



Published in final edited form as:

*Drug Alcohol Depend.* 2012 June 1; 123(1-3): 160–166. doi:10.1016/j.drugalcdep.2011.11.003.

## Measurement of Gender-Sensitive Treatment for Women in Mixed-Gender Substance Abuse Treatment Programs<sup>†</sup>

Zhiqun Tang<sup>\*</sup>, Ronald E. Claus, Robert G. Orwin, Wendy B. Kissin, and Carlos Arieira Westat, 1600 Research Boulevard, Rockville, Maryland 20850, USA

### Abstract

**Background**—Gender-sensitive (GS) substance abuse treatment services have emerged in response to the multidimensional profile of problems that women display upon admission to substance abuse treatment. The present study examines the extent to which treatment programs vary in GS programming for women in real-world mixed-gender treatment settings, where most women are treated.

**Methods**—Data were collected through site visits using semi-structured interviews with program directors, clinical directors, and counselors in 13 mixed-gender treatment programs from Washington State. Rasch modeling techniques were used to analyze the data.

**Results**—Naturally occurring variation was revealed within and across the treatment programs, and demonstrated that reliable measures of three GS domains (Grella, 2008) can be constructed despite a small number of programs.

**Conclusions**—This is the first study to quantify GS treatment for substance abusing women. The identified treatment services and practices and the way they clustered together to form scales have practical implications for researchers, service providers, clinicians, and policy makers. The scales can be used to study treatment outcomes and to evaluate the effectiveness, cost-effectiveness, and cost-benefit of GS programming for women.

### Keywords

Gender Sensitivity; Substance abuse treatment; Mixed-Gender; Short Term Residential (STR); Rasch modeling

---

<sup>†</sup>Supplementary material can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

© 2011 Elsevier Ireland Ltd. All rights reserved

<sup>\*</sup>Corresponding author, Westat, 1600 Research Boulevard, Rockville, MD 20850, United States. Tel.: +1 240 456 2656; fax: +1 301 610 5140. [zhiquntang@westat.com](mailto:zhiquntang@westat.com).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Contributors** Dr. Robert G. Orwin is the principal investigator and designed the study. Dr. Wendy Kissin is the project manager and led the protocol development. Dr. Ronald E. Claus was the site visit leader, contributed to data collection, reviewed selected analytic results, and wrote the second draft of the manuscript. Dr. Carlos Arieira was the data manager and conducted quality checks of the data. Dr. Zhiqun Tang designed the analysis plan, performed the statistical analyses, reviewed all analytic results, and wrote the first draft of the manuscript. All authors reviewed the instruments and related materials, reviewed and agreed on the analytical procedures and criteria used, and contributed to and have approved the final manuscript.

**Conflict of Interest** All the authors declare that they have no conflicts of interest.

## 1. Introduction

A substantial body of research has demonstrated gender differences in the treatment needs of substance abusers (Brady and Randall, 1999; Pelissier and Jones, 2005) due to different patterns of alcohol and drug use (Greenfield et al., 2010; Grella and Lovinger, 2011; Weiss et al., 2003), progression to dependence (O'Brien and Anthony, 2005; Zilberman et al., 2003), treatment access and recovery processes (Green, 2006; Green et al., 2002; Greenfield et al., 2007; Grella et al., 2008; Westermeyer and Boedicker, 2000). Recognition of the different treatment needs of women has led to the increased provision of gender-sensitive (GS) treatment services for women since the 1990s (Greenfield and Grella, 2009; Grella, 2008; Grella and Greenwell, 2004) and to advocacy for women-only treatment (Hodgins et al., 1997). (While the term “gender-sensitive” is often used generically (e.g., by the United Nations to address women's basic or special needs) other terms commonly used include women- or gender-responsive, women-focused, and gender-specific; the distinctions are not always clear (Grella, 2008)). Funding constraints currently limit the number of women-only treatment programs, and most women in the United States continue to receive care in mixed-gender settings (Grella and Greenwell, 2004). Substance abusing women may be treated as effectively in mixed-gender programs as in women-only programs (Kaskutas et al., 2005) when treatment is designed to address women's specific needs (Greenfield et al., 2007). Anecdotal evidence and the lack of a gold standard for GS treatment suggest that mixed-gender programs vary widely in the extent to which they provide GS services, but the study of gender sensitivity is challenged by the lack of adequate and reliable measures (Greenfield and Grella, 2009; Grella, 2008).

As research on women's treatment has matured, GS treatment has been described as a set of comprehensive, family-focused interventions provided in a strengths-based, relational, trauma-informed fashion within a safe and affirming environment (Bloom, 1999; Bloom et al., 2003; Covington, 2008; Covington and Surrey, 1997; Grella, 2008). Recently, specific guidelines for the provision of GS treatment for substance abusing women were summarized (Women's Services Practice Improvement Collaborative, 2007) in response to the need to clarify what constitutes GS treatment. Drawing upon the past three decades of research, Grella (2008) proposed a multidimensional model for measuring the gender sensitivity construct. However, little empirical research has demonstrated the degree to which GS programming is embedded in mixed-gender settings, how gender sensitivity is related to treatment outcomes, or whether the value added by GS treatment offsets the costs.

