Measuring Coaching Effectiveness: Validation of the

Workplace Outcome Suite (WOS) for Coaching

Richard D. Lennox, Ph.D.[1]

David Sharar, Ph.D.[1]

Francine Miller,[2]

[1] Chestnut Global Partners, 1003 Martin Luther King Drive, Bloomington, IL 61701

[2] Health Advocate, San Francisco, CA

ABSTRACT

Chestnut Global Partners developed the Coaching version of the Workplace Outcome Suite (WOS) to measure the effectiveness of coaching interventions to address common workplace issues. The original 25-item version of the WOS was slightly modified where outcomes questions were tailored to directly relate to the coaching process. Although the changes were modest, there was still some concern that they might adversely affect the psychometric characteristics of the 25 items by compromising their validity and reliability. Standard scientific prudence demands that we evaluate the coaching version of the WOS to assure its equivalence with the original WOS. Data from 309 clients using coaching services as part of disease management for depression or diabetes was used, and various analytical methods, including reliability analysis, confirmatory factor analysis, and correlational studies, were included in the study. The results of these analyses showed that it performed slightly better than it did in the original WOS. The instrument was shown to be extremely reliable, particularly for a short scale. We concluded that the 25-item coaching version of the WOS can be used to test the efficacy and effectiveness of the coaching program without concerns for significant measurement error. The results suggest that psychometric studies of the original WOS can be applied to the coaching version of the WOS, that it can be adapted to and reflect clinical change, and that the WOS is able to detect changes effected by workplace health programs with samples as small of 50 clients.

Measuring Coaching Effectiveness: Validation of the

Workplace Outcome Suite for Coaching

A growing body of literature offers promising support for the efficacy of coaching-based disease management (DM) programs aimed at supplementing clinical treatment for many chronic diseases, including heart disease, hypertension, depression, and diabetes (Goetzel et al., 2007; Pelletier, 1999, 2005; Merrick, Horgan, Garnick, Hodgkin, & Morley, 2008).This research demonstrates that call centers organized around a case-management strategy can produce statistically detectable clinical change across time. Moreover, the use of randomized clinical trial (RCT) designs in many of these studies has provided convincing evidence that it is the management itself that causes the observed improvement in clinical status, rather than the other way around. However, this body of research is not without its criticism; indeed, Lewis and Khanna, (2014) assert that these studies are so methodologically flawed that the field cannot attest to the effectiveness of the DM approach.

These concerns notwithstanding, recent corporate-based disease management has pushed forward with coaching DM models for controlling the workplace costs of chronic conditions such as depression, diabetes (Petrak & Herpertz, 2009) and heart disease. Focusing on important heath related costs such as absenteeism, productivity and company-based health insurance costs, these programs have employed call centers and case management activities to impact behaviors such as medication compliance, treatment adherences, and other wellness activities that serve to support the initial clinical treatment. Some studies have supported workplace disease management (WDM) in improving clinical status and cost-effectiveness, and a few have found positive effects on workplace outcomes such as presenteeism (Burton, Chen, Conti. Schultz & Edington, 2006; Johns, 2008: Burton, Chen, Conti, Schultz, Pransky, & Edington, 2005), and a few have even supported long-term outcomes such as health care utilization and costs (Goetzel et al., 2007; Pelletier, 1999, 2005).

Studies of the impact of coaching-based WDM programs have focused largely on specific effects within the broader health services system (Pelletier, 1999, 2005; Pelletier, Boles, Lynch, 2004). For example, the majority of the research has sought to establish the efficacy of WDM programs in terms of their ability to improve specific clinical markers of diseases, with less attention given to the impact of these programs on workplace functioning and general health status (Pelletier, 1999). For depression, this means focusing on depression symptoms, rather than its impact on work performance. Often, this focus has been accompanied by the use of RCT methodologies that place a high value on being able to assert casual direction at the cost of a broader, longer range analysis of workplace functioning and health care costs. Unlike typical clinical bench research, health services research often builds on previous tests of efficacy to assert causal direction. This approach allows for the use of random-effects models that analyze the variance in existing measures rather than assigning subjects to conditions. While this research may be limited in its ability to assert causal direction, it provides a stronger perspective on the real-world aspects of the health services system.

There have been few attempts to track the effectiveness of coaching programs and in the actual operational environment. With the proliferations of coaching in EAP programs, it would be prudent to examine these coaching models within the context of broader long-term effectiveness research rather than short-term efficacy models designed largely to isolate evidence that an intervention is simply capable of producing some change in clinical markers.

There has been a growing interest in WDM programs for depression and other mental health problems (Pelletier, 2005). Recent RCT studies have reported that depression DM programs resulted in reduced symptoms of depression, improved job retention and work productivity, and reduced absenteeism (Rost, Smith, & Dickinson, 2004; Wang et al., 2007; Wang, Simon, & Kessler, 2008). Although these results are very encouraging, little is known about the inter-relationships among improvements in health and improvements in workplace function, and the impact of these improvements on future health care costs; that is, whether

changes in depression symptoms are associated with reductions in absenteeism and presenteeism, and whether those changes are associated with reductions in future health care costs. The systemic model shown in Figure 1 tracks the collinear effects of both depression treatment and WDM through these mediating variables.

