3-2014

Refining the Evidence-Based Practice Attitude Scale (EBPAS): An Alternative Confirmatory Factor Analysis

David A. Patterson Silver Wolf (Adelv unegv Waya) PhD
Washington University in St Louis, Brown School, dpatterson22@wustl.edu

Catherine N. Dulmus PhD
University at Buffalo, SUNY, Buffalo Center for Social Research

Eugene Maguin PhD
University at Buffalo, SUNY, Buffalo Center for Social Research

Nicole Fava
University at Buffalo, SUNY, School of Social Work

Refining the Evidence-Based Practice Attitude Scale (EBPAS): An Alternative Confirmatory Factor Analysis

Evidence-based practices (EBP) are considered the gold standard in human service organizations with funders, as well as empowered consumers, who often demand them. Tax-payer supported grants and program budgets should not support unproven practices and consumers should offered services that have been scientifically evaluated. However, introducing and integrating an EBP into an agency setting can engender substantial costs related to training, supervision and developer consultation. The history of EBPs in social work, along with the prevalence and magnitude of EBPs in practice settings, has been thoroughly reviewed by Thyer and Myers. (see Thyer & Myers, 2011). Therefore, it is important to consider the different factors that may contribute to the effectiveness and successful implementation of EBPs into agency practice. A growing literature suggests that community-based mental health organizational factors and individual worker-level factors affect decisions about whether or not health care agencies implement EBPs (Aaron, 2005; Glisson & Hemmelgarn, 1998; Glisson & James, 2002; Hemmelgarn et al., 2001; Hemmelgarn et al., 2006). Some possible explanations for why workers are not implementing EBPs in practice consist of worker's years of work experience (Aarons, 2004; Pignotti & Thyer, 2009), educational attainment (Aarons, 2004; Ogborne et al., 1998; Stahmer & Aarons, 2009), educational discipline (Aarons, 2004; Stahmer & Aarons, 2009), and students completing an internship (Aarons, 2004; Garland, 2003). All of these investigations suggested several factors that contribute to the successful implementation of EBPs. One factor, in particular, interesting for further investigation is the attitudes of the workers. The workers' opinions and beliefs regarding EBPs can greatly influence their utilization and implementation of new EBPs (Aarons, 2004; Aarons, McDonald, Sheehan, & Walrath-Greene, 2007; Aarons et al., 2010).

Evidence-Based Practice Attitudes Scale Development

Based on the literature regarding the implementation of evidence based practices related to worker attitudes, Aarons (2004) identified four factors as influential inworkers' attitudes towards the acceptance and use of EBPs: 1) openness to implementing new interventions (Openness); 2) the intuitive appeal of the new intervention (Appeal); 3) willingness to using required interventions (Requirements); and 4) conflict between clinical experience and research results (Divergence). An initial pool of 18 items to measure these four constructs, developed from the literature and from consultation with workers and researchers, was reduced to the 15 items in the final measure (see Table 1).

Three studies (Aarons, 2004; Aarons, McDonald, Sheehan, & Walrath-Greene, 2007; Aarons, Glisson, Hoagwood, Kelleher, Landsverk, & Cafri, 2010), summarized in Table 2, used confirmatory factor analysis (CFA) to evaluate the measurement structure of the EBPAS in samples of child and adolescent mental health workers. Each of these studies reported a significant chi-square value and was judged to have adequate model fit by comparing the values of the CFI (Confirmatory Fit Index), TLI (Tucker-Lewis Index), RMSEA (Root Mean Square Error Of Approximation), and SRMR (Standardized Root Mean Square Residual) to the criterion for determining goodness of fit that existed at the time of the analysis (Bollen, 1989). Since that time, the criterion used to determine goodness of fit have become more stringent (e.g., Schermelleh-Engel, Moosbrugger, & Muller, 2003; Marsh, Hau, & Wen, 2004). The new standards would render the model fit found in the three previous studies marginal to poor. Our review of the EBPAS items suggests that further scale development is warranted in order to better measure the identified constructs. For example, the Divergence scale consists of two negatively-worded items (items 5 and 7) and two positively-worded items (items 3 and 6). Exploratory factor analyses of theoretically unidimensional scales with a combination of negatively and positively worded items, such as the Rosenberg Self-Esteem Scale (Rosenberg, 1965), often have yielded a factor of positive items and a factor of negative items (see Marsh, Scalas, & Nagengast, 2010 for a review). Furthermore, items 5 ("Research-based therapies are not clinically useful") and 7 ("Won't use manualized therapy/interventions") are more logically and semantically similar to each other than to items 3 ("Know better than researchers how to care for clients") and 6 ("Clinical experience more important than manualized interventions"), which, in turn, are more similar to each other.

We think the quantitative results also support the general conclusion that the Divergence scale needs further development. Item-total correlations for the Divergence scale were between .32 and .43 for both Aarons (2004) and Aarons et al. (2007), implying that none of the four items had a strong relationship with the total scale score (Aarons et al. [2010] did not report these data). The factor loadings for Divergence seem to show an unusual pattern, as the standardized loadings for items 5 and 7 are numerically larger than those for items 3 and 6. The pattern is most clear in Aarons (2004) (.76 and .65 versus .34 and .42) and Aarons et al. (2007) (.70 and .68 versus .43 and .50) but less so for Aarons et al. (2010) (.68 and .57 versus .55 and .49). These factor loadings imply fairly high correlations between the two high-loading items, low correlations between the two low-loading items and, probably, moderate correlations between the two sets of items. Unfortunately, none of the articles reported the item covariance matrix.