Despite the burgeoning literature on the importance of GS treatment and what constitutes GS treatment, measurement of the construct and its essential components is in its infancy. Research has focused more on straightforward program contrasts (i.e., mixed-gender compared to women-only programs; Niv and Hser, 2007; Prendergast et al., 2011) and research examining the organizational components underpinning the delivery of GS treatment services is scarce (Hser and Niv, 2006). Semi-qualitative inventories for women-only and mixed-gender programs have been developed (Claus et al., 2007; Covington and Bloom, 2008), yet a clear method for translating these tools into quantitative measures that can be used to place programs along a continuum of gender sensitivity has not been established. GS treatment models treat all treatment services as equally important, equally likely to be implemented, and of the same value to an organization. Indeed, “the field currently lacks adequate measurements of the therapeutic and programmatic components of women-focused treatment, which are essential to empirical validation of its effectiveness and dissemination of effective treatment approaches” (Greenfield and Grella, 2009, p.881). The methodological challenges of quantifying this construct are amplified by the small sample sizes common to organizational research.

This study uses a Rasch modeling approach to develop measures of GS treatment. The aim of the Rasch model (Rasch, 1960) is to examine whether observed data can be fit to create scales with strong measurement properties. Rasch analysis independently orders both items (here, GS practices) and respondents (here, treatment programs) along the same continuous, interval-scaled latent trait (here, program's self-reported gender-sensitivity). The Rasch model assumes the probability of endorsing an item is a function of the distance between a program's and an item's value or location on the latent variable. That is, when a program's gender sensitivity level on this continuum is equal to the "difficulty" of a particular item, the probability of endorsing the item is 50%. GS practices that are used more widely receive lower scores, while less used practices receive higher scores. Likewise, programs using more GS practices receive higher scores and those using fewer receive lower scores along an underlying GS dimension. Since item and respondent difficulties are independently estimated, Rasch-derived measures are considered sample-independent (Embretson and Reise, 2000) and are more likely to generalize to other similar populations, which has advantages over methods based on classical test theory (CTT) that is sample dependent and requires large samples (e.g., factor analysis) to effectively separate measurement error from true score (Kline, 2005). The Rasch approach permits small sample sizes (Linacre, 1994) and offers multiple fit indices and calculates item information functions for model, item, and item response category diagnosis, allows for the identification of item redundancies, and can readily be used to show differences in the relative ordering of items. These properties enable the comprehensive examination and distillation of a large number of items drawn from semi-structured interview data into a meaningful, psychometrically sound scale.

This article focuses on the assessment of naturally occurring variation in GS programming for substance abusing women in mixed-gender short-term residential treatment programs. Specifically, the objectives are to use organization-level empirical data to quantify the gender sensitivity of programs based on Grella's (2008) multidimensional model and to present a methodology for the measurement and evaluation of GS treatment and associated cost benefits.

## 2. Methods

### 2.1 Sample

Thirteen eligible mixed-gender short-term residential (STR) substance abuse treatment programs in Washington State (WA) received site visits from August 2008 to March 2009. Eligibility criteria included (1) serving publicly funded men and women over age 18, (2) offering 30-day residential treatment services, and (3) having a patient gender ratio no smaller than 1:11 in either direction, thus affording appropriate variability in a mixed-gender environment. All of the eligible programs (called intensive inpatient programs in WA) agreed to participate.

### 2.2 Instrumentation and procedure

GS measurement instruments were developed based on protocols used in a prior study, which focused on long-term residential treatment for women in women-only and mixed-gender settings (Claus et al., 2007) and were loosely based on protocols used to interview staff at long-term residential programs serving pregnant women and their children (Chen et al., 2004; Dowell et al., 2003; Greenfield et al., 2004). Draft protocols were created by two doctoral-level clinicians. To incorporate developments since the previous study and adapt the protocols to STR programs, new items were added from the Gender-Responsive Program Assessment (Covington and Bloom, 2008). After review by external experts and subsequent revision, the protocols and other relevant interview materials were pilot-tested at one urban and one rural mixed-gender STR program in a different state.

The interview protocols include open- and closed-ended questions on treatment orientation and philosophy, staffing and training, services provided, program challenges, and program environmental factors. Specifically, the *Program Director Interview* is designed to collect information on program structure and philosophy, admission patterns, children's services, staff competencies and training, program challenges, and program costs. The *Clinical Director Interview* asks about treatment philosophy and caseloads, population served, assessment and treatment engagement, treatment planning, services provided to patients, children and other family members, discharge planning, post-treatment housing services, continuing care services, and costs. The *Counselor Interview* asks about patient counseling, patient access to children, continuing care services, perceived patient satisfaction, barriers to treatment, and general environmental features. An *Observational Protocol* was used by the interviewers to record observations about cleanliness, safety, security, and privacy of the program's physical environment. Interview responses were binary (e.g., program director gender), 5-point ordinal (e.g., the extent that training addressed women's mental health issues), or count variables (e.g., staff with addiction certification). For analytic purposes, some polytomous variables were generated from count or continuous variables (e.g., annual training hours related to women's recovery was recoded as a 3-level variable).<sup>1</sup> (Instruments are available upon request.)

The program director, the clinical director, and a counselor at each program were interviewed as part of a 1.5 day site visit by two researchers. At three agencies, an additional staff person was interviewed to obtain information that other respondents could not provide, resulting in 42 staff member interviews. Immediately following each site visit, the researchers completed the observational protocol. All study procedures were approved by the appropriate ethical review boards.