Although there is obvious appeal to the idea of employing problem-focused one-to-one guidance to a variety of personal issues, the empirical support for coaching is limited and somewhat fragmented. Most of the studies focus more on the one specific disease than on the coaching methodology itself, and as a whole the collection suffers from a lack of scientific foundations that support the overall efficacy of the approaches. For example, most of the evaluations of coaching programs were not designed to assess comprehensive workplace outcomes (Schneider et al. 2011) conducted an evaluation of a coaching program for managing diabetes, but the evaluation was limited to satisfaction with the program and clinical goal attainment and did not include a control group. On the other hand, Merrell, et al, (2011) designed a comprehensive clinical outcomes study for a wellness program but it did not include a coaching component, nor did it include workplace oriented functional measures or a control group. Hawk, Athernot, Taylor, Scuffham, Eadie, Miller, & Oldenburg (2009), conducted a randomized clinical trial of the scripted telephone coaching program for treating cardiovascular disorders. The study included measures of quality of life from the SF36 and physical activity, but did not include workplace functioning or life satisfaction. Linden, et al (2010) employed a motivation interviewing-based coaching approach to compare participants to non-participants, finding that higher levels of self efficacy, patient activation, lifestyle change and perceived health status, lowers levels of increased stage change and risk over time. Aoun, et al. (2011) tested the effectiveness of a BMI coaching program, finding improvement in weight loss, dietary habits and quality of life as measured with the SF36 and life satisfaction. Neither of these studies examined workplace outcomes. Patja, et al., (2012) tested the effectiveness of telephonic coaching for the self-care of hypertension disease in a randomized clinical trial, but failed to

support the hypothesis when testing primary clinical outcomes; however, the study failed to consider broader quality of life and work functioning measures.

Some studies were unable to find consistent support for Coaching. For example, Margolis, et al (2012) studied hypertension treatment for low income in minority populations, and although there was a significant reduction in systolic blood pressure, there was no significant difference between treatment arms of the study (home titrations versus no-home titration conditions). However, there was an interesting dose response effect indicating that the more coaching sessions a patient received the better the outcomes. This result offers some limited support for the effectiveness of coaching in general by showing the added benefits of continued coaching sessions.

These studies suggest that the coaching field would benefit from a standardized outcomes measure with which to test the efficacy and effectiveness of various health promotion activities. The article tests the validity of previously validated outcome measures using the Workplace Outcome Suite (for details see Lennox, et al., 2011), which was slightly modified for use with coaching interventions.

A better understanding of the dynamic effects of WDM as they pass through the proximal and distal effects will give counselors a better understanding of the value of WDM for monitoring the effects of their program on job performance, and hence the value of monitoring a wide range of workplace outcomes. The mediating model will also provide a clearer picture of the value of WDM for bottom line effects, including more proximal effects associated with absenteeism and presenteeism and more distal effects involving health care costs. Finally, system-wide tests of depression WDM will provide policy makers with an evidence-based impression of the value of depression WDM for controlling societal health care costs.

Figure 1 illustrates a model for using structural equation modeling with latent variables (Bollen, 1989) to examine the impact of coaching for depression on future health care costs. The model shows how three exogenous variables related to coaching sessions and propensity

scores that capture pre-existing differences between treatment and controls group affect clinical

measures of depression, which further affects absenteeism and presenteeism as medial

outcomes when then ultimately impact the distal outcomes in future health care costs. The

various β terms in the figure illustrate the structural coefficients that can be isolated by the
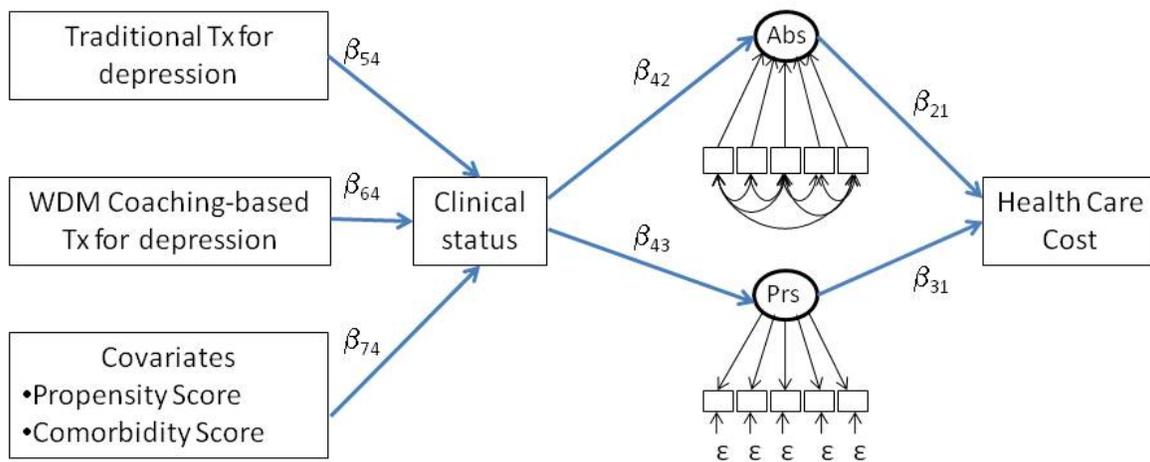
model. For example

Figure 1: A path model of Coaching-based Depression WDM

β_{21} and β_{31}, show the impact of absenteeism and presenteeism respectively on future health

care costs in 2015 and, more importantly, the impact on changes in depression caused by the

coaching assignment, to treatment impacts health care costs through the other variables in the

model. The primary objective of a model like this would be to test the effectiveness of DM on

three classes of outcome measures: 1) ***proximal*** clinical outcomes, 2) ***medial*** workplace outcomes, and 3) ***distal*** future healthcare cost. The Structural Equation Model (SEM) presented in Figure 1 illustrates the systemic impact that depression and diabetes WDM programs have on the workplace. The path model shows that depression treatment has one impact -- to reduce depression symptoms -- while WDM has a statistically unique impact as well, after statistically controlling for differences along the depression and WDM continua via the propensity score. The model shows that depression symptoms influence workplace functioning in terms of absenteeism and presenteeism (Sanderson, Tilse, Nicholson, Oldenburg, and Graves, 2007), which in turn have unique effects on future health care costs observed in 2018.

The Workplace Outcome Suite (WOS) was designed to evaluate the effectiveness of EAPs and other wellness activities. It was validation of studies of the effectiveness of traditional programs including call centers and other therapeutic intervention, but not on coaching per se. The brief description of the design and validation of the WOS which is published elsewhere (Lennox, et al., 2010) is prescription below.