Within the context of a CFA, we think the evidence suggests two alternative models. The first model involves the addition of a residual covariance between either items 5 and 7 or items 3 and 6 to account for the positive-negative wording or meaning effects (see Marsh, Scalas, & Nagengast, 2010 for a review related models). The second model involves splitting Divergence into two factors, one factor defined by items 3 and 6 and the second by items 5 and 7. Although two item factors are not desirable, we think there is logical and empirical support for such a model. To the extent that this model is supported, it also highlights the need for further scale development.

Within Organization Measurement Invariance

A remaining unexamined question is whether attitude scores or evidence-based practices attitude structures differ among clinicians working in different types of programs, or with different client populations, or in different organizational structures. Although the three Aarons et al. studies (Aarons, 2004; Aarons, et al., 2007; Aarons, et al., 2010) sampled community child mental health agencies, the degree of program variety across the agencies is unknown. While Table 2 shows variation in the factor loadings, corresponding loadings are numerically similar across studies, excluding items 9 and 10 where a residual covariance was added. A comparison of scale mean score indicates larger variation across studies for the Requirements and Openness scales (*ES* = .29 and .36, respectively) and smaller variation for the Appeal and Divergence scales (*ES* = .13 and .17, respectively).

Purpose of Study

The purpose of the present study is to investigate the factor structure of the EBPAS using data from 'front-line' employees of a single, large private, not-for-profit child and family service agency providing a wide range of programs in both residential and community settings. We will use CFA to replicate the four-factor model previously investigated (Aarons, et al., 2007; Aarons, et al., 2010) and compare to that model two alternative models: (a) a model that adds a residual covariance between either items 5 and 7 or items 3 and 6 and (b) a model that splits the current Divergence scale into two 2-item factors (items 5 and 7 and items 3 and 6). Using the best-fitting model, we then will test to what extent the measurement structure of EBPAS and the factor means, variances and covariances are the same for clinicians in two different types of programs by means of a multiple group analysis.

Method

Setting

The setting for this study was Hillside Family of Agencies (HFA), the largest child and family human service agency in Western and Central New York State (NYS). HFA has helped children and their families for more than 170 years, and currently employs more than 2,200 staff within six affiliate organizations located in 40 sites across 30 New York counties and in Prince George’s County, Maryland. Affiliates of this $140+ million network provide services to children from birth to age 26 in more than 9,000 families each year. HFA provides 120 services in six major categories including child welfare, mental health, juvenile justice, education, youth development, and developmental disabilities/mental health. HFA is accredited by the Council on Accreditation and holds NYS licenses from the Office of Children and Family Services, Office of Mental Health, Office for People with Developmental Disabilities, Department of Health, and State Education Department (M. Cristalli, personal communication, March 18, 2010).

Study Sample

All participants in this study were 'front-line' employees (i.e., those employees having direct contact with the children and families served by this agency) from the four affiliates that provide direct services. Given this criterion, participants represented a number of different work roles in the agency including, but not limited to: Direct care workers in residential settings, therapists, and mentors. However, we did not ask participants to identify their work roles nor did we ask the agency to assemble workers by work role. The rate of participation for this study was over 80%, yielding a total sample of 1,273 child and family front-line workers from 55 workgroups. The number of participants per workgroup ranged from 3 to 84 (Median = 15, *M* = 23, *SD* = 19). About 75% of the sample worked at one affiliate, about 13% at the second, around 8% at the third and about 5% at the fourth. Approximately 42% of respondents worked in residential programs, 22% worked in community-based programs, 12% worked in school-based programs that were part of a residential facility, and 11% worked in day treatment programs. The remaining respondents worked in four much smaller services.

The final sample of participants had a mean age of 35 years (*SD* = 11; range: 19-73), 59% were female, and 74% self-identified as white, 17% as African American and 5% or less for any other category (multiple categories were allowed). At the time of this survey's administration, participants had worked in the human service field for a mean of 9.6 years (*SD* = 8.5; range: 0-50; median = 7) and at their current agency for an average of 5 years (*SD* = 5.6; range: 0-36; median = 3). Seventeen percent had completed high school, 17% had earned an associate's degree, 38% had received a bachelor's degree, 27% had obtained a master's degree, and only 1% had earned a doctoral degree. The predominant discipline for participants' degrees was education (23%), followed by social work (18%), psychology (16%), nursing (4%), and medicine (0.4%); the category of "other" made up for the bulk of the distribution (39%); however, we were not able to determine the contributing disciplines.

Measure

Worker attitudes were measured using the Evidence Based Practice Attitudes Scale (EBPAS: Aarons, 2004). The EBPAS consists of 15 items rated on a 5-point Likert scale, ranging from 0 (*not at all)* to 4 (*to a very great extent*). Previous reports on the EBPAS indicate an acceptable overall reliability ranging from .76 to .79 and subscale reliabilities ranging from .59 to .93 (Aarons, 2004; Aarons et al., 2007; Aarons et al., 2010).

Data Collection Procedure

Upon IRB approval, the EBPAS and a companion measure that included a set of demographic questions were administered to participants in paper and pencil format. The EBPAS measure and the companion measure were linked at the workgroup level only. Data collection occurred in groups, with no agency administration present. Each group was read instructions assuring subjects that their responses were anonymous and data would only be reported back to the organization in aggregated form. All subjects were volunteers, signed informed consent, and received no compensation.