## 2.3 Analytical procedures

**2.3.1 Gender sensitivity measurement domains**—The strategy of using both substantively-driven and empirically-based methods was adopted considering the small program-level sample size. Items from the three interview protocols and observational protocol were grouped into seven measurement domains based on Grella's (2008) multi-dimensional taxonomy: (1) treatment orientation/processes, (2) administration and staff, (3) organization characteristics, (4) women's services, (5) general services, (6) children's services, and (7) physical environment. Two doctoral-level clinicians independently classified each item into a GS domain and met to resolve minor assignment differences. To ensure broad content coverage and increased scale stability, domains with item pools of 20 or more were selected for separate Rasch analysis (see Table 1). Subsequently, we combined domains based on conceptual overlap and the empirical associations observed in preliminary analyses (e.g., Cronbach's alpha and item-total correlations; details available upon request). Specifically, Treatment Orientation and Administration and Staff items were analyzed together, and Women's Services and General Services items were analyzed together.

**2.3.2 Rasch modeling**—Rasch modeling was chosen as the primary statistical analysis method to create GS scales. Items were screened for their empirical utility; those without variability were removed from further analyses. Rasch analyses were performed using Winsteps 3.70.0.2 (Linacre, 2009). Within each domain, variables were grouped by the number of response categories and were specified as a grouped-item rating scale model with one measurement structure per item-group (Linacre, 2010).

---

<sup>1</sup>Descriptive statistics of the retained items can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

Model and item fit statistics show how the empirical data fit a theoretically determined model (Linacre, 2010; Wright et al., 1994). The overall goal for a set of Rasch analyses was to achieve reliability at 0.80 (for both item and programs) mainly by identifying and removing misfit items. The analyses involved six general stages and the judgment of model fit was based on the set of model- and item-level fit statistics described below. First, items that had a negative item-measure correlation were removed from further analyses. Second, items were excluded sequentially based on item fit statistics. Items diagnosed as misfitting were removed successively until no more misfitting items presented. Item outfit and infit are critical chi-square fit statistics that indicate whether an item is productive for Rasch measurement: they have an expected value of 1, where large values indicate underfit, in which noise and unmodeled variance may degrade measurement, and small values suggest overfit, in which there is too little random noise and reliability may be inflated. Underfit is often more of a concern than overfit and thus was prioritized in our study. In general, the selection of cut-off values for infit and outfit statistics is related to both the sample size and the intended use of a measure. In the present study, we retained items with mean square (MNSQ) infit and outfit statistics values less than 2 (Linacre, 2003; Wright et al., 1994). Given the small sample size, standardized chi-square fit statistics were also examined (Linacre, 2003). Third, items with acceptable fit but a low item-total correlation ( $r < 0.40$ ) were removed. Fourth, items that exhibited disordered empirical average measures in the response categories were investigated. Items with non-adjacent disordered categories were treated as invalid and removed (Linacre, 2010). Fifth, unidimensionality was examined with principal components analysis (PCA) of residual variance after the final model was fit (Linacre, 1998). To verify that unexplained variance showed the scale to be unidimensional, simulations were conducted to compare residuals of equivalent Rasch-fitting data with the empirical data (Linacre, 2010). Finally, we checked bivariate residual correlations to ensure that any remaining pair of items did not have a correlation above .70, which indicates potential local dependency of items (Linacre, 2010). If so, one item was removed and the model was rerun to verify fit.

For each final model, we report reliability and separation statistics. Separation is defined as the ratio of the standard deviation of the sample, adjusted for inflation due to error (i.e., the true variance), to the standard error of measurement (i.e., the root mean square error). Although similar to coefficient alpha, separation can more precisely determine whether a shorter test has maintained adequate measurement properties since alpha suffers from a ceiling effect (Mallinson et al., 2004). The separation statistic is used to calculate  $G$ , or the number of strata into which programs can be reliably separated (Linacre, 2010). We reviewed the distributions, means and ranges of item and program scores; ideally, mean item and program scores are closely located on the latent variable. When a set of items is not well-matched to the programs to which they are applied (i.e., many items are too easy or too challenging to achieve), reliability is affected. The hierarchical structures of items in the final models were reviewed for meaning and content validity.

### 3. Results

#### 3.1 Treatment orientation and staff training

Beginning with 48 variables from the Treatment Orientation/Processes and Administrator and Staff domains, we identified a good-fitting model containing 16 variables that met all analytic and fit criteria described in the previous section. Item MNSQ fit statistics and standardized fit statistics ( $Z$ ), with the GS characteristics ordered from the most to least challenging to achieve, are presented in Table 2. Item-scale correlations were also provided. PCA showed that the items and programs explained 72.4% of the total raw variance, and the first factor extracted from the residuals was small (6.7%). Additional simulations of residuals did not provide evidence to disconfirm the unidimensionality of this scale. The

measure's reliability was very good for both items (0.86) and programs (0.92), and the program separation statistic (3.34) showed that 4 groups of programs could be reliably discerned ( $G = 4.79$ ). The item and program measures displayed wide ranges and showed a substantial overlap<sup>2</sup>, indicating broadly varying program standards that were appropriately targeted by the GS items. This set of item- and model-level statistics suggested that the model fit the data well and vice versa. Consistent with Grella's (2008) taxonomy, the Treatment Orientation and Staff Training scale includes items that address a program's orientation toward behavioral health and the provision of behavioral health services as well as staff preparation and training in areas specific to treatment for women.

