University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Special Education and Communication Disorders
Faculty Publications

Department of Special Education and
Communication Disorders

2013

Measurement Properties of the Motivation for Youth Treatment Scale with a Residential Group Home Population

Matthew C. Lambert
University of Nebraska-Lincoln, mlambert2@unl.edu

Kristin Duppong-Hurley
University of Nebraska–Lincoln, kristin.hurley@unl.edu

M. Michele Athay Tomlinson
Vanderbilt University, michele.athay@vanderbilt.edu

Amy L. Stevens
Father Flanagan's Boys Home, amy.stevens@boystown.org

Follow this and additional works at: http://digitalcommons.unl.edu/specedfacpub

Part of the Special Education and Teaching Commons

Measurement Properties of the Motivation for Youth Treatment Scale with a Residential Group Home Population

Matthew C. Lambert,
University of Nebraska-Lincoln, 211 Barkley Memorial Center, Lincoln, NE 68583, mlambert2@unl.edu

Kristin Duppong Hurley,
University of Nebraska-Lincoln, 204F Barkley Memorial Center, Lincoln, NE 68583, kristin.duppong@unl.edu

M. Michele Athay Tomlinson, and
Vanderbilt University, 151 Peabody, Nashville, TN 37203, michele.athay@vanderbilt.edu

Amy L. Stevens
Father Flanagan's Boys Home, 100 Crawford Drive, Boys Town, NE 68010, amy.stevens@boystown.org

Measurement Properties of the Motivation for Youth's Treatment Scale with a Residential Group Home Population

A client's motivation to receive services has long been identified as a highly relevant component of mental health treatment. In fact, ample evidence demonstrates that client motivation is significantly related to seeking services, remaining in services, and improved client outcomes (e.g., Broome, Joe, & Simpson, 2001; Ryan, Plant, & O'Malley, 1995; Schroder, Sellman, Frampton, & Deering, 2009). Additionally, it has been recognized that motivation is a "dynamic" characteristic that changes throughout treatment (Melnick, De Leon, Hawke, Jainchill, & Kressel, 1997; Schroder et al., 2009). In this way, motivation is an important client factor to assess and monitor throughout the treatment process.

The broad construct of motivation is comprised of two separate, but related components conceptualized as *motivation to change* and *motivation for treatment*. As defined by DiClamente, Schlundt, and Gemmell (2004), motivation to change refers to a willingness to recognize problematic behavior and take steps toward change, whereas motivation for treatment refers to a willingness to seek help and remain compliant with an intervention program. In other words, a motivated person not only perceives the importance of changing, but also has confidence that they are able to be successful at making the change (Burke, Arkowitz, & Menchola, 2003).

Although motivation may be relevant for all client age groups (i.e., children, adults, etc.) in any treatment setting (i.e., outpatient, inpatient, etc.) there has been limited research investigating its importance for youth receiving services in residential settings for primarily behavioral and emotional needs. A systematic review of residential care effectiveness pointed to client qualities, in general, that might be predictive of treatment success (Betmann & Jasperson, 2009); however, none of the reviewed studies directly addressed the role of motivation to treatment in residential group care. Recently, Harder, Knorth, and Kalverboer (2012) examined treatment motivation with 22 youth in a secure residential setting serving juvenile offenders in the Netherlands. They found that service satisfaction was correlated with higher motivation for treatment and that treatment motivation increased from admission to discharge. However, results from their small sample also suggested that youth with high

motivation for treatment at intake tended to deteriorate by discharge. A study by Edelen and colleagues (2007) assessed aspects of motivation in adolescents treated in a therapeutic residential community for substance abuse. They found that adolescents who displayed increased problem recognition (i.e., motivation to change) were more likely to remain in treatment for 90 days or more, and were then more likely to maintain compliance with post-treatment plans such as attending 12-step meetings. Not only has research consistently confirmed this relationship between motivation and treatment retention (e.g., DeLeon, Hawke, Jainchill, Kressel, & Melnick, 1999; Schroder et al., 2009), evidence also indicates that motivation is significantly related to therapeutic involvement and engagement, as well as compliance behaviors within treatment (Broome, Joe, & Simpson, 2001; Hiller, Knight, Leukefeld, & Simpson, 2002; Laurier, Lafortune, & Collin, 2010). Nonetheless, there has been minimal research on motivation for treatment for adolescents in residential group homes serving youth with behavioral and emotional issues, aside from programs specifically targeting substance use (e.g., Edelen et al., 2007; Kelly, Urbanoski, Hoeppner, & Slaymaker, 2012) or secure settings for juvenile offenders (Harder et al., 2012). Accordingly, little is known regarding the measurement properties of scales that measure motivation for treatment in residential group care settings. To date, only one scale, the *Circumstances, Motivation, Readiness, and Suitability Scale* (DeLeon et al., 1994) has been evaluated with a residential group care population of youth (substance abuse treatment); however, it should be noted that the focus of this particular evaluation was primarily on the psychometrics of the *suitability of residential care* subscale.

Motivation for Youth Treatment Scale (MYTS)

There are currently several measures used to assess youth motivation for mental health treatment such as the Problem Recognition Questionnaire (PRQ; Cady, Winters, Jordan, Solberg, & Stinchfield, 1996), the University of Rhode Island Change Assessment Scale (DiClemente & Hughes, 1990) and the Treatment Motivation Questionnaire (TMQ; Ryan et al., 1995). However, a review of available measures by researchers at Vanderbilt University found none that were psychometrically sound while also being suitable for frequent administration in diverse treatment settings (Breda & Riemer, 2012). Therefore, the Motivation for Youth Treatment Scale (MYTS; Bickman et al., 2010; Breda & Riemer, 2012) was developed. The MYTS is one of the 11 clinical measures found in the Peabody Treatment Progress Battery (PTPB) used to assess the process and progress of mental health treatment for youth (Bickman et al., 2010; Riemer et al., 2012). Containing only 8-items, the MYTS was created to be a brief measure that can be administered frequently throughout services to assess motivation.