Design of the Original Workplace Outcome Suite

The WOS was designed to provide assessment on relevant individual differences that focus on outcomes that are related specifically to EA interventions and are likely to change across time if the intervention works. Unlike most prior workplace outcome measures, the scales in the WOS were designed using a common theoretical framework built around more than 100 years of applied psychometric research and practice. The result is a set of short measures that are designed around the same measurement theory and validated using the same set of psychometric principles (Nunnally, 1978; Nunnally & Bernstein, 1994). Starting with the explicit measurement model that Bollen (1989) prescribed for all scales, we constructed 5-item scales from which a user can select those best suited for his or her research and

assessment purposes. None of the existing measures of the constructs addressed in Cluster I of the WOS were designed around a formal measurement model, so we felt that our scales were an improvement in this regard.

The WOS was constructed on two formal measurement models: an effect-Indicator model and a formative measurement model (Bollen & Lennox 1991). As shown in Figure 3, in the effect-indicator model, the measured indicators are thought to emerge from a single underlying latent variable (illustrated by the ellipse). The small arrows under the boxes represent the random errors in the items that are effectively removed by adding the items together. The mechanics of the effect indicator model can be characterized as:

$$Y_i = \lambda_{i1}\xi_1 + \varepsilon_i \qquad\qquad [1]$$

Where Y_i is the ith item in the scale, ξ_1 is the single latent construct presumed to underlie the covariance among the item set, λ_{i1} is the factor loading of the ith item on the single latent factor, and ε_i is the random measurement error in the ith item**.** The random measure in ε_i is at the heart of the effect indicator model because it provides the basis for offsetting the random errors, which are expected to be half in the positive direction and half in the negative direction, thus offsetting one another in the sum or average of the items.

The formative measurement model uses different items that combine linearly to create a broad construct. The right side of Figure 3 illustrates the manner in which the items combine to form the latent construct. The formative model can be expressed as

$$\eta_1 = \gamma_{11}X_1 + \gamma_{21}X_2 + \gamma_{31}X_3 \ldots \gamma_{i1}X_4 + \zeta_1 \qquad\qquad [2]$$

Where η_1 is the latent construct and $\gamma_{11}X_1 + \gamma_{21}X_2 + \gamma_{31}X_3 \ldots \gamma_{i1}X_4$ are coefficients that indicate the

contribution that each item makes to the formulation of the latent construct in a manner similar

to multiple regression. The ζ_1 term reflects the variance in the latent construct not explained by

the specific set of indicators. Notice that there is no random measurement error for the

individual items; for this reason, coefficient alpha is not appropriate for assessing reliability in

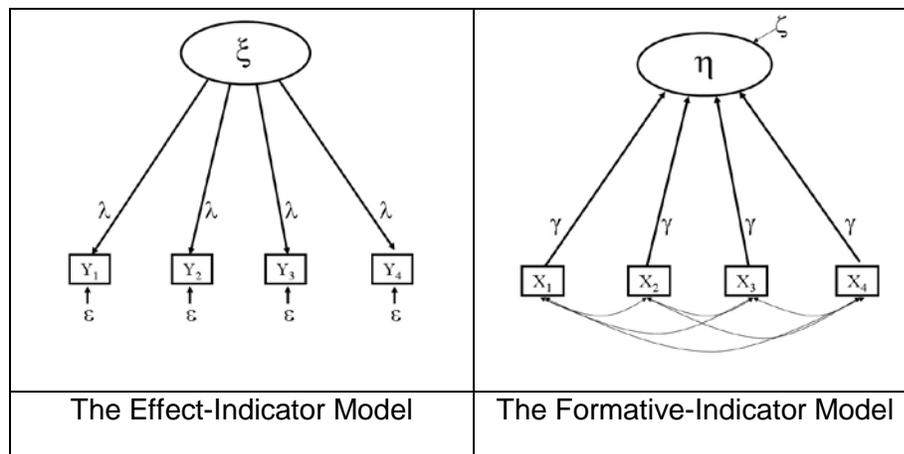this type of model (Bollen 1984, 1989; Lennox & Bollen, 1991).

| The Effect-Indicator Model | The Formative-Indicator Model |

Figure 3: Formal Measurement Models

Results of the psychometric analysis of the WOS provide some initial support for its

reliability, structural validity, and construct validity. Two separate validation studies of the suite,

one with a paper-and-pencil modality ($N = 220$) and the other with a telephone interview

modality ($N = 228$), tested the reliability of the scales, the structural validity of the items, and the

construct validity of the unit-weighted scale scores. (Note that the Absenteeism scale is based

on a formative measurement model that does support coefficient alpha as an index of reliability

or factor analytic procedures, so it was not included in these analyses.) The effect-indicator

scales were found to have moderate (coefficient alphas in the range of .75) to excellent

(coefficient alphas in the range of .90) levels of internal consistency. All of the items produced factor loadings of .30 or higher in the paper-and-pencil study. Only two items in the Work Engagement scale produced low factor loadings in the telephone interview study; however, since these low loadings did not replicate in the paper-and-pencil study, they were considered spuriously low for the time being. Correlations of the scale scores with self-reported measures of relevant behavior and emotions provided limited evidence of construct validity for all five scales. The results suggest support for the use of the WOS to evaluate EAP services and interventions

The coefficient alphas for the paper-and-pencil study were as follows: (a) Presenteeism, α = .90; (b) Work Engagement, α =.74; (c) Life Satisfaction, α =.76; (d) Workplace Distress, α = .90. The coefficient alphas for the Telephone Interview study were similar: (a) Presenteeism, α = .92; (b) Work Engagement, α =.63; (c) Life Satisfaction, α =.78; (d) Workplace Distress, α = .88. Item-total correlations for all 20 effect indicator items were substantial and significant beyond p < .01. Analysis of the alpha coefficients created after iteratively removing a single item showed that there was very little improvement to be gained by removing any one item. There are some items in the Workplace Engagement scale that suggest some room for improvement, which will be better understood in the context of the confirmatory factor analysis in the next section.

