel providing services and supports to children with special health needs and their families. Washington, DC: Georgetown University Center for Child and Human Development; 2002.
- Hampton MD. The role of treatment setting and high acuity in the over diagnosis of schizophrenia in African Americans. *Archives of Psychiatric Nursing*. 2007; 21(6):327–335. [PubMed: 18037443]
- Harper, M.; Hernandez, M.; Nesman, T.; Mowery, D.; Worthington, J.; Isaacs, M. Organizational cultural competence: A review of assessment protocols (Making children's mental health services successful series, FMHI pub. No 240-2). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health; 2006.
- Institute Of Medicine (IOM). Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academics Press; 2002.
- Isaacs, MR.; Huang, LN.; Hernandez, M.; Echo-Hawk, H. The road to evidence: The intersection of evidence-based practices and cultural competence in children's mental health. Alliance of Multi-ethnic Behavioral Health Associations; 2005.
<http://systemsofcare.samhsa.gov/resourceguide/docs/Road%20to%20Evidence.pdf>
- Kraemer HC. The small sample non-null properties of Kendalls's coefficient of concordance for normal populations. *Journal of the American Statistical Association*. 1976; 71(335):608–613.

- Lucas T, Michalopoulou G, Falzarano P, Menon S, Cunningham W. Healthcare provider cultural competency: Development and initial validation of a patient report measure. *Health Psychology*. 2008; 27(2):185–193. [PubMed: 18377137]
- Mason, JL. Cultural competence self-assessment questionnaire: A manual for users, research and training center on family support and children's mental health. Portland, OR: Portland State University; 1995.
- Mcguire TG, Alegria M, Cook BL, Wells KB, Zaslavsky AM. Implementing the Institute of Medicine definition of disparity: An application to mental health care. *Health Services Research*. 2006; 41(5):1979–2005. [PubMed: 16987312]
- New York State Office of Mental Health. Cultural competence performance measures for managed behavioral healthcare programs. NY: Albany; 1998.
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. Vol. 3. New York NY: McGraw-Hill; 1994. p. 232
- Roper W, Mays G. Performance measurement in public health: Conceptual and methodological issues in building the scientific base. *Journal of Public Health Management and Practice*. 2000; 6(5):66–77. [PubMed: 11067663]
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86(2):420–428. [PubMed: 18839484]
- Siegel C, Chambers ED, Haugland G, Bank R, Aponte C, McCombs H. Performance measures of cultural competency in mental health organizations. *Administration and Policy in Mental Health*. 2000; 28(2):91–106. [PubMed: 11194126]
- Siegel C, Haugland G, Chambers ED. Performance measures and their benchmarks for assessing organizational cultural competency in behavioral health care service delivery. *Administration and Policy in Mental Health*. 2003; 31(2):141–170. [PubMed: 14756197]
- Siegel, C.; Haugland, G.; Schore, R. The interface of cultural competency and evidence based practices. In: Drake, RE.; Merrens, MR.; Lynde, DW., editors. Chap. 12 evidence-based mental health practice: A textbook. New York: WW Norton; 2005.
- Siegel, C.; Laska, E.; Wanderling, J.; Baker, S.; Harrison, G.; Rank, R., et al. Study methodology. In: Hopper, K.; Harrison, G.; Aleksander, J.; Sartorius, N., editors. *Recovery from Schizophrenia—An international perspective*. Vol. Chap 2. New York, NY: Oxford University Press Inc; 2007.
- Sue DW, Arredondo P, McDavis RJ. Multicultural counseling competencies and standards: A call to the profession. *Journal of Counseling and Development*. 1992; 70(4):477–483.
- USPHS. Mental Health: Culture, Race and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General. Rockville, MD: Department of Health and Human Services, US Public Health Service; 2001. United States Public Health Service, Office of the Surgeon General.
- Whaley AL. Cultural mistrust of White mental health clinicians among African Americans with severe mental illness. *American Journal of Orthopsychiatry*. 2001; 7(2):252–256. [PubMed: 11347366]