The MYTS measures motivation by assessing both problem recognition and readiness to participate in treatment. Therefore, completion of the MYTS yields a total score for overall motivation, as well as subscale scores for problem recognition and readiness to change. With the goal of maintaining a brief measure that captures as much reliable and valid information at possible, the MYTS has undergone two comprehensive rounds of psychometric testing utilizing methods from classical test theory (CTT), item response theory (IRT), and factor analysis (Bickman et al., 2007, 2010; Breda & Riemer, 2012). Most recently, the item and scale properties of the 8-item version of the MYTS were evaluated in a large sample of clinically referred youth, and their caregivers, receiving community-based mental health services (Breda & Riemer, 2012). Results found both MYTS forms to be psychometrically sound with internal reliability estimates ranging from 0.82 to 0.89 for total and subscale scores, adequate item properties, and a two-factor structure confirmed with confirmatory factor analysis. Additionally, consistent with the literature, Breda and Riemer (2012) found youth symptom severity to be a significant predictor of treatment motivation. In addition, within the context of residential group care, Lambert, Duppong Hurley, Pick, and Thompson

(2013) suggested that greater motivation for treatment was linked to greater youth symptom severity (at intake) as well as greater treatment expectations, greater therapeutic alliance, and more substantial perceived counseling impact. Taken together, these results provide evidence that the MYTS is a brief, reliable, and valid (e.g., evidence of convergent and predictive validity) assessment tool that can be used regularly throughout a variety of treatments and, potentially, across several mental health and child welfare populations.

While several assessments of motivation for treatment exist for counseling and substance abuse treatments, as noted above, none have been thoroughly evaluated for psychometric quality with a sample of youth receiving residential group care services. Therefore, the purpose of this study was to conduct a comprehensive psychometric analysis of the youth MYTS measure in a residential facility focused on serving youth with emotional and behavioral needs. We hypothesized that the measurement properties for youth in residential group care would be acceptable and largely similar to those found for out-patient counseling samples although there is no existing, empirical evidence to support this hypothesis. Following the model of prior reports on the MYTS (Bickman et al., 2007, 2010; Breda & Riemer, 2012) which have presented results of classical test theory analyses, confirmatory factor analyses, and Rasch model analyses, we also used these three analytic approaches to evaluate the MYTS.

Method

Setting

The study was conducted at a residential group home facility that serves over 500 girls and boys in 70 family-style group homes in a large Midwestern city. The agency serves youth from around the nation with most youth referrals coming from child welfare and juvenile justice agencies with funding streams that closely match the referral sources (Mason, Chmelka, & Thompson, 2012). The primary goal of the program, which is accredited by the Council on Accreditation (COA), is to "provide care for at-risk youth ages 10 – 18 who cannot live at home and whose behavioral and emotional needs require more intensive treatment" (Mason, Chmelka, & Thompson, 2012). To this end, the agency uses an adaptation of the Teaching Family Model (TFM; Davis & Daly, 2003; Wolf et al., 1976) which focuses primarily on developing social skills and building positive relationships between peers and with adults. The agency employs married couples as the primary service delivery agents; these couples live in a family-style home with up to eight adolescent girls or boys on the agency's campus. The facility is staff-secured; youth are closely monitored, but they are free to move about to school and campus activities (ROLES = 12; group homes).

Participants

Youth eligible to participate in the study were identified with a disruptive behavior diagnosis (e.g., oppositional defiant disorder, conduct disorder, or attention deficit hyperactivity disorder) via a professional diagnosis, Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), or the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001). Forty-eight percent of the sample was identified with conduct disorder, 34% with oppositional defiance disorder, and 18% with ADHD. Other inclusion criteria included were at least 10 years old, were experiencing their first admission to the agency, and were assigned to service providers that consented to participate e in the study. Based on these criteria, over a two year period, 170 youth were eligible for participation and 145 (85%) had guardian consent and youth assent to participate. The sample of youth consisted of 62 girls and 83 boys. Thirty-two youth indicated that they were Hispanic, 76 Caucasian, 53 African American, and 30 other. Age at enrollment ranged from 10 to 17 years, with a mean age of 15.2 years ($SD = 1.39$). Mean CBCL scores for the

sample, collected at 2 months into care, were 57.91 (*SD* = 9.74) and 51.44 (*SD* = 10.33) for externalizing and internalizing problems, respectively. These scores indicate that, on average, youth had problem severity in the *normal* range (Achenbach & Rescorla, 2001); however, 50% of the sample had *borderline* or *clinical* level externalizing severity, and 19.8% had borderline or clinical level internalizing severity. It should be noted that since the scores were collected 2 months into care, the scores are likely to lower compared to the level of severity upon entry into care.

Procedures

Youth completed the MYTS as a part of an assessment battery for a larger, longitudinal non-experimental study on common therapeutic process factors and implementation fidelity in residential group care (see Duppong Hurley et al., 2013). Youth completed the MYTS, via a web-based survey, between admission into the facility and one month into care. Most youth completed the MYTS within 2 weeks of admission (*M* = 9.3 days, *SD* = 7.7 days). All recruitment and consent procedures for youth were approved by the University of Nebraska-Lincoln IRB and the service agency IRB.

Analysis Plan

To investigate the psychometrics of the MYTS, we used classical tests theory (CTT), Rasch modeling, and confirmatory factor analysis (CFA) approaches. Each approach yields unique indicators of the measurement quality of the individual items as well as the overall scale. The CTT approach provides typical indicators of psychometric quality such as Cronbach's alpha, and summary and central tendency statistics. Rasch modeling provides comprehensive item-level properties such as item difficulty, item discrimination, and item fit in addition to indicators of scale functioning (i.e., category characteristic curve plots). CFA modeling yields sample dependent indicators of psychometric quality, but provides a comprehensive evaluation of the internal structure of the assessment. SPSS v20 was used to calculate CTT statistics, Winsteps v3.73 (Linacre, 2012a) was used to fit a rating-scale Rasch model (Andrich, 1978), and Mplus v6 (Muthén & Muthén, 1998–2009) was used to fit a CFA model.