Statistical Analyses

The CFA models were estimated using Mplus 6.12 (Muthen & Muthen, 2010). Since respondents worked together in organizationally distinct groups, the data were clustered and the Mplus Complex option was used to adjust the chi-square value and model standard errors for the effects of clustering. EBPAS items were modeled as categorical to account for the ordinal nature of the response scale. As a result, the default WLSMV estimator was used. Cases with missing data were included since the WLSMV estimator permits the inclusion of cases with missing data. All analyses use the default delta formulation, rather than the alternative theta formulation. In the delta formulation, the latent residual variances are fixed at 1.00 and the item thresholds and scale factors are free to be estimated.

As Hu and Bentler (1998, 1999) recommended, multiple indices summarized the model fit. In addition to the chi-square test statistic, the approximate fit indices provided in Mplus (CFI, TLI, RMSEA, and WRMR [Weighted Root Mean Square Residual], due to analyzing the data as categorical) were used. Although Hu and Bentler (1999) and Schermelleh-Engel, et al. (2003) have recommended higher threshold values for fit indices, these threshold values are based on continuous data. To our knowledge, only Yu (2002) has studied this question for categorical models and she recommended .95 for CFI and TLI, .05 for RMSEA, and 1.00 for WRMR. Lastly, a number of authors (e.g., Schreiber, Nora, Stage, Barlow, & King, 2006) have recommended plotting model residuals and examining modification indices.

Nested models were compared by means of the chi-square difference test for the WLSMV estimator (Asparouhov & Muthen, 2006). Although the chi-square difference (likelihood ratio) test has long been used to compare nested models (e.g., Joreskog, 1971), a little-appreciated assumption is that the less restrictive model is correctly specified, as a non-significant chi-square value shows (Chou & Bentler, 1990; Yuan & Bentler, 2004). A significant chi-square for the baseline model, as is expected here, reduces the power of the chi-square difference test (Yuan & Bentler, 2004). Although several authors, (e.g., Cheung & Rensvold, 2002) have investigated the use of approximate fit indices to compare tests of measurement invariance in multiple group models, to our knowledge the applicability of their work to nested comparisons of single group models or models for categorical is not known.

Multiple group model comparisons were made following the test sequence that Muthen and Muthen (2011) recommended, which differs from the sequence Millsap and Yun-Tien (2004) described. After estimating the model in each group separately, the model was then estimated in both groups simultaneously as the baseline analysis. In the baseline analysis, factor means were fixed at 0.0 in both groups and factor loadings; scale factors and thresholds were free to vary. The next analysis, which tested measurement invariance (the measurement invariance model), freed the scale factors in both groups and the factor means in the second group, and constrained the factor loadings and item thresholds to equality across the two groups. A chi-square difference test was performed to test whether the measurement invariance model differed significantly from the baseline model. If the difference test was significant, modification indices were examined to identify the source(s) of the invariance. Invariance of the structural parameters was tested by constraining separately the variances, the covariances, and the means to equality in the measurement model and then comparing each constraint set to the unconstrained measurement invariance model.

Results

Of the 1,273 participants in this study, 1,190 had complete data for all 15 items on the EBPAS, another 70 had missing data for one or more items and 13 chose not to complete the EBPAS at all. Thus, usable data were available for 1,260 participants. Covariance coverage ranged between 96.8% and 99.8%. Item distributions tended to be skewed towards the favorable end of the response scale and to range between somewhat flattened (i.e., less peaked) and overly peaked; however, both skewness values (-0.92 to 0.01) and kurtosis values (-0.62 to 0.76) were small for all items. Except for Items 5, 7, 13 and 15, which had “J” shaped distributions, the remaining items all had interior modes. Table 3 presents the response category proportions, item means (data treated as continuous) and correlations, computed as polyconics, and the ICCs. The ICCs ranged from .016 to .059. The magnitudes of the ICCs indicate that the workgroup mean item scores vary only slightly across workgroups and the majority of variation is among workers.

Model Comparisons

Table 4 presents the results for the Aarons, et al. (2007) and the two alternative models: Added Covariance (AC) and Five Factor (FF). The AC model added a residual covariance between the positively-worded items 3 and 6 of the Divergence factor. (We also estimated a model that added a residual covariance between items 5 and 7, the negatively-worded items of the Divergence factor, but found that the estimation yielded an inadmissible solution because the estimated standardized residual covariance, i.e., correlation, between the two items was greater than 1.00). The FF model divided the Divergence factor into 2 two-item factors, items 3 and 6, labeled Clinical, and items 5 and 7 labeled Research. As the fit statistics data in Table 4 shows, while all three models had CFI and TLI values above the .95 value recommended by Yu (2002), each model had a significant chi-square value at $p < .001$ and WRMR values that were considerably over the 1.00 value recommended by Yu (2002). Rounded to two significant digits, all RMSEA values were at the .05 recommended value, with both the Aarons, et al. (2007) and the AC models slightly above .05 and the FF model slightly below.