### 3.2 Women's and General Services

Beginning with 65 variables from the Women's Services and General Services domains, we identified a good-fitting model containing 25 variables (see Table 3). Adjacent responses were combined to eliminate disordered categories in 16 variables. PCA conducted on the final model explained 67.5% of total raw variance, and the first factor extracted from residuals was small (7.6%). Additional simulations of residuals did not provide evidence to disconfirm the unidimensionality of this scale. The measure's program reliability was excellent (0.93) and item reliability was adequate (0.70). Program separation (3.78) showed that five groups of programs could be reliably discerned ( $G = 5.37$ ). The items for this scale covered a moderate but adequate range of GS ability (5.1 logits), whereas the range of program scores was substantially larger (10.05 logits, from  $-4.83$  to  $5.22$ ). Given the close location of item and program means, this suggests the need to develop both more challenging and less challenging items in this domain. Content review showed that this scale focused on treatment assessment and services, including items which address issues more specific to women (e.g., women's health information and domestic violence) and items which are more general (e.g., stage of change assessment), consistent with the delivery of women's and general treatment services.

### 3.3. Physical environment

Beginning with 25 items from the Physical Environment domain, we identified a good-fitting model with 16 items (see Table 4). This scale drew upon items from staff interviews and interviewer observations, which were grouped separately in analyses. PCA on the final model explained 58.5% of total raw variance, and the first factor extracted from residuals was acceptable (10.5%). Additional simulations of residuals did not provide evidence to disconfirm the unidimensionality of this scale. The scale showed good item reliability (0.80) and excellent program reliability (0.92). The program separation statistic (3.46) indicated that 4 groups of programs could be reliability discriminated ( $G = 4.95$ ). Item and program mean scores were adjacent (0.42 logits apart) with moderate ranges (5.45 and 6.02 logits) that were largely coincidental. Consistent with Grella's (2008) taxonomy, the Physical Environment scale contained items related to safety, privacy, and livability features in the treatment program and its surrounding area.

## 4. Discussion

This is the first study to systematically identify, measure, and quantify the characteristics of GS treatment for substance abusing women in mixed-gender programs. The study used Rasch modeling techniques to comprehensively assess scale properties at the scale, item and category levels, beyond the examination of item-scale correlations. Beginning with a rich set of items, reliable and valid scales were created to measure Treatment Orientation and Staff

---

<sup>2</sup>Program-item maps, with distributions of both item and program measures can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

Training; Women's and General Services; and Physical Environment. The scales map onto a multidimensional taxonomy that catalogues GS treatment (Grella, 2008). The findings are likely to generalize to other residential treatment programs, given the Rasch modeling approach used here.

#### 4.1 GS variation

The objective of this study was to assess and quantify the delivery of GS programming for women admitted to mixed-gender substance abuse treatment programs. Our findings identified naturally occurring variation in GS programming across the participating programs. The range of each scale establishes wide differences in the delivery of GS treatment. This variation is manifested in different ways. Consistent with Grella's multidimensional classification, more than one distinct scale emerged to measure gender sensitivity. Statistical power to demonstrate that the GS scales measure independent constructs is low; given the available sample size, correlations must reach .55 to be statistically different than zero ( $p < .05$ ). The observed associations between the scales were low to moderate in size:  $r = .02$  for Treatment Orientation and Staff Training and Women's and General Services;  $r = -.25$  for Treatment Orientation and Staff Training and Physical Environment;  $r = .50$  for Women's and General Services and Physical Environment. These findings suggest that the provision of GS services may be related to a program's physical environment. However, not all of Grella's (2008) domains emerged as orthogonal in this study. We found that items from the Treatment Orientation/ Process and the Administrative and Staff domains overlapped, and items in the Women's Services and General Services domains overlapped, to the extent that each pair formed one scale instead of two. Hence, the resulting three scales appear to capture different and unique aspects of the GS construct and support a multidimensional taxonomy.

The scales are intended to study treatment outcomes and the effectiveness, cost-effectiveness, and cost-benefits of GS programming. Given adequate fit of the models, as observed here, the underlying GS trait is measured on a true interval scale. The strong program reliabilities obtained for the scales allow us to confidently group the study programs into strata distributed along the GS continuum. This sets the stage for examination of the relationship between GS treatment and patient outcomes. In addition, agencies, clinicians, and policy makers may wish to consider the relative value of specific GS services and practices to a treatment organization. The challenges and value of implementing a particular service can now be considered by programs wishing to increase their ability to provide GS treatment.

#### 4.2 Limitations

The generalizability of the current findings could be impacted by policy and practice standards for mixed-gender STR elsewhere. Although Rasch findings are widely viewed as sample independent (Embretson and Reise, 2000), the extent to which the observed multidimensional structure would be replicated in other treatment systems is unclear. Additional studies are needed to systematically evaluate naturally-occurring GS treatment services and practices in other locations.

The contribution of children's services, an important aspect of GS care (Ashley et al., 2003; Grella, 2008), could not be evaluated in the present study. In STR, mothers are not allowed to have their children with them; consequently, programs offered limited services for children and little variability in children's services was observed across programs. Future studies should include modalities that are more likely to provide children's services, such as long-term residential treatment programs, so that this aspect of the GS taxonomy can be examined. In the Organizational Characteristics domain described by Grella (2008),

relatively few items were available for analysis. In part, this reflects a degree of structural similarity among the STR programs, since many of the interview items targeting this domain showed little variability. Further, the remaining items in this area (e.g., the proportion of women's STR beds and proportion of women referred by the child welfare system) did not appear strongly related. Future research should consider additional criteria to rate the organizational characteristics of GS programs.