We conducted a confirmatory factor analysis on the two data sets to test the overall structure of the items as they relate to their respective latent constructs and one another. This analysis allowed us to evaluate the item-level internal consistency and discriminant validity of the items in terms of loading on their respective factor but not on other factors. In confirmatory factor analysis we test the hypothesis that the item defines its appropriate factor and is unrelated to the other factors. This is achieved by fixing the "off factor" loading to zero and testing the goodness-of-fit for the model of the four-correlated factors. Figure 3 illustrates this factor structure with the four ellipses representing the respective latent factors for Presenteeism,

Work engagement, Life Satisfaction, and Workplace Distress. We tested the model using Maximum Likelihood estimation. For the telephone interview sample of 228 respondents, modeling the observed covariance matrix to the hypothesized four-factor model yielded a Bentler-Bonnet fit index of .86, a Bollen fit index of .88, and a comparative fit index of .88. Similarly, modeling the observed covariance matrix from the paper-and-pencil data set ($N = 220$) yielded a Bentler-Bonnet Fit index of .83, a Bollen fit index of .86, and a comparative fit index of .85. Both models yield respectable fits to their respective covariance matrices.

The results support the general fit of the 25-item scale to the hypothesized structure of the four effect-indicator latent variables. This paper presents a validation of the slightly revised version of the Workplace Outcome Suite (Lennox, et al., 2010) modified for use with coaching interventions. Changes were focused on instructions to the respondents as they relate to coaching rather than to EAP services. All 25 items and subscales in the Coaching version of the WOS and the validations is focused on the suitability of the revised instruction. The basic methodology is the same as was used in validating the original WOS.

METHOD

Data Sources

Data for the study were provided by two EAPs offering coaching services as part of their program. In Study I Chestnut Global Partners provided data from 309 clients using coaching service as part of the disease management program for depression or diabetes. In Study II Wellcall N=311

RESULTS AND DISCUSSION

Analysis

Descriptive statistics. Central tendency and dispersion in the individual items in the Workplace Outcome Suite will be examined for severe departure from normality in distribution that are expected to be normal, which might create problems for traditional statistical analysis.

Variation around the mean will also be monitored for evidence of small variance and thus limited discriminability and/or large variance that might indicate confusion and cause excessive measurement error.

   Reliability analysis. The basic structure of the effect indicator measurement models will be assessed using an analysis of internal consistency. Under the assumption of the traditional unidimensional effect indicator model, the reliability of a multi-item scale of parallel items can be used to assess the reliability of scale scores as indicated by the common variance reflected in the first factor. Coefficient alpha (Cronbach, 1951) will be used to assess the internal consistency of the respective item-sets in the individual scale. The alpha coefficient will be used as a standard for evaluating the random measurement error of the individual scales.

   Confirmatory factor analysis. Scales measuring Presenteeism, Work Engagement, Life Satisfaction and Workplace Distress are based on the effect indicator models, and are expected to define a unidimensional factor that reflects their respective underlying construct. Figure 3 illustrates the theoretical structure of the four effect-indicator latent variables. The

-----Insert Figure 1 about here-----

figure shows that the individual scales are thought to emerge from their respective single underlying constructs, their unique variance is considered to be random measurement error, and that the four latent constructs are also expected to be intercorrelated with one another.

   In contrast to the absenteeism that assesses the individual components of total time away from work, these four items contain highly intercorrelated elements that share a common factor reflected in the common variance. Confirmatory factor analysis will be used to test the internal structure of the 20 items in these scales as being capable of defining four correlated, yet unique factors within their covariance matrix. Maximum likelihood estimations will be used to fit a four-factor model to the covariance matrix, with a χ^2 test of goodness-of-fit used to test the

residual covariance matrix against the model matrix. We hypothesize the four-correlated-factor model will account for the majority of the reliable variance in the covariance matrix. We also expect that standardized factor loading connecting the individual items to their respective latent variable will be significant and substantial.

Bivariate correlations of scale scores. To provide some limited construct validity we will correlate scale scores with some self-reported pseudo-behavioral and emotional measures that are expected to correlate with the constructs we are measuring. For example, the item "I have a hard time getting out of bed" negatively correlates with work engagement because the engaged worker would be passionate about getting to work. A positive correlation or an insignificant correlation might indicate a problem with our work engagement measures. Similarly, the item "I often feel sad" would be expected to correlate positively with workplace distress because the distressed worked would be expected to feel sad even when away from the job. The following items are used as pseudo-criterion related targets for the five individual outcome measures.

- I have a hard time getting out of bed (i.e., item should correlate negatively with Life Satisfaction).

- I often feel sad (i.e., item should correlate positively with Workplace Distress).

- I keep falling behind schedule at work (i.e., item should correlate negatively with Work Engagement).)

- I am rarely late for work (i.e., should correlate negatively with Absenteeism)

- I often get home late from work. (i.e., should correlate positively with Work Engagement).

Results

Descriptive statistics. Table 2 presents the means and standard deviation for the 25

-----Insert Table 2 about here-----

items in the Outcome Suite across the two administration modalities. Wrapping the 95%

confidence interval around the means in both groups shows that none of the difference in the

items are significantly different from zero, showing a basic comparability of the two modalities in

terms of central tendency and dispersion. The comparability is somewhat surprising for the

absenteeism items that used a 30-day recall period for the paper-and-pencil study, and a 7-day

recall for the telephone interview study.   While there are a slightly higher number of average

hours for the 30-day recall, the difference is not significant. This suggests that there may be little

to gain by forcing employees to try to remember absenteeism beyond 7 days. On the other

hand, the results also suggest that it doesn't make too much difference when recall is used in

terms of estimating the effects. The means for presenteeism, work engagement, life

satisfaction, and workplace distress are centered onthe middle of the Likert response format

and the standard deviation indicates a healthy level of variability within each item. Together, the

means and standard deviation indicate sufficient discriminability and symmetry consistent with a

normal distribution of the individual ratings.