Comparing the factor loading coefficients shows that values for items loading on the Requirements, Appeal, and Openness factors are within +/-.001 of each other across the three models. The residual covariance added between items 3 and 6 was significant ($r = .160$, $p < .001$) and this addition resulted in numerically smaller loadings compared to those for the Aarons, et al. (2007) model. Dividing the Divergence factor into two factors resulted in larger values (about 30% and 50%) for items 3 and 6 but smaller values (about 2%) for items 5 and 7 compared to the Aarons, et al. (2007) model.

The Aarons, et al. (2007) model is nested within both the AC and FF models. Comparing the AC model to the Aarons et al. model demonstrated that constraining the residual covariance between items 3 and 6 to 0.0 resulted in a significant difference test value ($\chi^2[1] = 22.253$, $p < .001$) and indicated that the AC model is a better fitting model. Constraining the FF model to the

Aarons et al. model also resulted in a significant difference test value ($\chi^2[4] = 39.669$, $p < .001$) and indicated that the FF model is also a better fitting model.

Since the FF model is not nested within the AC model, these two models cannot be directly compared by a chi-squared difference test. When the data are modeled as continuous, information measures (BIC, AIC) offer an alternative, but these measures are not available for categorical data. Comparing the approximate fit indices (RMSEA, CFI, and TLI) in Table 4 shows them to be numerically smaller for the FF model than for the AC model. However, the differences occur in the third decimal point. While the WRMR value for the FF model is about 8% lower than that for the AC model, the substantive meaning is not known.

Table 5 reports the factor correlations and variances for the three models being compared. Except for two, all variances and correlations are significant. Adding the residual covariance in the AC model decreases the variance of the Divergence factor by about 15% and increases the correlations between Divergence and the other three factors by less than .01. It is worth noting that the variance of the Divergence factor, irrespective of the specific four-factor model, is between about one-third and one-fifth of the variance of the other factors. The effect of the splitting the Divergence factor shows that, while Clinical and Research are strongly correlated ($r = .771$), these two new factors do not have similar relationships with the other three factors, and, in fact, Clinical has small, non-significant correlations with both Appeal and Openness. In contrast, Research has moderate, similarly-sized, significant correlations with both Appeal and Openness (r = -.397 and -.325, respectively). Both Clinical and Openness have a commonality in the Requirements factor, a small, similarly-sized, significant correlation with Requirements (r = -.192 and -.168, respectively). Lastly, the variance of Research is similar to three original factors(Requirements, Appeal, and Openness), while Clinical shows a numerically much smaller variance.

Unstandardized residuals were examined since standardized residuals are not available with the WLSMV estimator. Across all models examined, residuals were normally distributed with a mean of from -0.0012 (*SD* = 0.0449) for the five factor model to a mean of 0.0050 (*SD* = 0.0501) for the Aarons, et al. (2007) model. Depending on the model, two to four residuals fell outside the 1.5 interquartile range on a box plot. A review of modification indices (MIs) with values greater than 10 showed between 18 to 23 possible cross-loadings, one of which had a standardized expected parameter change value greater than 0.30, and from 12 to 14 possible residual covariances. Aarons (2004) reported using a threshold of .30 to select items in the initial factor analysis and it would not be surprising if small cross loadings or residual covariances were present for some of the items and to be identified by the modification indices.

Summary. The comparisons of the two alternative models showed that both were improvements, in terms of fit, over the Aarons, et al. (2007) model. However, neither alternative model can be shown to be significantly better than the other. The five factor model makes clear that items 3 and 6, taken as a factor, and items 5 and 7, taken as a factor, are not parallel measures of each other. Otherwise, the two factors would have similar relationships with both Openness and Appeal, which they do not.

Between Group Comparisons

We chose to compare the responses of workers in community-based and residential programs. We think the program structure and client populations offer interesting differences. The other option of comparing affiliates was less appealing because programs are partially crossed with affiliates; multiple affiliates provide the same type of program. In addition, one affiliate accounts for about 75% of respondents and a large majority of programs.

As reported earlier, about half (*n* = 540: 42%) of respondents worked in residential programs and about one-quarter (*n* = 282: 22%) worked in community-based programs. While about 12% of respondents worked in school-based programs in residential facilities, these workers were judged to be programmatically distinct from other workers in the residential facilities and, thus, we did not combine them with the other residential workgroups. In addition, community-based and residential have 17 and 18 workgroups, respectively, a number we judged sufficient to permit an accounting for clustering.

A comparison of the demographic data for respondents working in community based or residential programs found that while the mean and median years of experience did not differ, community based workers were significantly older (M = 35.2 vs. 33.4, $t(751) = 2.27$, $p = .024$) but had fewer years of experience at Hillside (M = 3.75 vs. 4.70, $t(768) = 2.44$, $p = .015$). Analyses of the medians showed the same results. Community based workers were significantly better educated as a higher percentage had either a bachelors (49.4% vs. 37.3%) or masters degree (29.1% vs. 13.3%) rather than a high school diploma (5.7% vs. 29.4: $X^2[4, n = 760] = 79.28$, $p < .001$); significantly more likely to be African American (26.4% vs. 19.1%: $X^2[1, n = 798] = 5.31$, $p = .021$); and significantly more likely to be female (73.0% vs. 48.5%: $X^2[1, n = 764] = 41.23$, $p < .001$). We chose to treat the two alternative models as equal, but different, and planned to examine how each functioned in a multiple group comparison analysis. However, we found that the Five Factor model yielded an inadmissible solution in the form of a correlation greater than 1.00. We present results for the AC model only.