Due to the limited number of programs in this study, it was not possible to fit multidimensional Rasch (Briggs and Wilson, 2003) or more complicated item response theory models (Baker and Kim, 2004) to our data (personal communication, Linacre and Wilson, 2010). Instead, Rasch analysis was performed for each domain. This approach allowed us to ensure that analytic assumptions were met, that the included items demonstrated adequate fit, and the measurement properties of each scale were strong. The small sample size may have led us to drop potentially significant items due to restricted statistical power or to outlier effects. Although the study generated scales with sound measurement properties, the addition of items near the floor and ceiling of each domain would provide increased sensitivity.

Rasch evaluation of rating scale effectiveness ideally proceeds with at least 10 observations in each response category (Linacre, 1994) in order to obtain stable measure estimates. To reach toward this standard and to confirm program score stability, we dichotomized the item responses and re-analyzed the data, as suggested by Linacre (personal communication, July 2010). Although identical program rankings were not maintained, the reanalyzed scales placed programs in comparable strata, thus supporting the utility of the selected item categories and program scores. In order to reliably maximize the GS differences between programs, several over-fitting items (infit and outfit statistics  $< .50$ ) were retained. In each case, the variable had been recoded to combine adjacent disordered categories. These items were retained because they provided new information to differentiate the programs, which forwarded the main objective of the study. Future studies would benefit from a larger sample of programs and consideration of additional treatment modalities.

## 5. Conclusions

Short instruments that maintain high quality are needed to measure inter-program variability in gender sensitivity, both for researchers and administrators. Empirically-derived scales measuring GS treatment were developed using Rasch modeling techniques for Treatment Orientation and Staff Training, Women's and General Services, and Physical Environment. The study findings demonstrate for the first time a naturally occurring variation of GS programming in mixed-gender programs. These quantifiable program differences occur in qualitatively different domains. The measures provide a link from GS programming to treatment outcomes and allow the evaluation of the effectiveness, cost-effectiveness, and cost-benefit of GS programming for women. Moreover, these scales and items offer useful tools and lay the foundation for future studies assessing and measuring GS services. The identified treatment services and practices can inform program administrators, clinicians and policy makers about the state of GS programming and what aspects may be improved to better serve the needs of substance abusing women.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



## Acknowledgments

The authors wish to thank NIDA for its financial support of this study. Sincere thanks also go to Dr. Kevin Campbell, Dr. Alice Huber, and Sue Green at Washington State's Department of Social and Health Services, Division of Behavioral Health and Recovery for their administrative, data support, and consultation regarding program operations, to Dr. Christine Grella at UCLA for her expert consultation on gender-sensitive treatment, and to Dr. Shelly Greenfield at McLean Hospital/Harvard Medical School for her thoughtful and constructive review of the interview protocols.

**Role of Funding Source** Funding for this study was provided by NIDA Grant R01DA020082. NIDA had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

## References

- Ashley OS, Marsden ME, Brady TM. Effectiveness of substance abuse treatment programming for women: a review. *Am. J. Drug Alcohol Abuse*. 2003; 29:19–53. [PubMed: 12731680]
- Baker, FB.; Kim, S. *Item Response Theory: Parameter Estimation Techniques*. 2nd Ed.. CRC Press; New York: 2004.
- Bloom B. Gender-responsive programming for women offenders: guiding principles and practices. *Correctional Service of Canada*. 1999; 11:3.
- Bloom, B.; Owen, BA.; Covington, S. *Gender-Responsive Strategies: Research, Practice, and Guiding Principles for Women Offenders*. US Dept. of Justice, National Institute of Corrections; Washington, DC: 2003.
- Brady KT, Randall CL. Gender differences in substance use disorders. *Psychiatr. Clin. North Am*. 1999; 22:241–252. [PubMed: 10385931]
- Briggs DC, Wilson M. An introduction to multidimensional measurement using Rasch models. *J. Appl. Meas*. 2003; 4:87–100. [PubMed: 12700433]
- Chen X, Burgdorf K, Dowell K, Roberts T, Porowski A, Herrell JM. Factors associated with retention of drug abusing women in long-term residential treatment. *Eval. Program Plann*. 2004; 27:205–212.
- Claus RE, Orwin RG, Kissin W, Krupski A, Campbell K, Stark K. Does gender-specific substance abuse treatment for women promote continuity of care? *J. Subst. Abuse Treat*. 2007; 32:27–39. [PubMed: 17175396]
- Covington SS. Women and addiction: a trauma-informed approach. *J. Psychoactive Drugs Suppl*. 2008; 5:377–385.
- Covington, SS.; Bloom, B. [Accessed on 1-11-2011] Gender-Responsive Program Assessment. 2008. <http://www.centerforgenderandjustice.org/pdf/GRProgramAssessmentTool%20CJ%20Final.pdf>
- Covington, SS.; Surrey, JL. The relational model of women's psychological development: implications for substance abuse. Wilsnack, RW.; Wilsnack, SC., editors. *Rutgers Center of Alcohol Studies*; New Brunswick, N.J.: 1997. p. 335-351.
- Dowell, K.; Orwin, RG.; Herrell, JM.; Deang, L.; Bernichon, T. *Lessons learned: Residential substance Abuse Treatment for Women and Their Children (DHHS Publication No. (SMA) 03-3787)*. Center for Substance Abuse Treatment; Rockville, MD: 2003. A portrait of CSAT's women and children's projects and the clients they serve.
- Embretson, SE.; Reise, SP. *Item Response Theory for Psychologists*. Lawrence Erlbaum; Mahwah, NJ: 2000.
- Green CA. Gender and use of substance abuse treatment services. *Alcohol Res. Health*. 2006; 29:55–62. [PubMed: 16767855]
- Green CA, Polen MR, Dickinson DM, Lynch FL, Bennett MD. Gender differences in predictors of initiation, retention, and completion in an HMO-based substance abuse treatment program. *J. Subst. Abuse Treat*. 2002; 23:285–295. [PubMed: 12495790]
- Greenfield L, Burgdorf K, Chen X, Porowski A, Roberts T, Herrell J. Effectiveness of long-term residential substance abuse treatment for women: findings from three national studies. *Am. J. Drug Alcohol Abuse*. 2004; 30:537–550. [PubMed: 15540492]