Reliability analysis. Table 3 presents the internal consistency for the four effect

indicator

-----Insert Table 3 about here-----

measures across the two studies. Coefficient alpha for the paper-and-pencil study are as

follows Presenteeism = .90; Work Engagement =.74; Life Satisfaction =.76, and Workplace

Distress = .90. For the Telephone Interview study, Presenteeism = .92; Work Engagement

=.63; Life Satisfaction =.78, and Workplace Distress = .88. Item-total correlations for all 20

effect indicator items are significant beyond the .01 level and substantial. Analysis of the alpha

coefficients that would be created by removing a single item shows that there is very little

improvement to be gained by removing any one item. There are some items in the workplace

engagement that suggest some room for improvement, which will be better understood in the

context of the confirmatory factor analysis in the next section.

     Confirmatory factor analysis. We conducted a confirmatory factor analysis on the two

data sets to test the overall structure of the items as they relate to their respective latent

constructs and the others. This allows us to evaluate the item level internal consistency and

discriminant validity of the items in terms of loading on their respective factor and not on other

factors. In confirmatory factor analysis we test the hypothesis that the item defines its

appropriate factor and are unrelated to the other factors. This is achieved by fixing the "off

factor" loading to zero and testing the goodness-of-fit for the model of the four-correlated factor.

Figure 3 illustrates this factor structure with the four ellipses representing the respective latent

factors for Presenteeism, Work engagement, Life Satisfaction and Workplace Distress.

The curved lines connecting the ellipses represent the correlations among the factors that are to

be estimated. The arrows connecting each item to its respective ellipse represent the factor

<div align="center">-----Insert Figure 3 about here-----</div>

loading of the items. The small arrows at the bottom of the items represent the random

measurement error in each item that is removed from the latent factor.

     We tested the model in Figure 3 using Maximum Likelihood estimation for the telephone

interview sample involving 228 respondents. Modeling the observed covariance matrix to the

hypothesized four-factor model yields a Bentler-Bonnet fit index of .86, a Bollen fit index of .88,

and a comparative fit index of .88. Modeling the observed covariance matrix from the paper-

and-pencil data set yields a Bentler-Bonnet Fit index of .83, a Bollen fit index of .86, and a

comparative fit index of .85. Both models yield respectable fits to their respective covariance matrices.

Table 4 contains the standardized loading for the items on their respective latent factor.

-----Insert Table 4 about here-----

The absenteeism items are not included in this analysis. Virtually all of the factor loadings for the telephone interview data were significant and substantial, suggesting that they are all appropriately related to their respective latent factor. The two exceptions are item 15, "I often find myself thinking about my work at home," and item 12, "I often think about work on my way to the work site," which are hypothesized to be core items of the work engagement scale. These loadings suggest that respondents affirming these items may not necessarily be highly engaged in their work; it may be that these respondents are distressed by their work. Still, all other items have high loading, and the two negatively worded life satisfaction items produced the expected negative loading on their respective factor. The standardized factor loading for the telephone interview data set produced significant and substantial for all items, including the two items in the Work Engagement scale. This suggests that the problem seen in the first sample may not replicate and it may be premature to consider modifying the scale for these items.

Table 5 presents the correlations among the latent factors as estimated in the

-----Insert Table 5 about here-----

confirmatory factor analysis. These factors are estimated using the precise weights used in the analysis and should not be confused with correlations with the unit-weighted scale scores. Correlations for the paper-and-pencil data set are presented below the diagonal and the telephone interview is presented above the diagonal. The results show a generally parallel pattern of positive and negative correlations with the difference in direction being attributed to the direction of scoring in the items that indicate which is associated with a high score. Workplace distress and work engagement produce a very high correlation in the telephone interview data set (r = -.80), but it is considerably lower in the paper-and-pencil data set (r = -

.57). The two are not interchangeable, but it remains to be seen what level of differential prediction can be found with such highly correlated items. All other correlations are consistent with their respective theoretical constructs.

Taken together, these two sets of results support the structure of the 20 effect indicator items in the Workplace Outcome Suite and measure four correlated latent constructs. When considered with the alpha coefficients, the results suggest that the scale can be expected to produce reliable measures of their respective constructs without excessive random measure error or overlap with other related constructs in the suite.

Bivariate correlational analysis. Table 5 presents the results of a limited test of construct validity. In the analysis we created five single-item measures of behavior and/or emotions that should be theoretically related to the Outcome Suite scales in one form or another. The first two items were directed at a behavior and an emotion associated with negative states.

-----Insert Table 5 about here-----

In the paper-and-pencil data set, both measures are negatively associated with life satisfaction and workplace distress. They are also related to presenteeism in a direction suggesting that individuals whose personal problems adversely affect their job performance also have some level of sadness and difficulty that is consistent with depression. By way of discriminant validity, not all of the measures are associated with the Workplace Outcome Suite. For example, Work Engagement is not associated with feeling sad and absenteeism is not associated with having a hard time getting out of bed. As expected, falling behind schedule at work is associated with absenteeism, presenteeism, and workplace distress, but not with work engagement, and only moderately with life satisfaction. The behavior of not being late for work is negatively associated with absenteeism, and getting home late from work is positively associated with work engagement, suggesting high scorers do not watch the clock.

In the telephone interview data set, the same pattern of significance was found for the two depression indicators and falling behind schedule at work. All provide the same level of limited construct validation for the scale. The items referring to being late for work and staying late at work also showed the same support for work engagement. The slight positive correlation between workplace distress and getting home late from work is consistent with a complex interpretation of the thinking about work at home that may be more related to work distress than to work engagement. Still the association is very small and probably not indicative of a large problem with those items.

The correlation results provide evidence that the Outcome Suite does measure its intended construct. However, because the items are all self-reported, they are likely affected to some degree by common method variance that limit their value as indicators of true convergent validity.