Two CFA models were fit to the data: (a) one with all items loading onto a single latent factor; and (b) a second model with items loading onto one of two correlated latent factors, treatment readiness or problem recognition. Both models were specified without correlated residual variances across items. Maximum likelihood with robust standard errors (MLR) was used to estimate the models and the factors were scaled using an effects-coding method (Little, Slegers, & Card, 2006) where unstandardized factor loadings were constrained to average to 1 across items. This approach to scaling produces latent variables on the same scale as the items (that is, factor scores range from 1 to 5). The fit of the models was compared using the Satorra-Bentler scaling correction chi-square difference test (Satorra & Bentler, 2010). Missing data for the CFA models were minimal (< 1%), treated as *missing at random* (MAR) and included in the analysis by using a model-based missing data approach (i.e., full information maximum likelihood).

Results

Table 1 lists the mean, standard deviation, skewness and kurtosis as well as Rasch item-level indicators and CFA factor loadings for each item on the MYTS. All of the items demonstrated acceptable distributional characteristics (< |2| for skewness and < 2 for kurtosis; Harlow, 2005), indicating that youth responded to the items with a variety of ratings. Responses to a majority of items exhibited a negative skew indicating that the majority of youth rated the items using the higher response options while only a few youth

rated the items using the lower response options. The *Total Score* mean (3.18) suggests that youth had a moderate level of motivation for treatment following admission into residential group care (Bickman et al., 2010). The subscale means indicate that, for the sample of youth, treatment readiness ($M = 3.60$) and problem recognition ($M = 2.75$) were moderate (Bickman et al., 2010). Twenty-two percent of youth reported low treatment readiness, 54% reported medium and 24% reported high treatment readiness. On the other hand, 24.1% of youth reported low problem recognition, 51.7% reported medium and 23.1% reported high problem recognition.

The internal consistency of the MYTS was acceptable (.80; Kline, 1999) at the subscale level, but less than acceptable for the total score. The total scale, treatment readiness and problem recognition subscales had Cronbach's alphas of .74, .81, and .83, respectively. The alphas suggest that the subscales are internally consistent measures, but that the total motivation score is less reliable due to only a slight correlation between the two subscales ($r = .10$, $p = .25$).

Rasch

Rasch modeling was used to estimate item difficulty and discrimination parameters, item fit indices and category characteristic curve plots. Item difficulty refers to the degree to which youth endorse the item using high ratings; the more difficult the item the less likely youth are to endorse the item with a high rating. On the MYTS, a difficult item would represent a higher-level domain of treatment readiness or problem recognition that few youth would rate using the *agree* or *strongly agree* response option. Item discrimination refers to the degree to which items can be used to identify youth with high motivation from youth with low motivation. Item fit (infit and outfit) refers to the degree to which observed responses to an item correspond to the expected responses given the difficulty of the item and the motivation level of the youth (i.e., the youth's measure). Infit statistics represent the fit of items for examinees in the middle of the distribution of the sample and outfit statistics represent the fit of items for examinees in the tails of the distribution of motivation (Linacre, 2012a). For evidence of item-level psychometric quality, we focused on item difficulty, discrimination and fit. For evidence of rating scale functioning, we focused on the category characteristic curves plot and the corresponding statistics (e.g., item and person reliability, mean person measure, category fit indexes, etc.).

The results of the Rasch measurement model reported only non-extreme cases (n = 142) because the standing on the underlying latent trait for extreme cases (those with maximum or minimum scores) cannot be estimated reliably (Bond & Fox, 2002). In this case, the underlying latent trait (also referred to as *measure* in Rasch output) represents motivation for treatment. For person parameters, higher measure values indicate greater motivation to change. For item parameters, higher measure values indicate more difficult items. For precise measurement along the entire latent continuum, effective tests include a range of easy to difficult items (Bickman et al., 2010).

Item-level Rasch indicators (i.e., measure, discrimination, infit, and outfit) are presented in Table 1. All of the items on the MYTS except for the third Problem Recognition item in the table (*feelings bother me*) demonstrate robust item fit and discrimination properties (Bond & Fox, 2002; Linacre, 2006). Acceptable unstandardized infit and outfit values range from 0.6 to 1.4 (Wright & Linacre, 1994). The discrimination parameter was centered at 1 for the analyses, so items with values below 1 indicate less discriminating efficacy. The third Problem Recognition item has acceptable fit, but a relatively low discrimination parameter (. 50) indicating that the item is not as useful for distinguishing between youth with high motivation and youth with low motivation (Bond & Fox, 2002).

In addition to largely acceptable item-level Rasch properties, scale-level properties were also predominately adequate. The MYTS items, as a set, were appropriate for measuring the youths' level of motivation as indicated by the near zero mean person measure ($M = 0.13$, $SD = 0.75$). The item facet was centered at zero to aid with the interpretation of the mean difficulty of the assessment (Linacre, 2012b). In such a model, a test with a fairly even mix of difficult and easy items that align with the latent trait range of the examinees will yield a mean measure statistic around zero. In other terms, a mean around zero indicates that the spread of persons on the latent trait continuum is relatively normally distributed across the range that is measured by the MYTS items.

The item (.75) and person reliability statistics (.98) were nearly acceptable and acceptable (.8; Linacre, 2002), respectively, indicating that the model can place items more accurately than persons on the latent trait continuum. The respective separation indexes for items and persons were 1.73 and 6.56, which indicate the relative number of statistically distinguishable groups of items or examinees. For example, a person separation index of 1.73 indicates that the items on the MYTS can identify between one and two statistically distinguishable groups of examinees; these groups might represent youth with high or low motivation for treatment.

