IJHP

INTERNATIONAL JOURNAL OF HEALTH & PRODUCTIVITY

SPECIAL EDITION • December 2018 • Volume 10, Number 2

FROM THE EDITOR'S DESK
***International Journal of Health & Productivity (IJHP)* Expands its Thought Leadership Role**

ARTICLES
EAP Works: Global Results from 24,363 Counseling Cases with Pre-Post Data on the Workplace Outcome Suite© (WOS)

Demonstrating Value: Measuring Outcome & Mitigating Risk: FOH EAP Study Utilizing the Workplace Outcome Suite©

Development and Validation of a Critical Incident Outcome Measure

Validation of the 5-item Short Form Version of the Workplace Outcome Suite©

Measuring Coaching Effectiveness: Validation of the Workplace Outcome Suite© for Coaching

A PUBLICATION OF THE INSTITUTE FOR HEALTH AND PRODUCTIVITY MANAGEMENT

Every workplace is unique. Behavioral health solutions should be too.

Workplace well-being is more important than ever but finding and choosing the best solutions to protect and cultivate it is challenging.

At R3, we partner with you to understand the unique needs of your organization and co-create customized solutions that fit perfectly.

People are the heart of your business. Learn how our experts can help your people and your business thrive.

Visit us at **R3c.com or Call 866-927-0184**

R3 CONTINUUM

Crisis Response I Consulting I Evaluations

PUBLISHING, EDITORIAL MANAGEMENT AND PRODUCTION SERVICES
Douglas Murphy Communications, Inc.
P.O. Box 71895
Richmond, VA 23255-1895
Tel. 804.387.7580 **Fax**. 703.997.5842
www.douglasmurphy.com

Grant W. Murphy, President,
grant.murphy@douglasmurphy.com

Barry Barnum, Editorial Director,
barry.barnum@douglasmurphy.com

Alyson Plante, Graphic Director,
alyson@plantecreativestudio.com

CUSTOMER SERVICE AND SUBSCRIPTIONS
Subscription Prices, USA/Canada: print and online, $140 per year. Other countries: print and online, $195 per year. To order or change a subscription, contact Douglas Murphy Communications, Inc., Customer Service, P.O. Box 71895, Richmond, VA 23255-1895 or e-mail barry.barnum@douglasmurphy.com For prompt and accurate service, please enclose your current address label containing your unique subscription number.
Reprints, International Journal of Health & Productivity article reprints are available in any quantity. To obtain information or to place an order, contact Douglas Murphy Communications, Inc., Customer Service, P.O. Box 71895, Richmond, VA 23255-1895 or e-mail barry.barnum@douglasmurphy.com

Electronic Availability, Online acces to the International Journal of Health & Productivity is included with regular individual subscriptions. To activate individual online subscriber-level access, or inquire about institutional or corporate access contact Douglas Murphy Communications, Inc., Customer Service, P.O. Box 71895, Richmond, VA 23255-1895 or e-mail barry.barnum@douglasmurphy.com

MANUSCRIPT SUBMISSIONS
The International Journal of Health & Productivity is a peer-reviewed publication. All articles must be submitted through our editorial submission system.

For more information, please contact barry.barnum@douglasmurphy.com

INTERNATIONAL JOURNAL OF HEALTH & PRODUCTIVITY

SPECIAL EDITION

THE INTERNATIONAL JOURNAL OF HEALTH & PRODUCTIVITY (ISSN 2164-4373) is published by the Institute for Health and Productivity Management (IHPM), with corporate offices located at 17470 N Pacesetter Way, Scottsdale, AZ 85255 USA. Periodical postage paid at Richmond, VA, and at additional mailing offices.