DISCUSSION

These studies support the basic psychometric properties of the Workplace Outcome Suite as measures of EAP outcomes both at the workplace and elsewhere. The formal measurement models provide a transparent description of the manner in which the individual items impact the latent variable. The two potential problems in the work engagement scale found in the telephone interview studies were not found to be a problem in the paper-and-pencil study, suggesting that those smaller coefficients did not replicate, and as such should be considered spurious for the time-being. However, the items will be carefully examined in another study of the Internet modality that is currently underway. The items will be considered for replacement should the problem reappear.

Taken together, the results of the reliability analysis, the confirmatory factor analysis, and the correlational studies all support the use of the Workplace Outcome Suite in its current form. Alpha coefficient for the Presenteeism Scale and the Workplace Distress scales are in the .90 range, showing them to be extremely reliable, especially for short scales. The mid-70's

alpha for the Life Satisfaction scale in both studies shows to have an acceptable level of reliability. The one low alpha coefficient for the Work Engagement scale does not replicate in the paper-and-pencil studies, also suggesting that the scale has acceptable reliability for a short scale in its early development (Nunnally, 1978). The basic psychometric characteristics of the Workplace Outcome Suite show it capable of measuring individual difference relevant to EAP intervention; however, further research on the construct validation, looking at actual behavior and records, will provide greater confidence on the value of the instrument.

One of the main purposes of creating the Workplace Outcome Suite was to provide the EAP researcher with a set of short measures found under a single validation umbrella. The evaluator does not need to search the literature for measures, nor comb through the proprietary catalogues, many of which do not include the syntax of the items before the purchase. The transparency of the measurement models coupled with the public nature of the validation material is designed to streamline the evaluation process and facilitate the empirical testing of EAP interventions across the board. The standardized approach to the development and validation of the five constructs facilitate evaluation of the measures for specific use. Finally, Chestnut Global Partners is making the measures available for specific intervention testing at no cost. Again, it is our intention to facilitate the use of the scientific method for evaluation and comparison of EAP services.

Authors' Notes

Although the measures are copyrighted by Chestnut Global Partners, they are available for

unqualified public use. We request that those interested contact us prior to using the measures

so we can keep track of how the measures are being used by the industry. Questions regarding

the use of the measures should be directed to Drs. Richard Lennox (rlennox@chesnut.org;

(919) 942-0448) or David Sharar (dsharar@chesnut.org; (309) 820-3570. Correspondence

should be sent to Dr. Richard Lennox, 2404 Western Park Lane, Hillsborough, North Carolina

27278.

References

Attridge, M. (2009). Measuring and managing employee work engagement: A review of the research and business literature. *Journal of Workplace Behavioral Health: Employee Assistance Research and Practice, 24*(1), 1-6.

Aoun, S., Osseiran-Moisson, R., Shahid, S., Howat, P., & O'Connor, M. (2011). Telephone Lifestyle Coaching: Is it Feasible as a Behavioural Change Intervention for men? *J. Health Psychol.* Retrieved from http://hpq.sagepub.com/content/early/2011/07/08/1359105311413480

Bollen, K. A. (1984). Multiple Indicators: Internal Consistency or no necessary relationship? *Quality and Quantity*, *18*, 377-385.

Bollen, KA (1989) Structural Equations with Latent Varialbes. New York: Wiley.

Bollen, K, A., & Lennox, R, D. (1991). Conventional Wisdom on Measurement: A Structural Equations Perspective. *Psychological Bulletin, 110:2,* 305-314.

Bollen, KA, Lennox, RD, & Dahly, DL. (2009). Practical application of the vanishing tetrad test for causal indicator measurement models: An example from health related quality of life. *Statistics in Medicine.*

Bray, J., Bray, LM., Lennox, R., Mills, M., McRee, B., Goehner, D., Higgins-Biddle, J., (2009). Evaluating Web-Based Training for Employee Assistance Program Counselors on the Use of Screening and Brief Intervention for At-Risk Alcohol Use. *Journal of Workplace Behavioral Health.*

Cronbach, LJ. (1951). Coefficient alpha and the internal consistency of tests. *Psychometricka,* 16, 297-334.

Diener, E., Emmons, RA., Larsen, RJ., Griffin. (1985). The Satisfaction with Life Scale. *Journal of Personality* 49:1.

Hawkes, AL., Atherton, J., Taylor, CB., Scuffham, P., Eady, K., Houston-Miller, N. & Oldenburg, B. (2009). Randomised controlled trial of a secondary prevention program for myocardial infarction patients ('ProActive Heart'): study protocol. *BMC, Cardiovascular Disorders,* 9:6.

Holder, H. D., Lennox, R. D., & Blose, J. O. (1992). The Economic Benefits of Alcoholism Treatment: A Summary of Twenty Years of Research. *Journal of Employee Assistance Research*, 1:1, 63-82.

Lennox, RD., & Cecchini, MA. (2008). The NARCONON drug education curriculum for high school students: a non-randomized, controlled prevention trial. *Substance Abuse Treatment, Prevention, and Policy, 3:8.*

Lennox, RD, Dennis, ML, Ives, M, & White, M (2006). The construct and predictive validity of different approaches to combining urine and self-reported drug use measures among adolescents in substance abuse treatment, *American Journal on Addictions.*

Lennox, RD., Dennis, ML., Scott, C,. & Funk, R. (2005) Combining Psychometric and Biometric Measures of Substance Abuse. *Drug and Alcohol Dependence.*

Lennox, R.D., Zarkin, G., & Bray, J. (1996). Latent Measurement Models of Alcohol-Related Constructs. *Journal of Substance Abuse.*

Lennox, R.D. (1995). Outcomes of Mental Health and Drug Abuse Treatment. *The Journal of Mental Health Administration*, 22:3, 212-213.

Lennox, R. D., & Dennis, M. L. (1994). Measurement Error Issues in Substance Abuse Services Research: Lessons from Structural Equation Modeling and Psychometric Theory. *Evaluation and Program Planning.* 17:4, 399-407.