Added Covariance (AC) Model. The fit statistics for the baseline model were $\chi^2(164) = 309.844$, $p < .001$; RMSEA = 0.047; CFI = 0.983; TLI = 0.978; and WRMR = 1.391. Each group contributed nearly equally to the overall chi-square value. Comparing the factor loadings for the two program groups showed that most coefficients differed by 5% to 15% between the two groups—the largest differences were 50% for item 6 and 25% for item 14, with no trend favoring one program. After setting the constrained and freed parameters for the measurement invariance model (Neither the residual covariance for items 9 and 10 nor that for items 3 and 6 were constrained equal between groups), the chi-square difference test of the measurement invariance model to the baseline model was not significant ($\chi^2[52] = 53.884$, $p = .402$). As expected, inspection of the modification indices revealed no problems with thresholds and no large differences in factor-by-item indices compared to the baseline model. However, the residual covariances between items 9 and 10 (Community: $r = .552$; Residential: $r = .314$) and between items 3 and 6 (Community: $r = .031$; Residential: $r = .200$) did appear to differ between the groups and this was confirmed by subsequent chi-square difference tests: $\chi^2(1) = 10.799$, $p = .001$ for items 9 and 10 and $\chi^2(1) = 6.127$, $p = .013$ for items 3 and 6.

Table 6 reports the factor correlations, variances, and unstandardized means for the two groups as estimated by a model that did not constrain the (9, 10) and (3, 6) residual covariances to equality between groups. Except for the Requirements factor, factor variances were similar between the two groups and, as a set, the four factor variances did not differ between groups [$\chi^2(4) = 4.307$, $p = .366$]. The Requirements factor variance, which had the largest numerical difference (Community: 0.751, Residential: 0.954), also did not differ when tested alone. Except for the Requirements-Divergence correlation, the remaining factor correlations for the residential group were numerically larger than those for the community group by about 10% to 400%. The correlation with the largest difference was the Openness-Divergence correlation (community: -.098; residential: -.396). As a set, the six factor covariances did differ between groups [$\chi^2(6) = 16.336$, $p = .012$]; however, effect is attributable to the Openness-Divergence correlation difference as the remaining five correlations, taken as a set, did not differ between groups [$\chi^2(5) = 9.728$, $p = .083$]. Looking lastly at the factor means, with the community group's means fixed at 0.0, the residential group's means for Requirements ($M = 0.298$) and Openness ($M = 0.336$) respectively, differed significantly at $p < .001$ from the community group's means for those factors.

In summary, measurement invariance was demonstrated between community-based and residential programs but structural invariance was not. Differences between the two groups were found for the Openness-Divergence correlation and the means for the Requirements and Openness factors.

Discussion

The attitudes of workers are an important component that influence the implementation and utilization of EBPs (Aarons, 2004; Aarons et al., 2007; Aarons et al., 2010). Assessment of those attitudes is largely dependent on a new assessment tool, the EBPAS, which has undergone limited psychometric evaluation. One purpose of this study was to replicate the previously reported four-factor structure for the EBPAS and then to examine two alternative models based on our analysis of the items in the current Divergence factor, which consists of two positively-worded items and two negatively-worded items. One of the alternative models used a residual covariance to model wording effects while the other used two factors. The second purpose was to compare the measurement structure and the factor variances, means and correlations for clinical employees of community-based programs and residential programs located in a single, large child and family services agency.

We expected to replicate the current published EPAS model (Aarons et al., 2007; Aarons et al., 2010) and we did, albeit with slightly better approximate fit statistics than previous analyses had reported. We found the pattern identified in prior analysis in our factor loadings for the Divergence factor. More importantly, we found that adding a residual covariance between the two positively worded Divergence items significantly improved the fit. We also found that splitting the current Divergence scale into 2 two-item factors also significantly improved the fit.

Although we were unable to choose one model over the other on statistical grounds, the fact that both models showed an improved fit indicates that any revisions to the EBPAS measure need to consider carefully the development of questions about the relative and absolute values of clinical experience and research-derived knowledge. For the purpose of measure development only, we think the results of the five factor model, made by splitting the Divergence scale into two factors, are more important. This is because the results show that the two positively-worded items, as items and as a factor, Clinical Knowledge, are either uncorrelated or weakly correlated with other items and with the Appeals and Openness scales. A pool of new items better assessing how clinical knowledge informs treatment decisions may be useful. The two negatively-worded items, the Research factor, might well form a core around writing three to five new items that assess the value of research-derived knowledge in treatment decisions, with a goal of having at least three, but preferably four or five, items in the factor (Kline, 2005).

We found that the alternative four factor model supported measurement invariance, but we were unable to evaluate the five factor model. The results for the alternative four factor model show that both community-based and residential workers shared the same measurement structure, which then made comparisons between the groups meaningful. Both groups yielded approximately the same correlations between factors, with one exception. Both residential and community-based workers who rejected research-derived knowledge and interventions also rated themselves as less open to new therapies from different sources; however, the relationship was considerably stronger for residential workers than for community-based workers. Comparisons of the factor means indicated that residential workers who were more responsive to requirements used evidence-based practices and that they were more open to using them than community-based workers.

Because the analysis took into account the clustering of employees into workgroup, ICCs were computed and these offer insight into the location of the variability in evidence base practice attitudes. The small values, almost all less than .05 and similar in magnitude to prior studies, indicate the variability between persons more than between workgroups. Thus, any efforts to influence attitudes would seem to need to be addressed to individual employees across most workgroups, rather than all employees in some workgroups.