Fig. 1.
Frequency distribution of total CCAS scores

Fig. 2.
Proportion of group having 2nd visit in one month

Fig. 3.
Proportion of group having four services in six months

Table 1

CCAS criteria and their rationale

Criterion	Description	Rationale
1. Agency's commitment to cultural competency (CC) (agency commitment)	The agency (or its parent organization (PO)) has a management level person responsible for CC and a dedicated budget for CC activities, a CC plan and procedures for updating the CC plan.	A management level person who has primary responsibility for CC ensures that CC will be addressed. Without a dedicated budget, only limited activities can be conducted. A written plan concretizes the agency's commitment to CC. Review and updating ensures that there is a feedback loop and that corrective actions can be taken. If an agency is part of a larger organization, the PO, these activities can take place there.
2. Data to assess service needs (data)	The agency obtains current data on its service users and its target community that enable identification of their cultures and language needs	Particular data items need to be collected for all clients in a consistent manner so that the cultures and language needs of the population being served by the agency can be determined. Information on the target community allows the agency to tailor its outreach and services to the needs of community cultural groups.
3. Agency CC group with community input addresses cultural group needs (having a CC committee)	The agency (or its PO) has a CC Committee or other group that addresses cultural issues and has participation from cultural groups of the target community	A committee dedicated to CC will enhance the likelihood that activities appropriate to the cultures served are introduced and carried out. The committee can be stand alone or part of existing boards, advisory groups and committees. Input is required from representatives of the most prevalent cultural group of the target community. More than one person from a cultural group is desirable to ensure active participation.
4. Integration of CC group in agency (integration of CC committee in agency)	<p>The CC group is integrated within agency evidenced by functions:</p> <ul style="list-style-type: none"> • Reviews services/programs with respect to CC • Reports to Quality Assurance/Quality Improvement program of the agency/PO • Participates in planning and implementation of services • Directly transmits recommendations to executive level of agency/PO 	The extent to which the functions of the CC Committee are reported and used in the agency provides a measure of the likelihood of change with respect to CC. Four key committee functions are expected to take place, but they may be introduced at different stages.
5. CC training of staff (conducting staff training)	The agency (or PO) offers to staff educational activities in which cultural issues are addressed and requires staff to have an adequate amount of specific training on CC	Training and educating staff in CC enhances the likelihood of the delivery of appropriate services to cultural groups. Ideally, educational activities should take place every year and that all staff members who have face-to-face contact with and provide direct clinical care to agency clients receive CC training. Administrative staff should also be knowledgeable about CC.
6. Recruitment, hiring and retention of staff from/or experienced with the most prevalent cultural groups of service users (recruitment, hiring and retention)	The agency is committed to hiring and retaining CC staff who are from or who have had experience working with the most prevalent cultural groups of its service users	Having direct service, supervisory and administrative staff with relevant experience with the most prevalent cultural groups enhances the likelihood of incorporating CC practices. Hiring and retaining professional staff members who are from the cultures of service users provides positive role models for clients of the agency and affords additional opportunities to increase knowledge about the cultures.
7. Language capacity: interpreters (having interpreters)	The agency (or PO) accommodates persons who have limited English proficiency (LEP) by using interpreter services or bilingual staff	Language needs of persons with LEP must come to the attention of the agency at point of first contact in order to identify their needs and link them to the appropriate services. In service, language needs must continue to be addressed. Ideally interpreters are formally trained and certified or are bilingual staff members who have received CC training.
8. Language capacity: Bilingual staff (having bilingual staff)	The agency has staff who speak the language of the most prevalent cultural group of service users with members who have LEP	Having bilingual staff members who can work directly with persons with language needs is ideal – and likely to be cost effective as well. Many of these staff members may also be bicultural so that cultural groups special styles and needs are better understood.

Criterion	Description	Rationale
9. Language capacity: key forms (having translated key forms)	The agency has translated versions of key documents and forms in the language of the most prevalent cultural groups of its service users	Legal documents and critical care information must be understandable to persons with LEP to ensure that such information is provided equally to all service users.
10. Language capacity: service descriptions and educational materials (having translated/formatted service and educational materials)	The agency accommodates persons with LEP by translating service descriptions. It accommodates persons with limited reading skills by providing service descriptions or educational materials in formats they can understand.	Using different formats and media for descriptions of available services and for educational materials ensures that literacy level is not a barrier to accessing services. Service-related materials translated into the most prevalent non-English language will be especially helpful to overcome treatment barriers for service users and family members.
11. Assessment, adaptation and introduction of new services for cultural groups (assessing, adapting, and having new services)	The agency (or PO) reviews and adapts or introduces services suitable to the most prevalent cultural groups of service users	A comprehensive CC-based assessment of services may result in the adaptation of existing services and the introduction of new services to better serve members of the most prevalent cultural groups.

Table 2

Sample description

	All	Black	Hispanic	Other	White
<i>N</i>	10193	2876	1061	286	5970
Percent					
All	100	28	10	3	59
Diagnostic group					
Psychotic	7	11	5	8	6
Mood	44	39	50	31	45
Anxiety	10	6	10	10	11
Other	40	44	35	51	38
Gender					
Female	55	55	54	51	56
Setting					
Clinic adult	73	72	80	61	73
Clinic children and youth	22	23	17	37	22
Day treatment	3	4	2	<1	3
Psychiatric rehabilitation	1	1	1	2	1
Mean age (SD)					
Clinic adult	39.5 (15.3)	37.1 (11.9)	36.9 (12.5)	35.9 (13.2)	41.1 (17.0)
Clinic children and youth	11.3 (3.9)	11.2 (3.8)	10.8 (4.0)	9.8 (4.3)	11.6 (3.9)
Day treatment	39.8 (12.8)	37.6 (10.8)	37.3 (9.0)	57 (0)	41.2 (14.0)
Psychiatric rehabilitation	40.4 (10.6)	41.2 (9.6)	33.3 (12.4)	43 (10.1)	41.1 (10.5)

Table 3CCAS and CCSI checklist scores means (SD) and Pearson correlation of scores ($N = 27$ sites)

CCAS criteria	CCAS score mean (SD)	CCSI checklist score mean (SD)	Pearson correlation
Average total	3.30 (0.06)	0.80 (0.10)	0.37*
1. Agency commitment	3.90 (1.6)	0.83 (0.17)	0.47*
2. Having data	4.56 (0.6)	0.97 (0.08)	0.35
3. Having a CC committee	3.96 (1.3)	0.85 (0.23)	0.57*
4. Integration of CC committee in agency	4.06 (1.0)	0.73 (0.25)	0.16
5. Conducting staff training	3.50 (1.1)	0.52 (0.34)	0.58*
6. Recruitment, hiring, retention	3.20 (0.8)	0.84 (0.34)	-0.07
7. Having interpreters	3.44 (1.1)	0.83 (0.27)	-0.10
8. Having bilingual staff	2.94 (1.1)	0.41 (0.50)	0.44*
9. Having translated key forms	2.78 (1.1)	0.70 (0.34)	0.48*
10. Having translated/formatted service and educational materials	2.09 (1.1)	0.87 (0.26)	0.23
11. Assessing, adapting and having new services	2.15 (1.2)	0.66 (0.30)	0.30

* Pearson correlation significantly different from 0 ($P < 0.05$)

CCAS scores range from 1 to 5; CCSI scores range from 0 to 1. Higher scores indicate greater CC