The quality of the rating scale was evaluated in the Rasch model and is presented in Figure 1 as the category characteristic curve plot. The curves in the plot represent the mean likelihood (cumulated across all items) that a youth, with a given level of motivation for treatment, would endorse items using a certain rating category. There are several features of the category characteristic curve plot that indicate proper functioning: (a) each curve should peak at a unique place on the continuum representing that each category is distinguishable from one another; (b) curves should be ordered monotonically from *strongly disagree* (1) to *strongly agree* (5); and (c) ideally, curves should peak at or above .5 on the probability scale indicating that at least 50% of youth in that range of motivation are providing ratings using that response option (Linacre, 2002). These features are achieved for two of the five categories, *strongly disagree* (1) and *strongly disagree* (5), and partially achieved for *agree* (4). This suggests that youth completing the MYTS differentiate the best between the more extreme rating categories, but cannot differentiate between the middle three categories with *disagree* (2) and *neither agree nor disagree* (3) demonstrating the poorest properties.

CFA

The indicators used to assess goodness-of-fit were chi-square (X^2), comparative fit index (CFI; Bentler, 1990), Tucker-Lewis index (TLI; Tucker & Lewis, 1973), root mean square error of approximation (RMSEA; Steiger & Lind, 1980), and standardized root mean square residual (SRMR). Multiple fit indexes are reported since each indicator gauges fit or misfit in a unique way and multiple indexes can help triangulate and strengthen conclusions on the appropriateness of fit. X^2 represents an exact test of fit and a non-significant value indicates that the model fits the data acceptably; however, X^2 is typically not used to assess model fit in applied research because it is often too conservative (Browne & Cudeck, 1993). The chi-square difference test (X^2) was computed to evaluate the fit of a one factor compared to a two-factor model. The Satorra-Bentler scaling correction was used to calculate the difference test because the models were estimated using MLR estimation (Satorra & Bentler, 2010). CFI and the TLI indexes are comparative fit indexes representing the degree of improvement over the worst fitting model (Boomsma, 2000). Both indexes are scaled from 0 to 1 with values closer to 1 indicating better fit. An acceptable fitting model has a TLI or CFI greater than or equal to .90 (Browne & Cudeck, 1993). RMSEA and SRMR represent the degree of model misfit and are reported on a scale of 0 to 1; values closer to zero indicate better fit. Values less than .08 are considered acceptable (Hu & Bentler, 1999).

Table 2 lists the model fit indicators for the one-factor model as well as the two-factor model. All of the indicators of goodness-of-fit favor the two-factor model over the one-factor model. In addition, the fit indicators for the two-factor model are largely acceptable compared to standard cutoffs for these indexes (CFI = .92, TLI = .88, RMSEA = .11, SRMR = 0.07). The RMSEA was slightly larger than the acceptable level and the TLI was slightly lower than the acceptable level, but, taken together, fit appears to be acceptable. In addition to fit indicators, all standardized factor loadings ( ) were positive and large in magnitude (> .40; see Table 1).

In the two-factor model, the correlation between treatment readiness and problem recognition was positive, small in size (Cohen, 1988), and non-significant ($r = .05$, $p = .59$). Factor score determinacies (see Lawley & Maxwell, 1971), the average correlation between the factor score matrix and the raw data matrix, were .94 and .95 for the treatment readiness and problem recognition, respectively, indicating that the factors were measured well by the items. The X^2 indicated that the two-factor model fit the data significantly better than the single factor model. Given the acceptable model fit indicators and adequate factor loadings (in addition to the robust factor score determinacies), we can be moderately confident that, for this sample, the factor structure of MYTS aligns with the hypothesized structure that consists of two factors (treatment readiness and problem recognition).

Discussion

Overall, we found moderately strong evidence of the psychometric quality of the MYTS for youth in this residential group home setting. The internal consistency of the MYTS is slightly low for the total scale, but each subscale represents a reliable measure of treatment readiness or problem recognition. None of the items indicate ceiling or floor effects and youth tend to endorse ratings across the entire scale. Although youth use all the different response options available, the rating scale does not function entirely as intended. Other Rasch indicators suggest that items exhibit strong qualities in terms of fit to the measurement model and discrimination, and, as a set, the items align well with the examinees' level of motivation (i.e., there were only three extreme cases). In addition to the acceptable findings from the Rasch model, the hypothesized 2-factor structure was supported over a one factor model and fit the data acceptably according to the CFI and SRMR fit indicators; however, other indicators (e.g., TLI and RMSEA) suggest slightly worse fit to the data.

These findings, taken as a whole, are similar to prior psychometric evaluations of the MYTS conducted with outpatient counseling populations (Bickman et al., 2007, 2010; Breda & Riemer, 2012). Prior evaluations also found that the MYTS, when rated by youth, demonstrated: (a) acceptable internal consistency at the scale and subscale levels, (b) acceptable item-level descriptive statistics, (c) largely acceptable Rasch item fit and discrimination, and (d) acceptable fit to a 2-factor CFA structure. However, while the findings of prior psychometric evaluations were largely replicated with a different population, there are a few areas where the findings of the current study diverge from the findings of prior research.

One area in which our findings are substantively different from prior findings is the small correlation between the subscales. We found that the correlation between readiness for treatment and problem recognition was $r = .10$, 95% CI [−.07, .26]. This stands in stark difference to correlations found with other populations of youth ($r = .43$, 95% CI [.36, .50]; Breda & Riemer, 2012). The low correlation between subscales is likely a significant reason as to why the reliability of the total score is borderline low for youth in residential group care compared to youth in outpatient counseling settings.