Although this study achieved a response rate of over 80%, which resulted in a robust sample size of 1,273 participants, we do not know how workers who participated compared to those who did not. All of our participants were from the same organization, which might result in a greater similarity of attitudes than if participants had come from multiple organizations. However, the work groups themselves are dispersed over all of western and central New York State and we think that this dispersal would minimize, to some extent, the equalization of EBP attitudes compared to a situation in which work groups shared a common location. At the same time, the participating agency may have provided a broader range of services than at least some of the agencies had in the prior studies. Because we were unable to link respondent demographics from the companion measure to the respondent's EBPAS responses, we were unable to examine how participants' demographic characteristics related to their attitudes towards evidence-based practice, both at the factor level, to understand how demographic differences related to factor variation and at the item level, to understand how demographic differences related to differential responding to the items in a factor.

While the results of this study are useful for investigators using the EBPAS in studies, what can agency managers take away from our results? We think the most important point is that, while attitudes about evidence-based practices are likely to be similar from workgroup to workgroup, some attitudes, represented by Divergence and Appeal, may vary less and those represented by Requirements and Openness may vary more. Although this study has identified a problem with the Divergence scales and how that scale might be improved, we doubt that there is anything for managers to gain by scoring the Divergence scale differently. However, managers may be interested to know that Aarons and colleagues have added subscales of important worker attitudes to the original measure (Aarons et al., 2010b). Although we think our concerns about the Divergence scale remain, the additional scales may be useful to provide a broader and more nuanced picture of attitudes about evidence-based practices.

References

Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61-74.

Aaron, G.A. (2005). Measuring provider attitudes toward evidence-based practice: Organizational context and individual differences. *Child and Adolescent Psychiatric Clinic of North America, 14*, 255-271.

Aarons, G. A., Cafro, G., Lugo, L., & Sawitzky, A. (2010b). Expanding the domains of attitudes towards evidence-based practice: The Evidence Based Practice Attitude Scale-50. *Administration and Policy in Mental Health*.

Aarons, G. A., Glisson, C., Hoagwood, K., Kelleher, K., Landsverk, J., & Cafri, G. (2010). Psychometric properties and U.S. national norms of the Evidence-Based Practice Attitude scale (EBPAS). *Psychological Assessment, 22*(2), 356-365.

Aarons, G. A., McDonald, E. J., Sheehan, A. K., & Walrath-Greene, C. M. (2007). Confirmatory factor analysis of the Evidence-Based Practice Attitude Scale (EBPAS) in a geographically diverse sample of community mental health providers. *Administration and Policy in Mental Health, 34*, 465-469.

Asparouhov, T. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics*. Retrieved April 2, 2012 from http://statmodel.com/download/webnotes/webnote10.pdf.

Bollen, K. A. (1989). *Structural equation models with latent variables*. New York: Wiley.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 235-255. u, L., & Bentler, P.

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

Chou, C.P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research, 25*, 115-136.

Cristali, M. (2010, March 18). Personal Communication.

Garland, A.F., Kruse, M., & Aarons, G.A., (2003). Clinicians and outcome measurement: What's the use? *Journal of Behavioral Health Services and Research, 30*, 393–405.

Glisson, C., & James, L.R. (2002). The cross-level effects of culture and climate in human services teams. *Journal of Organizational Behavior, 23*, 767-794.

Glisson, C., & Hemmelgarn, A.L. (1998). The effects of organizational climate and interorganizational coordination on the quality and outcomes of children's' services systems. *Child Abuse and Neglect, 22*(5), 401-421.

Hemmelgarn, A.L., Glisson, C., & Dukes, D. (2001). Emergency room culture and the emotional support component of family-centered care. *Children's Health Care, 30*(2), 93-110.

Hemmelgarn, A.L., Glisson, C., & James, L.R. (2006). Organizational culture and climate: Implications for services and interventions research. *Clinical Psychology: Science and Practice, 13*(1), 73-89.

Hu, L. & Bentler, P.M. (1998). Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

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

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.

Kline, R.B. (2005). *Principles and practice of structural equation modeling*. 2nd edition.New York: The Guilford Press.

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341.

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*, 366–381.

Millsap, R. E, & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*, 479-515.

Muthen, L.K., & Muthen, B.O. (1998-2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Authors.

Muthen, L.K., & Muthen, B.O. (2011). *Mplus short courses: Topic 2: Regression analysis, exploratory factor analysis, confirmatory factor analysis, and structural equation modeling for categorical, censored, and count outcomes*. Retrieved April 2, 2012 from http://statmodel.com/download/Topic%202-v20%20%5BCompatibility%20Mode%5D1.pdf

Ogborne, A.C., Wild, T.C., Braun, K., & Newton-Taylor, B. (1998). Measuring treatment process beliefs among staff of specialized addiction treatment services. *Journal of Substance Abuse Treatment, 15,* 301-312.

Pignotti, M. & Thyer, B. A., (2009). Use of novel unsupported and empirically supported therapies by licensed clinical social workers: An exploratory study. *Social Work Research, 33*(1) 5-17.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online 8*(2), 23-74.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*, 323–337.

Stahmer, A. C., & Aarons, G. A. (2009). Attitudes toward adoption of evidence-based practices: a comparison of Autism early intervention teachers and children's mental health providers. *Psychological Services*, *6*(3), 223-234.