Lennox, R. D., Scott-Lennox, J. A., & Bohlig, E. M. (1993). The Cost of Depression-Complicated Alcoholism: Health Care Utilization and Treatment Effectiveness. *The Journal of Mental Health Administration*, 20:2, 138-152.

Lennox, R. D., Scott-Lennox, J. A., & Holder, H. D. (1992). Substance Abuse and Family Illness: Evidence from Health Care Utilization and Cost-Offset Research. *The Journal of Mental Health Administration*, 19:1, 83-95.

Lewis, A., & Khanna, V. (2014) *Surviving Workplace Wellness With Your Dignity, Finances and Major Organs Intact The Health Care Blog;* 1 edition

Linden, A., Butterworth, SW, & Prochaska, JO (2010). Motivational interviewing-based health coaching as a chronic care intervention. *Journal of Evaluation in Clinical Practice,* 16, 166-174.

Margolius, D., Bodenheimer, T., Bennett, H., Wong, J., Ngo, V., Padilla, G., & Thom, DH. (2012). Health Coaching to Improve Hypertension Treatment in a Low-Income, Minority Population. *Annals of Family Medicine,* 10: 99-205.

Merrill, RM, Bowden, DE, & Aldana, SG (2010). Factors Associated with Attrition and Success in a Worksite Wellness Telephonic Health Coaching Program. *Education for Health,* 23:3.

Nunnally, JC (1978). *Psychometric Theory.* (2nd Edition) New York: McGraw-Hill.

Nunnally, JC, & Bernstein, IH., (1994) *Psychometric Theory*, (3rd Edition). New York: McGraw-hill.

Patja, K., Pilvikki, A., TOkula, K., Kyto, J., Kuronen, R., Ovaska, T., Harno, K., Nenonen, M., Wiklund, T., Kettunen, R., & Talja, M. (2012). Health coaching by telephony to support self-care in chronic diseases: clinical outcomes from The TERVA randomized controlled trial. *BMC Health Services Research,* 12:147.

Pelletier, K. R. (1999). A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: 1995-1998 update (IV). *American Journal of Health Promotion, 13*, 333-345.

Pelletier, K. R. (2005). A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: Update VI 2000-2004. Journal of Occupational and Environmental Medicine*, 47*, 1051-1058.

Petrak & Herpertz, (2009). Treatment of *Depression in Diabetes: as update*, Curr Opinion Psychiatry, 22(2) 211-217.

Rost, K., Smith, J. L., & Dickinson, M. (2004). The effect of improving primary care depression management on employee absenteeism and productivity: A randomized trial. *Medical Care, 42*, 1202-1210.

Schneider, JI., Hashizume, J., Heak, S., Maetani, L., Rude Ozaki, R., & Leong Watanabe, D. (2011). Identifying challenges, goals and strategies for success for people with diabetes through life coaching. *Journal of Vocational Rehabilitation,* 34: 129-139.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

Wang, P. S., Simon, G. E., Avorn, J., Azocar, F., Ludman, E. J., McCulloch, J., Petukhova, M. Z., & Kessler, R. C., (2007). Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes. *Journal of the American Medical Association, 298*, 1401-1411.

Wang, P. S., Simon, G. E., & Kessler, R. C. (2008). Making the business case for enhanced depression care: the National Institute of Mental Health-Harvard work outcomes research and cost-effectiveness study. Journal of Occupational and Environmental Medicine*, 50*, 468-475.

Table 1: The Coaching Version of the 25-items Workplace Outcome Suite	
Workplace Outcome Suite (Coaching)	©Chestnut Global Partners

General Instructions. Below is a series of statements that refer to aspects of your work and life experience *that may be affected by the **coaching health and lifestyle issue** you want to address with the coaching program during the past 30 days*. Please read each item carefully and answer as accurately as you can.

	*Instruction for items 1-4. Please report for the period of the past thirty (30) days the total number of **hours** your coaching health and lifestyle issue:*	Number of Hours
	Absenteeism	
1	caused you to miss work altogether.	
2	made you late for work.	
3	caused you to take off early.	
4	pulled you away from your normal work location while still at work	
5	required you to be on the phone, e-mail or internet while at work.	

	Instruction for items 6-25. The following statements reflect what you may do or feel on the job or at home. Please indicate the degree to which you agree with each of the statements for the past thirty (30) days. Use the 1-5 response key to the right.	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
	Presenteeism					
6	I had a hard time doing my work because of my health and lifestyle issue	1	2	3	4	5
7	My health and lifestyle issue kept me from concentrating on my work	1	2	3	4	5
8	Because of my *coaching* health issue I was not able to enjoy my work	1	2	3	4	5
9	My health and lifestyle issue made me worry about completing my tasks	1	2	3	4	5
10	I could not do my job well because of my health and lifestyle issue.	1	2	3	4	5
	Work Engagement					
11	I feel stimulated by my work.	1	2	3	4	5
12	I often think about work on my way to the work site.	1	2	3	4	5
13	I feel passionate about my job.	1	2	3	4	5
14	I am often eager to get to the work site to start the day.	1	2	3	4	5
15	I often find myself thinking about my work at home.	1	2	3	4	5

	Life Satisfaction					
16	My life is nearly perfect.	1	2	3	4	5
17	I am not very satisfied with my life as a whole.	1	2	3	4	5
18	So far, my life seems to be going very well.	1	2	3	4	5
19	There isn't anything about my life that I would change if I could.	1	2	3	4	5
20	I am very disappointed about the way my life has turned out.	1	2	3	4	5
	Workplace Distress					
21	I often feel anxious at work.	1	2	3	4	5
22	Thinking about being at work makes me upset.	1	2	3	4	5
23	I am unhappy most of time at work.	1	2	3	4	5
24	I dread going into work	1	2	3	4	5
25	I can't wait to get away from work.	1	2	3	4	5
	Validation Questions					
26	I have a hard time getting out of bed.	1	2	3	4	5
27	I often feel sad.	1	2	3	4	5
28	I keep falling behind schedule at work.	1	2	3	4	5
29	I am rarely late for work.	1	2	3	4	5
30	I often get home late from work.	1	2	3	4	5