- Greenfield SF, Back SE, Lawson K, Brady KT. Substance abuse in women. *Psychiatr. Clin. North Am.* 2010; 33:339–355. [PubMed: 20385341]
- Greenfield SF, Brooks AJ, Gordon SM, Green CA, Kropp F, McHugh RK, Lincoln M, Hien D, Miele GM. Substance abuse treatment entry, retention, and outcome in women: a review of the literature. *Drug Alcohol Depend.* 2007; 86:1–21. [PubMed: 16759822]
- Greenfield SF, Grella CE. What is “women-focused” treatment for substance use disorders? *Psychiatr. Serv.* 2009; 60:880–882. [PubMed: 19564216]
- Grella CE. From generic to gender-responsive treatment: changes in social policies, treatment services, and outcomes of women in substance abuse treatment. *J. Psychoactive Drugs Suppl.* 2008; 5:327–343.
- Grella CE, Greenwell L. Substance abuse treatment for women: changes in the settings where women received treatment and types of services provided, 1987–1998. *J. Behav. Health Serv. Res.* 2004; 31:367–383. [PubMed: 15602139]
- Grella CE, Lovinger K. 30-Year trajectories of heroin and other drug use among men and women sampled from methadone treatment in California. *Drug Alcohol Depend.* 2011; 118:251–258. [PubMed: 21549528]
- Grella CE, Scott CK, Foss MA, Dennis ML. Gender similarities and differences in the treatment, relapse, and recovery cycle. *Eval. Rev.* 2008; 32:113–137. [PubMed: 18198172]
- Hodgins DC, El-Guebaly N, Addington J. Treatment of substance abusers: single or mixed gender programs? *Addiction.* 1997; 92:805–812. [PubMed: 9293039]
- Hser YI, Niv N. Pregnant women in women-only and mixed-gender substance abuse treatment programs: a comparison of client characteristics and program services. *J. Behav. Health Serv. Res.* 2006; 33:431–442. [PubMed: 16688389]
- Kaskutas LA, Zhang L, French MT, Witbrodt J. Women's programs versus mixed-gender day treatment: results from a randomized study. *Addiction.* 2005; 100:60–69. [PubMed: 15598193]
- Kline, T. *Psychological Testing: A Practical Approach to Design and Evaluation.* Sage Publications, Inc.; Thousand Oaks, CA: 2005.
- Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions.* 1994; 7:328.
- Linacre JM. Structure in Rasch residuals: why principal components analysis. *Rasch Measurement Transactions.* 1998; 12:636.
- Linacre JM. Rasch Power Analysis: size vs. significance: standardized chi-square fit statistic. *Rasch Measurement Transactions.* 2003; 17:918.
- Linacre, JM. *Winsteps® (Version 3.70.0) [Computer Software].* Beaverton; Oregon: 2009. Winsteps.com
- Linacre, JM. *A User's Guide to WINSTEPS® and MINISTEPS: Rasch-Model Computer Programs: Program Manual 3.70. 0.* Beaverton; Oregon: 2010. Winsteps.com
- Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient in the creation of minimum item sets. *Med. Care Suppl.* 2004; 1:17–24.
- Niv N, Hser YI. Women-only and mixed-gender drug abuse treatment programs: service needs, utilization and outcomes. *Drug Alcohol Depend.* 2007; 87:194–201. [PubMed: 16996232]
- O'Brien MS, Anthony JC. Risk of becoming cocaine dependent: epidemiological estimates for the United States, 2000–2001. *Neuropsychopharmacology.* 2005; 30:1006–1018. [PubMed: 15785780]
- Pelissier B, Jones N. A review of gender differences among substance abusers. *Crime Delinq.* 2005; 51:343–372.
- Prendergast ML, Messina NP, Hall EA, Warda US. The relative effectiveness of women-only and mixed-gender treatment for substance-abusing women. *J. Subst Abuse Treat.* 2011:336–348. [PubMed: 21315540]
- Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests (Reprint, With Foreword and Afterword by B. D. Wright, 1980).* University of Chicago Press; Chicago: 1960.
- Weiss SRB, Kung HC, Pearson JL. Emerging issues in gender and ethnic differences in substance abuse and treatment. *Curr. Womens Health Rep.* 2003; 3:245–253. [PubMed: 12734036]

- Westermeyer J, Boedicker AE. Course, severity, and treatment of substance abuse among women versus men. *Am. J. Drug Alcohol Abuse*. 2000; 26:523–535. [PubMed: 11097190]
- Women's Services Practice Improvement Collaborative. *Treatment Guidelines: Gender Responsive Treatment of Women with Substance Use Disorders*. Connecticut Department of Mental Health and Addiction Services. 2007
- Wright BD, Linacre JM, Gustafson JE. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994; 8:370.
- Zilberman ML, Tavares H, Blume SB, El-Guebaly N. Substance use disorders: sex differences and psychiatric comorbidities. *Can. J. Psychiatry*. 2003; 48:5–13. [PubMed: 12635558]

**Table 1**

GS treatment domains.