There are a couple of possibilities for this unique finding. First, this study was limited to youth with a disruptive behavior diagnosis. Perhaps this subset of youth responds differently to the items than a broader sample of youth with mental health needs. Future research in residential care should include additional diagnoses to see if the findings are replicated across type of diagnosis. Second, there are important differences between outpatient services and residential care that may help frame the lack of correlation between the treatment readiness and problem recognition scales. For outpatient youth, the treatment readiness scale speaks of being ready for help or counseling services. However, for the residential version, the reference was a residential agency. There is a stark difference to endorsing an item indicative of being ready for counseling services versus being ready to enter out-of-home care. Thus, while youth may very well score high on problem recognition, they could have no desire to enter a residential facility away from family and friends. Further, for many youth, the element of "choice" in this placement type is relatively limited, with adults primarily making their placement decisions (e.g., guardians, courts, case-workers). There may also be a sub-group of youth that believe themselves lucky to be at this particular residential agency, (e.g., home-like atmosphere, excellent schools and athletic teams) versus alternative placement settings such as correctional facilities. As suggested by the mean scores in Table 1, there are also youth that sincerely believe they do not have any emotional or behavioral problems to address. Thus, it is not particularly surprising that the correlation between the two subscales is low in a residential youth setting. Future research would benefit from examining the two subscales in out-of-home populations to see if there are distinct profiles of youth, to examine how the subscales change over time, and their relationship to outcomes.

While the findings of recent CFA analyses (Breda & Riemer, 2012) and the findings of the current CFA analysis converge on similar conclusions regarding the factor structure of the MYTS, the prior evaluation reported stronger support compared to our findings. Breda and Riemer (2012) found that all the reported fit indicators (i.e., CFI, GLI, and SRMR) were within the acceptable range of scores indicating close fit to the data (Hu & Bentler, 1999), and strongly supported a two-factor structure. On the other hand, the findings of the current study suggest that fit to the data is likely, but not conclusively acceptable. However, these findings correspond more closely with slightly older evaluations of the MYTS (Bickman et al., 2007, 2010). Differences in model fit across the studies may be due to differences in the underlying factor structure across populations of youth, but could also, and perhaps more likely, be due to differences in the estimation procedures used to fit the CFA models. Breda and Riemer (2012) used a typical maximum-likelihood (ML) estimation technique while we used a robust maximum likelihood technique (referred to as MLR in Mplus) to account for the non-normal distribution of item responses. When a robust method is applied compared to a typical ML approach, the fit indicators will favor the typical ML analysis although the parameter estimates (e.g., factor loadings) might be slightly biased (Yuan & Bentler, 2000).

One final area where the findings of the currently study are differentiated from other psychometric evaluations is in reporting the ratings scale functioning of the MYTS. This is the first study to report on the functioning of the rating scale as evaluated in the Rasch model – all prior evaluations used Rasch modeling to yield item and scale level indicators of psychometric quality, but no one has reported the findings related to the use and functioning of the rating scale. As reported above, the MYTS rating scale performed less than adequately, as a whole, and may be revised, by modifying response option anchors or by reducing the number of response options, to better align with Rasch measurement properties. However, it should be noted that the functioning of the rating scale did not seem to confound the other indicators of psychometric quality at the item or scale-level. In addition, all of the findings of the Rasch model, not just those pertaining to the rating scale, should be interpreted with care given the relatively small size of the sample under investigation.

Limitations and Future Research

Although the study procedures were rigorous and findings were triangulated across multiple analytic methods, there are a number of limitations worth noting. First, the sample was drawn from a single residential agency, which weakens the external validity (i.e., generalizability) of the findings. In addition, the sample was relatively small when compared to the sample sizes used for other psychometric evaluations of the MYTS (Bickman et al., 2007, 2010); however, for studies pertaining to residential group care populations, the sample size was relatively large for a single agency. A second limitation was that the MYTS was completed after admission to care (for some youth up to a month into care). Youth likely responded differently to many of the MYTS items given that they had already been admitted to a 24/7 residential care facility. Future research should address this issue to see if youths' motivation for treatment, problem recognition and readiness for treatment differ before and after admission to services, as well as during the course of residential services. The manner in which motivation changes over the course of treatment is beginning to be addressed in community-based treatment settings (Hawley & Garland, 2008), but has yet to be studied in residential group care and deserves attention given the potential importance of this developmental process.

A number of analytic limitations also represent a threat to the internal validity of the findings. The most salient was the inconclusive fit of the CFA model. Evaluating the goodness-of-fit of CFA models is a 'gestalt' process – there is no single value that indicates acceptable fit – so researchers report multiple fit indicators to triangulate findings. In this case, two indicators suggest acceptable fit (CFI and SRMR) while the other two indicators suggest slightly less than acceptable fit to the data (TLI and RMSEA). The indeterminate fit of the CFA model calls into question the adequacy of the factor structure supported in this study as well as the correlations between subscale constructs (i.e., the reduction in correlation magnitude compared to the raw data). However, given the similarities between the findings of this study and those from other psychometric evaluations of the MYTS, there is considerable evidence supporting the factorial validity of the MYTS. Future CFA studies might consider treating items as categorical indicators of the latent factors rather than as continuous indicators (see Flora & Curran, 2004 for a discussion of this topic). This approach would better account for the non-normal distribution of the items as well as other violations of interval scaling that might have led to the questionable model fit observed in this study.

Additional directions for future research might include a rigorous study of the convergent validity of the MYTS with residential group care populations. Given the discrepancy of the relationship between problem recognition and readiness for treatment across this study and Breda and Riemer (2012), it would be wise for researchers to investigate the convergent validity of the MYTS to collect evidence that the MYTS is measuring the constructs that it purports to measure. Future research should also investigate the motivation of the youth's caregiver (e.g., birth or foster parent, grandparent, etc.) via a parallel MYTS form for the caregiver. While this study was unable to include the perspectives of caregivers, including a caregiver assessment of motivation to participate in the youth's treatment can be important given their central role in the youth's intervention process, particularly in outpatient service settings (Nock & Ferriter, 2005; Nock & Photos, 2006).