Table 2: Descriptive Statistics of the CGP Outcome Suite Candidate Items				
	Abbreviated Item text	M	SD	*N*
1	…caused you miss to work altogether.	.62	3.66	618
2	…made you late for work.	.23	2.01	615
3	…caused you to take off early.	.22	1.24	618
4	…pulled you away from your normal work location while still	.29	3.44	616
5	…required you to be on the phone, e-mail, or internet while at	1.36	8.84	612
6	I had a hard time doing my work because of my personal	1.52	0.96	609
7	My personal problems kept me from concentrating on my	1.54	0.96	610
8	Because of my personal problems I was not able to enjoy my	1.58	1.00	610
9	My personal problems made me worry about completing my	1.52	0.94	607
1	I could not do my job well because of my personal problems.	1.40	0.85	605
1	I feel stimulated by my work.	4.04	0.97	602
1	I often think about work on my way to the work site.	4.03	1.07	599
1	I feel passionate about my job.	4.15	0.98	603
1	I am often eager to get to the work site to start the day.	3.70	1.09	601
1	I often find myself thinking about my work at home.	3.76	1.14	603
1	My life is nearly perfect.	3.44	1.05	610
1	I am not very satisfied with my life as a whole.	1.96	1.07	609
1	So far, my life seems to be going very well.	4.05	0.90	612
1	There isn't anything about my life that I would change if I	2.64	1.23	606
2	I am very disappointed about the way my life has turned out.	1.60	0.93	612
2	I often feel anxious at work.	2.66	1.25	601
2	Thinking about being at work makes me upset.	2.00	1.18	601
2	I am unhappy most of time at work.	1.83	1.11	600
2	I dread going into work	1.85	1.15	596
2	I can't wait to get away from work.	2.28	1.25	601
V	I have a hard time getting out of bed.	2.04	1.16	290
V	I often feel sad.	1.76	1.03	292
V	I keep falling behind schedule at work.	1.97	1.20	291
V	I am rarely late for work.	4.24	1.19	293
V	I often get home late from work.	3.41	1.29	292

	Table 3: Reliability Analysis of the CGP Outcome Suite Coaching Items		
	Abbreviated Item text	Item-Total r	α if Deleted
1	caused you miss to work altogether.	NA[1]	NA[1]
2	made you late for work.	NA[1]	NA[1]
3	caused you to take off early.	NA[1]	NA[1]
4	pulled you away from your normal work location while still at	NA[1]	NA[1]
5	required you to be on the phone, e-mail, or internet while at	NA[1]	NA[1]
6	I had a hard time doing my work because of my personal	0.84	0.92
7	My personal problems kept me from concentrating on my	0.84	0.92
8	Because of my personal problems I was not able to enjoy my	0.80	0.93
9	My personal problems made me worry about completing my	0.85	0.92
1	I could not do my job well because of my personal problems.	0.86	0.92
1	I feel stimulated by my work.	0.64	0.77
1	I often think about work on my way to the work site.	0.60	0.78
1	I feel passionate about my job.	0.71	0.75
1	I am often eager to get to the work site to start the day.	0.66	0.77
1	I often find myself thinking about my work at home.	0.46	0.83
1	My life is nearly perfect.	0.69	0.74
1	I am not very satisfied with my life as a whole.	0.56	0.78
1	So far, my life seems to be going very well.	0.68	0.75
1	There isn't anything about my life that I would change if I	0.51	0.81
2	I am very disappointed about the way my life has turned out.	0.58	0.78
2	I often feel anxious at work.	0.58	0.90
2	Thinking about being at work makes me upset.	0.79	0.85
2	I am unhappy most of time at work.	0.76	0.86
2	I dread going into work	0.82	0.85
2 5	I can't wait to get away from work.	0.73	0.87
	Note: alpha coefficients for scales are: Presenteeism=938; Work Engagement=.816; Life Satisfaction=.807; Workplace Distress=.889		

	Table 4: Standardized Coefficients from the four-factor Confirmatory factor Analysis of the CGP Outcome Suite Coaching Items	
	Item Text	Standardized Loading
1	caused you miss to work altogether.	NA[1]
2	made you late for work.	NA[1]
3	caused you to take off early.	NA[1]
4	pulled you away from your normal work location while still at	NA[1]
5	required you to be on the phone, e-mail, or internet while at	NA[1]
6	I had a hard time doing my work because of my personal	0.87
7	My personal problems kept me from concentrating on my work.	0.87
8	Because of my personal problems I was not able to enjoy my	0.83
9	My personal problems made me worry about completing my	0.88
1	I could not do my job well because of my personal problems.	0.90
1	I feel stimulated by my work.	0.81
1	I often think about work on my way to the work site.	0.48
1	I feel passionate about my job.	0.90
1	I am often eager to get to the work site to start the day.	0.90
1	I often find myself thinking about my work at home.	0.35
1	My life is nearly perfect.	0.77
1	I am not very satisfied with my life as a whole.	0.65
1	So far, my life seems to be going very well.	0.77
1	There isn't anything about my life that I would change if I could.	0.58
2	I am very disappointed about the way my life has turned out.	0.66
2	I often feel anxious at work.	0.60
2	Thinking about being at work makes me upset.	0.82
2	I am unhappy most of time at work.	0.83
2	I dread going into work	0.90
2 5	I can't wait to get away from work.	0.80

Table 5: Correlations Among the four latent factors				
	Presenteeism	Work Engagement	Life Satisfaction	Workplace Distress
Presenteeism	1			
Work Engagement	-0.27	1		
Life Satisfaction	-0.37	0.35	1	
Workplace Distress	0.32	-0.45	-0.33	1

Figure 1:

Theoretical Structure of the Effect-Indicator Measurement Models of
the Presenteeism, Work Engagement, Life satisfaction and
Workplace Distress Scales