<b>Domain</b>	<b>Description</b>	<b>Items *</b>
Treatment Orientation and Processes	Women as priority, treatment model/approach, use of evidence-based practices	30
Administrator and Staff	Staff education and training, staff competencies, gender of program director, % female staff	37
Organizational Characteristics	Program age, capacity, accreditation, patient case-mix, referral sources	4
Women's Services	Prenatal/postnatal services, women only groups, parenting training/counseling, trauma/abuse counseling, women's health services	24
General Services	GS assessment, on-site mental health services, case management, individual counseling, family therapy, HIV education, 12-step groups, transportation, housing	43
Children's Services	Onsite child care, live-in accommodations, coordination with Children's Protective Services, counseling/mental health services	7
Physical Environment	Safety and security, cleanliness, spatial layout, social/recreation spaces	25

\* Excluding items without variability.

**Table 2**  
Gender-sensitivity estimates and fit statistics for Treatment Orientation & Staff Training items.

Training & Treatment Orientation	Gender Sensitivity			Infit			Outfit			Item-scale correlation
	Estimate	SE	MNSQ	Z	MNSQ	Z	MNSQ	Z		
Women's substance use and mental health	2.16	0.51	1.41	0.90	1.72	0.90	0.90	0.90	0.55	
Mean time staff spend per year on women-specific training	2.02	0.67	0.87	-0.10	0.76	0.30	0.30	0.30	0.61	
Women's sexuality	1.33	0.43	1.04	0.30	0.96	0.30	0.30	0.30	0.74	
Mission statement addresses behavioral health	1.12	0.76	1.27	0.80	1.28	0.70	0.70	0.70	0.46	
Role of co-occurring psychiatric disorders in women's recovery	1.01	0.45	0.56	-0.90	0.48	-0.30	-0.30	-0.30	0.82	
Effect on women of trading sex for drugs	0.91	0.42	1.08	0.30	0.91	0.20	0.20	0.20	0.76	
Role of parenting/caretaking in recovery	0.81	0.72	0.43	-1.50	0.29	-0.90	-0.90	-0.90	0.85	
Role of trauma/re-traumatization of women	0.02	0.61	0.46	-1.10	0.41	-0.90	-0.90	-0.90	0.80	
Community supports for women	0.01	0.39	1.42	1.00	1.31	0.70	0.70	0.70	0.75	
Cultural issues in women's treatment	-0.11	0.48	1.24	0.70	0.98	0.20	0.20	0.20	0.78	
Trauma/PTSD (% clinical staff with training)	-0.23	0.52	0.87	-0.20	0.56	0.10	0.10	0.10	0.78	
Sexual abuse	-0.29	0.39	0.70	-0.70	0.78	-0.20	-0.20	-0.20	0.84	
Family violence	-0.44	0.39	0.57	-1.10	0.56	-0.70	-0.70	-0.70	0.86	
Culturally-relevant treatment (% clinical staff with training)	-0.76	0.40	1.10	0.40	0.87	0.00	0.00	0.00	0.83	
Women's issues (% clinical staff with training)	-1.76	0.63	1.24	0.70	1.14	0.40	0.40	0.40	0.73	
Treat women with acute psychiatric conditions	-5.82	1.35	1.35	0.80	0.25	-0.30	-0.30	-0.30	0.50	
<b>Item Mean</b>	<b>0.00</b>	<b>0.57</b>	<b>0.98</b>	<b>0.02</b>	<b>0.83</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.73</b>	
<b>SD</b>	<b>1.81</b>	<b>0.23</b>	<b>0.33</b>	<b>0.80</b>	<b>0.39</b>	<b>0.50</b>	<b>0.50</b>	<b>0.50</b>	<b>0.13</b>	

Table 3

Gender-sensitivity estimates and fit statistics for Women's &amp; General Services items.