Thyer, B. A., & Myers, L. L. (2011). The quest for evidence-based practice: A view from the United States. *Journal of Social Work, 11*, 8-25.

Yuan, K.-H. & Bentler, P. M. (2004). On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, *64*, 737-757.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Doctoral dissertation, University of California, Los Angeles.

Table 1
EBPAS Four-Factor Model

Factor	Item	Item Text
Requirements	11	Required by supervisor
	12	Required by agency
	13	Required by state
Appeal	9	Intuitively appealing[a]
	10	Makes sense to you[a]
	14	Colleagues used intervention and were happy with it
	15	Received enough training to use it correctly
Openness	1	Like new therapies/interventions to help clients
	2	Would try new therapies even if they were manualized
	4	Would use research-based therapy
	8	Would try new/different therapies
Divergence	3	Know better than researchers how to care for clients
	5	Research-based therapies are not clinically useful
	6	Clinical experience more important than manualized interventions
	7	Won't use manualized therapy/interventions

Note. See Aarons (2004) for exact item wording, including item stems. [a]Residual covariance between items modeled by Aarons et al. (2007) and Aarons et al. (2010).

Table 2
Results of Prior Confirmatory Factor Analyses

	Aarons (2004)	Aarons et al. (2007)	Aarons et al. (2010)
Sample			
Respondents	163 clinical/case management workers	221 direct mental health workers	1,089 child mental health workers
Agencies	51 public child and adolescent mental health workers in one locale	Public and private agencies serving children with SED (N not stated)	100 public and private mental health clinics in 26 states
Fit Statistics			
Chi-square	114.92(84), $p < .001$	183.51(83), $p < .001$[a]	403.22(83), $p < .001$[a]
CFI/TLI	.93/.92	.92/.90	.94/.92
RMSEA	.067	.07	.060
SRMR	.077	.07	.058
Factor Loadings			
Requirements			
11: Supervisor	.88	.90	.88
12: Agency	.99	.99	.99
13: State	.78	.81	.77
Appeal			
9: Intuitive	.83	.45[b]	.49[b]
10: Makes sense	.89	.54[b]	.63[b]
14: Colleagues like it	.56	.76	.75
15: Been trained	.55	.68	.80
Openness			
1: Like new therapies	.62	.67	.70
2: Would use manualized	.61	.80	.78
4: Research-based ok	.81	.68	.85
8: Different from usual	.66	.70	.68
Divergence			
3: Know better	.34	.50	.49
5: Research not useful	.65	.68	.68
6: Clinical experience	.42	.43	.55
7: Won't use manualized	.76	.70	.57

Notes. CFI: Confirmatory Fit Index. TLI: Tucker-Lewis Index. RMSEA: Root Mean Square Error Of Approximation. SRMR: Standardized Root Mean Square Residual. [a]Fit statistics adjusted for clustering. [b]Residual covariance estimated for these two items.

Table 3
Correlations (Polychoric) Between EBPAS Items (N = 1,260)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000														
2	0.659	1.000													
3	0.044	-0.085	1.000												
4	0.578	0.649	-0.107	1.000											
5	-0.099	-0.164	0.336	-0.252	1.000										
6	0.052	-0.031	0.246	-0.009	0.301	1.000									
7	-0.123	-0.174	0.195	-0.234	0.527	0.279	1.000								
8	0.477	0.585	-0.124	0.571	-0.216	-0.029	-0.148	1.000							
9	0.365	0.392	-0.072	0.429	-0.245	-0.002	-0.212	0.470	1.000						
10	0.372	0.381	-0.059	0.420	-0.259	0.010	-0.214	0.461	0.732	1.000					
11	0.178	0.246	-0.097	0.200	-0.097	-0.006	-0.102	0.229	0.276	0.349	1.000				
12	0.191	0.259	-0.127	0.222	-0.102	-0.020	-0.104	0.233	0.271	0.330	0.873	1.000			
13	0.207	0.231	-0.109	0.189	-0.122	-0.052	-0.174	0.191	0.283	0.357	0.752	0.814	1.000		
14	0.282	0.303	-0.115	0.332	-0.188	-0.001	-0.172	0.338	0.510	0.540	0.375	0.363	0.398	1.000	
15	0.329	0.372	-0.075	0.393	-0.244	-0.014	-0.221	0.404	0.535	0.631	0.409	0.422	0.503	0.659	1.000
Response Category Proportions															
0	0.009	0.011	0.162	0.006	0.375	0.100	0.366	0.018	0.014	0.009	0.019	0.015	0.045	0.019	0.015
1	0.045	0.064	0.282	0.044	0.300	0.211	0.274	0.079	0.056	0.029	0.068	0.050	0.066	0.062	0.028
2	0.313	0.272	0.340	0.298	0.256	0.429	0.290	0.327	0.307	0.187	0.253	0.234	0.206	0.282	0.187
3	0.423	0.418	0.165	0.428	0.053	0.193	0.057	0.386	0.390	0.422	0.362	0.361	0.337	0.412	0.388
4	0.210	0.235	0.050	0.224	0.016	0.067	0.013	0.189	0.233	0.354	0.298	0.340	0.346	0.226	0.381
Mn	2.781	2.802	1.661	2.821	1.038	1.917	1.078	2.646	2.773	3.083	2.853	2.963	2.871	2.765	3.092
ICC	.025	.059	.054	.035	.045	.022	.020	.040	.035	.037	.023	.026	.016	.041	.030

Notes. Correlations are weighted least square mean and variance adjusted estimates and include cases with partial missing data. Response category proportions are proportions of respondents selecting that response scale option. Means computed by treating EBPAS responses as continuous. ICCs computed for ordinal data.