Implications

Taking the psychometric findings in total, it appears that the MYTS is acceptable for use with youth in a residential care setting. The majority of the psychometric properties found in outpatient settings (Bickman et al., 2007, 2010; Breda & Riemer, 2012) were replicated with youth entering residential services. Nonetheless, there remains concern regarding the lack of

correlation between the two subscales, treatment readiness and recognition of problems, for youth in residential care, whereas the subscales are correlated for youth in outpatient settings (Breda &, Riemer, 2012). Thus, it is recommended that future research examine the functioning of the MYTS in other residential settings, to see if the findings of this study are replicated. Meanwhile, practitioners and researchers in residential care should be certain to examine the subscales, as they may provide a different perspective than the total score. It is possible that distinct subgroups of youth in residential care utilize the two subscales in varied manners, which future research could also help elucidate. The Rasch model indicated that the rating scale was used inconsistently across items, which may be due to such distinct groups (e.g., youth that believe they have a problem but do not want residential services versus youth that do not believe they have any problems but prefer residential services over alternate placements such as jail). The Rasch models may also be improved by reducing the number of response options from five down to three. Since the extreme response categories function as expected, one might consider collapsing the three middle response options into a single response option. This option holds promise as evidence from the psychometric evaluations of other measures used in residential group care tend to indicate that a 3- versus 5-point rating scale may function better in this setting (Duppong Hurley et al., 2012a; Lambert & Duppong Hurley, 2013).

These findings suggest that there may be subgroups of youth that respond to the problem recognition and treatment readiness subscales in different fashions. Future research will need to explore the possibility of distinct groups of youth with regard to motivation to change. For service providers, knowledge about an individual youth's problem recognition and readiness for treatment may help to tailor services. Understanding the youth's perspective on these constructs may help staff in their efforts to enhance the therapeutic alliance with youth, thus improving youth engagement in the intervention as well as subsequent youth outcomes. Examining the role of motivation to change in regard to youth outcomes is especially relevant in residential care which often finds that positive outcomes achieved while in care are not maintained upon discharge (Bettmann & Japerson, 2009).

In sum, the findings of this study largely replicated the psychometric properties of the MYTS with youth in outpatient populations to those in residential care. Due to lower correlations among the subscales in the residential population, it is recommended that researchers and practitioners focus on the problem recognition and treatment readiness subscales to understand youth's motivation to change in out-of-home settings. Additional research is needed to better understand how a youth's motivation for treatment impacts their experiences in residential care and their ultimate outcomes upon the completion of care. The finding that the MYTS is suitable for use in residential care might be especially important given the current impetus in the study of common therapeutic process factors in residential group care as well as community-based treatment for youth with emotional and behavioral issues (Bickman et al., 2004; Duppong Hurley et al., 2012a; Duppong Hurley et al., under review; Hawley & Garland, 2008; Manso, Rauktis, & Boyd, 2008), and the relationship between motivation for treatment and therapeutic alliance, in particular (Lambert et al., 2013), and the subsequent link between alliance and outcomes (Duppong Hurley et al., 2012b; Duppong Hurley et al., under review).

Acknowledgments

References

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington: University of Vermont, Research Center for Children, Youth, & Families; 2001.

Andrich D. A rating formulation for ordered response categories. Psychometrika. 1978; 43:561–573.

Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin. 1990; 107(2):238–246. [PubMed: 2320703]

Bettmann JE, Jasperson RA. Adolescents in residential and inpatient treatment: A review of the outcome literature. Child and Youth Care Forum. 2009; 38:161–183.

Bickman, L.; Riemer, M.; Lambert, EW.; Kelley, SD.; Breda, C.; Dew, SE.; Brannan, AM.; Vides de Andrade, AR., editors. Manual of the Peabody Treatment Progress Battery. Nashville, TN: Vanderbilt University; 2007. [Electronic version]

Bickman, L.; Athay, MM.; Riemer, M.; Lambert, EW.; Kelley, SD.; Breda, C.; Vides Andrade de, AR., editors. Manual of the Peabody Treatment Progress Battery. 2nd ed. Nashville, TN: Vanderbilt University; 2010. (Electronic version)

Bickman L, Vides de Andrade AR, Lambert EW, Doucette A, Sapyta J, Boyd AS, et al. Youth therapeutic alliance in intensive treatment settings. The Journal of Behavioral Health Services and Research. 2004; 31:134–148. [PubMed: 15255222]

Bond, TG.; Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Erlbaum; 2002.

Boomsma A. Reporting analyses of covariance structures. Structural Equation Modeling. 2000; 7:461–483.

Breda C, Riemer M. Motivation for Youth's Treatment Scale (MYTS): A new tool for measuring motivation among youths and their caregivers. Administration and Policy in Mental Health and Mental Health Services Research. 2012; 39(1):118–132. [PubMed: 22407559]

Broome KM, Joe GW, Simpson DD. Engagement models for adolescents in DATOS-A. Journal of Adolescent Research. 2001; 16(6):608–623.

Browne, MW.; Cudeck, R. Alternative ways of accessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park: Sage; 1993. p. 136-162.

Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: A meta-analysis of controlled clinical trials. Journal of Consulting and Clinical Psychology. 2003; 71:843–861. [PubMed: 14516234]

Cady M, Winters KC, Jordan DA, Solberg KR, Stinchfield RD. Measuring treatment readiness for adolescent drug abusers. Journal of Child and Adolescent Substance Abuse. 1996; 5:73–91.

Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed.. Hillsdale, NJ: Lawrence Earlbaum; 1988.

Davis, JL.; Daly, DL. Girls and Boys Town long-term residential program training manual. 4th ed.. Boys Town, NE: Father Flanagan's Boys' Home; 2003.