Women's & General Services	Gender Sensitivity			Infit			Outfit			Item-Scale Correlation
	Estimate	SE	MNSQ	Z	MNSQ	Z	MNSQ	Z		
	Vocational needs assessed	2.63	0.62	0.35	-2.10	0.30	-0.30	0.30	-0.30	
Trauma counseling (% who receive)	2.57	0.64	1.34	0.90	1.41	0.70	1.41	0.70	0.55	
Life skills assessed	1.90	0.38	0.45	-1.40	0.57	0.10	0.57	0.10	0.83	
Women connected to social supports by discharge (how often)	1.44	0.81	0.54	-1.20	0.35	-0.10	0.35	-0.10	0.72	
Parenting skills assessed	1.31	0.72	0.83	-0.40	0.60	0.10	0.60	0.10	0.62	
Percentage of treatment plans strengths-based	0.81	0.70	1.36	1.10	1.17	0.60	1.17	0.60	0.46	
Percent of women with concrete post-treatment housing plan	0.81	0.70	1.17	0.60	0.99	0.50	0.99	0.50	0.52	
Spiritual, religious, or cultural needs assessed	0.80	0.50	0.89	-0.20	0.70	0.10	0.70	0.10	0.71	
Grief and loss history assessed	0.32	0.49	1.08	0.30	0.94	0.30	0.94	0.30	0.67	
Physical activities (% who participate)	0.32	0.49	1.53	1.30	1.28	0.60	1.28	0.60	0.58	
Medical services (other than psychiatric or perinatal) available onsite	-0.05	0.74	1.43	1.30	1.25	0.60	1.25	0.60	0.45	
Domestic violence history assessed	-0.11	0.46	0.36	-1.90	0.25	-0.30	0.25	-0.30	0.77	
Housing needs assessed	-0.18	0.71	0.56	-1.60	0.39	-0.10	0.39	-0.10	0.71	
Cultural or spiritual activities (% who participate)	-0.18	0.50	1.69	1.60	1.28	0.60	1.28	0.60	0.56	
Counselor availability	-0.21	0.83	0.98	0.20	1.24	0.50	1.24	0.50	0.60	
Safety concerns assessed	-0.29	0.41	0.35	-1.60	0.27	-0.30	0.27	-0.30	0.83	
Assertiveness/self-efficacy training (% who receive)	-0.63	0.54	1.85	1.70	1.80	0.90	1.80	0.90	0.55	
Family planning, Fetal Alcohol Spectrum Disorder education (% who receive)	-0.71	0.75	0.78	-0.60	0.48	0.00	0.48	0.00	0.64	
Social support assessed	-1.12	0.86	0.61	-0.90	0.32	-0.20	0.32	-0.20	0.70	
Counseling about healthy relationships (% who receive)	-1.12	0.54	1.39	0.90	1.02	0.50	1.02	0.50	0.60	
Sexuality assessed	-1.17	0.65	1.31	0.90	1.28	0.60	1.28	0.60	0.61	
Stage of change assessed	-1.40	0.44	1.14	0.40	0.53	0.10	0.53	0.10	0.76	
Social or recreational activities (% who participate)	-1.45	0.60	1.37	0.80	0.73	0.30	0.73	0.30	0.63	
Assessment sensitive to vulnerability to re-traumatization	-1.81	1.08	0.84	0.00	0.35	-0.10	0.35	-0.10	0.67	
Safer sex education (% who receive)	-2.47	0.89	0.77	0.10	0.54	0.10	0.54	0.10	0.77	
<b>Item Mean</b>	<b>0.00</b>	<b>0.64</b>	<b>1.00</b>	<b>0.01</b>	<b>0.80</b>	<b>0.23</b>	<b>0.80</b>	<b>0.23</b>	<b>0.65</b>	
<b>SD</b>	<b>1.29</b>	<b>0.17</b>	<b>0.43</b>	<b>1.10</b>	<b>0.43</b>	<b>0.40</b>	<b>0.43</b>	<b>0.40</b>	<b>0.11</b>	

**Table 4**

Gender-sensitivity estimates and fit statistics for Physical Environment items.

Physical Environment	Gender Sensitivity			Infit			Outfit			Item-Scale Correlation
	Estimate	SE	MNSQ	Z	MNSQ	Z	MNSQ	Z		
Bedrooms (inviting/livable vs. cold/institutional)	3.28	1.34	0.32	-0.80	0.08	-0.70	0.08	-0.70	0.80	
Dining area(s) (inviting/livable vs. cold/institutional)	1.95	0.72	0.45	-0.90	0.39	-0.70	0.39	-0.70	0.90	
Clinical rooms (inviting/livable vs. cold/institutional)	1.91	1.03	0.33	-1.00	0.14	-0.90	0.14	-0.90	0.91	
Safety and security of facility for women (e.g., isolated areas, supervision)	1.05	0.84	1.33	0.70	0.92	0.10	0.92	0.10	0.52	
Privacy meeting with visitors	1.00	0.44	1.32	0.80	1.27	0.70	1.27	0.70	0.70	
Bathrooms and showers (inviting/livable vs. cold/institutional)	0.27	0.41	1.44	1.10	1.31	0.80	1.31	0.80	0.72	
Privacy of showers	-0.05	0.39	0.90	-0.10	0.92	-0.10	0.92	-0.10	0.65	
Number of women per bathroom	-0.31	0.70	0.83	-0.60	0.70	-0.20	0.70	-0.20	0.58	
Lounge and recreational area(s) (inviting/livable vs. cold/institutional)	-0.49	0.38	0.59	-1.20	0.63	-1.00	0.63	-1.00	0.82	
Inviting, livable, and homey overall (vs. cold and institutional)	-0.63	0.37	0.93	-0.10	0.86	-0.30	0.86	-0.30	0.77	
Privacy while meeting with counselor or other clinical staff	-0.77	0.37	1.76	1.90	1.67	1.60	1.67	1.60	0.69	
Privacy of toilets	-0.77	0.37	0.48	-1.70	0.76	-0.60	0.76	-0.60	0.70	
Separation of men's and women's sleeping, bathing, residential areas	-1.04	0.37	1.00	0.10	0.90	-0.10	0.90	-0.10	0.67	
Safety of program location	-1.35	0.23	0.76	-0.60	0.51	0.10	0.51	0.10	0.53	
Overall cleanliness	-1.87	0.38	1.52	1.30	1.29	0.70	1.29	0.70	0.51	
Overall safety and security of neighborhood	-2.17	0.40	1.34	0.90	1.28	0.70	1.28	0.70	0.44	
<b>Item Mean</b>	<b>0.00</b>	<b>0.55</b>	<b>0.96</b>	<b>-0.01</b>	<b>0.85</b>	<b>0.01</b>	<b>0.85</b>	<b>0.01</b>	<b>0.68</b>	
<b>SD</b>	<b>1.45</b>	<b>0.29</b>	<b>0.44</b>	<b>1.00</b>	<b>0.43</b>	<b>0.70</b>	<b>0.43</b>	<b>0.70</b>	<b>0.14</b>	