01. Like new therapies/interventions
02. Willing to follow a treatment manual
03. Know better than researchers
04. Willing to use therapy developed by researchers
05. Research-based treatments are not useful
06. Clinical experience is more important
07. Will not use manualized therapy
08. Would try new/different therapy
09. Intuitively appealing
10. Makes sense to you
11. Required by supervisor
12. Required by agency
13. Required by state
14. Colleagues are happy with intervention
15. Have received enough training to use it

Table 4
Results of Confirmatory Factor Analyses (N = 1,260)

	Aarons et al. (2007)	Model AC	Model FF
Fit Statistics			
Chi-square	363.621(83), $p < .001$	346.684(82), $p < .001$	319.741(79), $p < .001$
CFI/TLI	.979/.974	.980/.975	.982/.976
RMSEA	.052 (CI=0.046-0.057)	.051 (CI=0.045-0.056)	.049 (CI=0.044-0.055)
WRMR	1.404	1.357	1.253
Factor Loadings			
Requirements			
11: Supervisor[a]	.897	.897	.897
12: Agency	.960	.960	.960
13: State	.852	.852	.852
Appeal			
9: Intuitive[a]	.705[b]	.705[c]	.706[d]
10: Makes sense	.762[b]	.762[c]	.763[d]
14: Colleagues like it	.725	.725	.724
15: Been trained	.842	.842	.842
Openness			
1: Like new therapies[a]	.719	.719	.720
2: Would use manualized	.830	.830	.830
4: Research-based ok	.788	.788	.789
8: Different from usual	.730	.730	.729
Divergence			
3: Know better[a]	.413	.388[e]	
5: Research not useful	.807	.815	
6: Clinical experience	.305	.268[e]	
7: Won’t use manualized	.685	.683	
Clinical			
3: Know better[a]			.541
6: Clinical experience			.454
Research			
5: Research not useful[a]			.789
7: Won’t use manualized			.668

Notes. Fit statistics adjusted for clustering. Standardized coefficients are reported. All coefficients significant at $p < .001$. CFI: Confirmatory Fit Index. TLI: Tucker-Lewis Index. RMSEA: Root Mean Square Error Of Approximation. SRMR: Standardized Root Mean Square Residual. [a]Factor fixed at 1.00, unstandardized. [b]Residual covariance = .423, $p < .001$. [c]Residual covariance = .423, $p < .001$. [d]Residual covariance = .422, $p < .001$. [e]Residual covariance = .160, $p < .001$.

Table 5
Factor Correlations for Aarons, et al. (2007), Added Covariance (AC), and Five Factor (FF) Models (N = 1,260)

	Aarons, et al. (2007) model				
	1	2	3	4	
Requirements	1.000	0.556***	0.308***	-0.191***	
Appeal	0.556***	1.000	0.644***	-0.359***	
Openness	0.308***	0.644***	1.000	-0.286***	
Divergence	-0.191***	-0.359***	-0.286***	1.000	
Variance	0.805***	0.497***	0.517***	0.171***	
	Added Covariance (AC) model				
	1	2	3	4	
Requirements	1.000	0.556***	0.308***	-0.193***	
Appeal	0.556***	1.000	0.644***	-0.364***	
Openness	0.308***	0.644***	1.000	-0.292***	
Divergence	-0.193***	-0.364***	-0.292***	1.000	
Variance	0.805***	0.497***	0.517***	0.150***	
	Five Factor (FF) model				
	1	2	3	4	5
Requirements	1.000	0.556***	0.308***	-0.192***	-0.168***
Appeal	0.556***	1.000	0.644***	-0.097	-0.397***
Openness	0.308***	0.644***	1.000	-0.098	-0.325***
Clinical	-0.192***	-0.097	-0.098	1.000	0.771***
Research	-0.168***	-0.397***	-0.325***	0.771***	1.000
Variance	0.804***	0.498***	0.519***	0.293***	0.623***

Note. ***p < .001.

Table 6
Factor Correlations and Unstandardized Means and Variances for the Added Covariance (AC) Model for Community-based Programs and Residential Programs

	1	2	3	4
Community-based (N = 277)				
Requirements	1.000	0.490***	0.212***	-0.288***
Appeal	0.490***	1.000	0.584***	-0.324***
Openness	0.212***	0.584***	1.000	-0.098*
Divergence	-0.288***	-0.324***	-0.098*	1.000
Variances	0.751***	0.460***	0.671***	0.276***
Means[a]	0.000	0.000	0.000	0.000
Residential (N = 538)				
Requirements	1.000	0.601***	0.308***	-0.165***
Appeal	0.601***	1.000	0.644***	-0.428***
Openness	0.308***	0.644***	1.000	-0.396***
Divergence	-0.165***	-0.428***	-0.396***	1.000
Variances	0.954***	0.420***	0.656***	0.234***
Means	0.298***	0.136	0.336***	-0.021

Note. ***$p < .001$. Coefficients estimated for model with residual covariances for items 9 and 10 for items 3 and 6 freed. [a]Factor means fixed at 0.0.