DeLeon, Hawke; Jainchill, Kressel; Melnick. 1999

DiClemente CC, Hughes SO. Stages of change profiles in alcoholism treatment. Journal of Substance Abuse. 1990; 2:217–235. [PubMed: 2136111]

DiClemente CC, Schlundt D, Gemmell L. Readiness and stages of change in addiction treatment. American Journal on Addictions. 2004; 13(2):103–119. [PubMed: 15204662]

Duppong Hurley, K.; Stevens, A.; Lambert, MC.; Sullivan, J. Bringing it all together: The relationships among implementation, therapeutic factors, and youth outcomes; Minneapolis, MN. A paper session presented at the annual conference for the American Evaluation Association; 2012.

Duppong Hurley K, Lambert MC, Van Ryzin M, Sullivan J, Stevens A. Therapeutic alliance between youth and staff in residential group care: Psychometrics of the therapeutic alliance quality scale. Children and Youth Services Review. 2013; 35(1):56–64. [PubMed: 23264715]

Edelen MO, Tucker JS, Wenzel SL, Paddock SM, Ebener PA, Dahl J, Mandell W. Treatment process in the therapeutic community: Associations with retention and outcomes among adolescent residential clients. Journal of Substance Abuse Treatment. 2007; 32:415–421. [PubMed: 17481465]

Flora D, Curran P. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods. 2004; 9:466–491. [PubMed: 15598100]

Harder AT, Knorth EJ, Kalverboer ME. A secure base? The adolescent-staff relationship in secure residential youth care. Child & Family Social Work. 2012 lAdvance online publication.

Hawley KM, Garland AF. Working alliance in adolescent outpatient therapy: Youth, parent and therapist reports and associations with therapy outcomes. Child and Youth Care Forum. 2008; 37:59–74.

Hiller ML, Knight K, Leukefeld C, Simpson DD. Motivation as a predictor of therapeutic engagement in mandated residential substance abuse treatment. Criminal Justice and Behavior. 2002; 29(1):56–75.

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6(1):1–55.

Kelly JF, Urbanoski KA, Hoeppner BB, Slaymaker V. Ready, willing, and (not) able" to change: Young adults' response to residential treatment. Drug and Alcohol Dependence. 2012; 121:224–230. [PubMed: 21955361]

Kline, P. The handbook of psychological testing. 2nd ed.. London, UK: Routledge; 1999.

Lambert MC, Duppong Hurley K. Measurement properties of the Symptoms and Functioning Severity Scale with a residential group care population. Journal of Behavioral and Emotional Disorders. 2013 Under review at the.

Lambert, MC.; Duppong Hurley, K.; Thompson, R.; Pick, R. Therapeutic alliance: A mechanism through which motivation to change impacts outcomes for youth in residential care; Tampa, FL. Paper session presented at the Children's Mental Health Research and Policy Conference; 2013.

Laurier C, Lafortune D, Collin J. Compliance with psychotropic medication treatment among adolescents living in youth care centres. Children and Youth Services Review. 2010; 32:67–73.

Lawley, DN.; Maxwell, AE. Factor analysis as a statistical method. 2nd ed.. London, UK: Butterworths; 1971.

Linacre JM. Optimizing rating scale category effectiveness. Journal of Applied Measurement. 2002; 3:85–106. [PubMed: 11997586]

Linacre, JM. Misfit diagnosis: Infit outfit mean-square standardized. 2006. Retrieved from http://www.winsteps.com/winman/diagnosingmisfit.htm

Linacre, JM. Winsteps® Rasch measurement computer program. Beaverton, OR: Winsteps.com; 2012a.

Linacre, JM. Winsteps® Rasch measurement computer program user's guide. Beaverton, OR: Winsteps.com; 2012b.

Little TD, Slegers DW, Card NA. A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. Structural Equation Modeling. 2006; 13:59–72.

Manso A, Rauktis ME, Boyd AS. Youth expectations about therapeutic alliance in a residential setting. Residential Treatment for Children and Youth. 2008; 25(1):55–72.

Melnick G, De Leon G, Hawke J, Jainchill N, Kressel D. Motivation and readiness for therapeutic community treatment among adolescents and adult substance abusers. American Journal of Drug and Alcohol Abuse. 1997; 23(4):485–506. [PubMed: 9366969]

Muthén, LK.; Muthén, BO. Mplus user's guide. 6th ed.. Los Angeles, CA: Authors; 1998–2012.

Nock MK, Ferriter C. Parent management of attendance and adherence in child and adolescent therapy: A conceptual and empirical review. Clinical Child and Family Psychology Review. 2005; 8:149–166. [PubMed: 15984084]

Nock MK, Photos VI. Parent motivation to participate in treatment: Assessment and prediction of subsequent participation. Journal of Child and Family Studies. 2006; 15(3):345–358.

Riemer M, Athay MM, Bickman L, Breda C, Kelley SD, Vides de Andrade AR. The Peabody Treatment Progress Battery: History and methods for developing a comprehensive measurement battery for youth mental health. Administration and Policy in Mental Health. 2012; 39:3–12. [PubMed: 22421933]

Ryan RM, Plant RW, O'Malley S. Initial motivations for alcohol treatment: Relations with patient characteristics, treatment involvement, and dropout. Addictive Behaviors. 1995; 20(3):279–297. [PubMed: 7653312]

Satorra A, Bentler PM. Ensuring positiveness of the scaled difference chi-square test statistic. Psychometrika. 2010; 75:243–248. [PubMed: 20640194]

Schroder R, Sellman D, Frampton C, Deering D. Youth retention: Factors associated with treatment drop-out from youth alcohol and other drug treatment. Drug and Alcohol Review. 2009; 28:663–668. [PubMed: 19930021]

Shaffer D, Fisher P, Lucas C, Dulcan M, Schwab-Stone M. NIMH Diagnostic Interview Schedule for Children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry. 2000; 39:28–38. [PubMed: 10638065]

Steiger, JH.; Lind, JC. Statistically-based tests for the number of common factors; Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City; May 30. 1980

Tucker L, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973; 38(1):1–10.

Wolf MM, Phillips E, Fixsen D, Baukmann CJ, Kirigin KA, Willner AG. Achievement place: The teaching-family model. Child Care Quarterly. 1976; 5:92–103.

Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Measurement Transactions. 1994; (l8)(3):370.

Yuan K-H, Bentler PM. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. Sociological Methodology. 2000; 30:165–200.
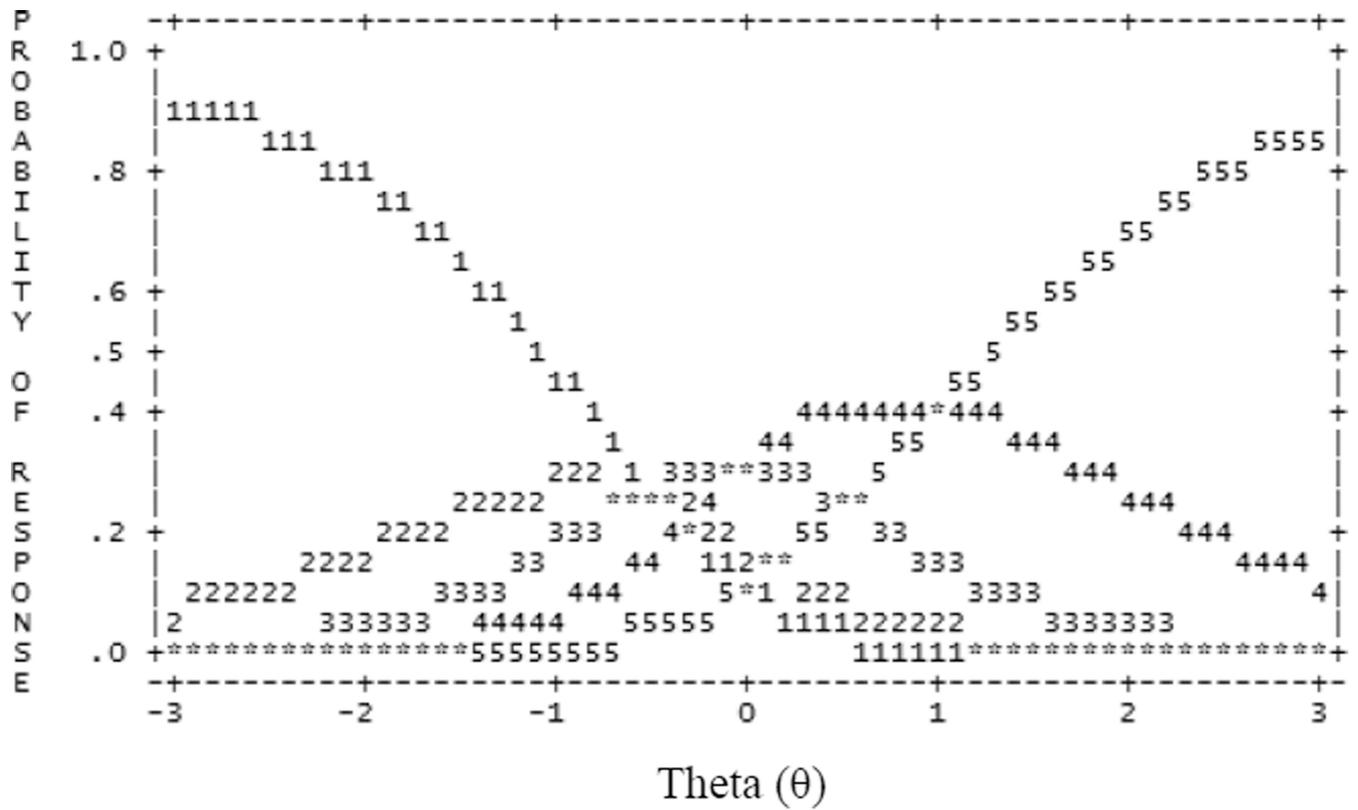
Figure 1. Rating Category Characteristic Curve Plot
Category Probabilities: Modes – Structure Measures at Intersections (1 = *Not at all*, 2 = *Only a little*, 3 = *Somewhat*, 4 = *Quite a bit*, 5 = *Totally*)

Table 1

Item-Level Descriptive Statistics

	M	SD	Skew	Kurtosis	Measure	Disc.	Infit	Outfit
Problem Recognition								
My behavior is causing problems at home, school, with my friends, or in other places	2.88	1.32	−0.19	−1.32	0.30	1.28	0.87	0.83	.89
My behavior is making my life worse	2.93	1.34	−0.17	−1.25	0.26	1.27	0.89	0.85	.80
Some of my feelings are really bothering me	2.60	1.24	0.30	−0.85	0.55	**0.55**	1.15	1.19	.52
My feelings are causing problems at home, school, with my friends, or in other places	2.61	1.31	0.30	−1.06	0.56	1.18	0.83	0.80	.74
Treatment Readiness								
I need help to deal with my problems	4.04	1.00	−0.97	0.49	−0.90	1.04	0.91	0.88	.44
Coming to _____ seems like a good idea to me	3.63	1.29	−0.72	−0.42	−0.41	0.99	1.09	1.02	.93
I want help finding solutions for my current problems	3.96	1.18	−1.00	0.13	−0.78	0.90	1.21	1.10	.86
I am coming to _____ because I want to	2.76	1.40	0.18	−1.22	0.42	0.78	1.18	1.19	.65

Note. Factor loadings are reported for the 2-factor MLR model; Disc. = Discrimination

Table 2

CFA Model Fit Indicators

	$X^2_{(df)}$	X^2	CFI	TLI	RMSEA [95% CI]	SRMR
1-Factor	266.17$_{(20)}$	–	0.40	0.16	0.29 [.26, .32]	0.20
2-Factor	53.21$_{(19)}$	122.7*	0.92	0.88	0.11 [.07, .15]	0.07

Note. X^2 was calculated as the Satorra-Bentler scaled chi-square difference test (Muthén & Muthén, 1998–2012). Significant X^2 indicates that the fit of the two models differ significantly.