William B. Bunn, III, MD, JD, MPH Editor-in-Chief, *wiliam.bunn@navistar.com*
Formerly Vice President, Health, Safety, Security, and Productivity, Navistar International Professor, Northwestern University School of Medicine

Sean Sullivan, JD President and CEO, *sean@ihpm.org*

Bill Williams III, MD, FAAFP Senior Vice President, *bill@ihpm.org*

Deborah Love Executive Vice President and COO, *deborah@ihpm.org*

Edward Jones, PhD Senior Vice President of Strategic Planning

Steve Priddy Executive Director, Value-Based Health, *steve@ihpm.org*

William B. Bunn, III, MD, Consulting Medical Director

TABLE OF CONTENTS

FROM THE EDITOR'S DESK

International Journal of Health & Productivity (IJHP) Expands its Thought Leadership Role
W. B. Bunn, III. **6**

ARTICLES

EAP Works: Global Results from 24,363 Counseling Cases with Pre-Post Data on the Workplace Outcome Suite© (WOS)
Mark Attridge, PhD, MA;
David Sharar, Ph.D;
Gregory DeLapp, MHS, CEAP;
Barbara Veder, MSW, RSW. **7**

Demonstrating Value: Measuring Outcome & Mitigating Risk: FOH EAP Study Utilizing the Workplace Outcome Suite©
Jeffrey Mintzer LICSW, CEA;
Veronica Y. Morrow , MSW, LCSW-C, CEAP;
Melissa Tamburo, Ph.D, LCSW-C;
David Sharar Ph.D;
Patricia Herlihy Ph.D., RN. **28**

Development and Validation of a Critical Incident Outcome Measure
Richard D. Lennox Ph.D;
Dave Sharar Ph.D;
Patricia A. Herlihy Ph.D, RN;
Matthew Mollenhauer, MS **35**

Validation of the 5-item Short Form Version of the Workplace Outcome Suite©
Richard D. Lennox, Ph.D,
David Sharar, Ph.D,
Eileen Schmitz,
David B. Goehner, LCSW **49**

Measuring Coaching Effectiveness: Validation of the Workplace Outcome Suite© for Coaching
Richard D. Lennox, Ph.D;
David Sharar, Ph.D; Francine Miller **62**

www.IHPM.org

© 2018 Copyright Institute for Health and Productivity Management (IHPM). Printed in the U.S.A. All rights reserved under International and Pan-American Copyright Convention. No part of this publication may be reproduced or transmitted in any way or by any means, electronic or mechanical, without prior written permission of the Institute for Health and Productivity Management (IHPM). Opinion expresed in the publication are those of the experts and authors and not necessarily those of the publisher or editorial advisory board members. Members are not responsible for the content of the publication.

Mental illness affects 1 in 5.
It's time to help 100% of your people.

Dear business leaders:

One in five people worldwide are suffering from mental health challenges* and almost a quarter of adults are not physically active enough to prevent illness. One in three employees are distracted by finances at work, and a further eight in 10 are disengaged from their jobs. In short, the majority of your workforce is suffering from one or more wellness issues at any one time, potentially leading to long-term absence or high employee churn.

The response to all these challenges is that employers like you have a duty of care, not just to their people, but to their shareholders and national economies. To invest in your employees' total well-being and help people build and sustain mental, physical, social and financial wellness. Anytime, anywhere.

Traditionally, employee support has been reactive, supporting people through an Employee Assistance Program (EAP). But a typical Employee Assistance Program sees only up to 5% usage. Even at Morneau Shepell - a market-leading EAP provider - where we see double this usage, this means that you are only supporting a fraction of your people. So, here's the real question: How do we help the other 90-95%? And, what would happen to your business if we enabled you to provide that support?

With our acquisition of LifeWorks, Morneau Shepell is combining industry leading clinical support and treatment with a solid focus on engagement and user experience. Between us, we have almost 100 years of employee assistance and clinical expertise and now, driven by digital innovation we can address the total well-being of your people at every stage of their employee journey.

Our mobile-first platform provides the opportunity to engage users across the health and well-being spectrum through regular personalized interactions, rewards, recognition, and communication. With usage rates of 50-65% we can deliver high participation, high persistency and better outcomes. Where traditional employee assistance acts as a safety net in a crisis, our proactive approach drives persistent participation and preventative wellness.

We believe it's not enough to serve employees in times of crisis alone, because reactive assistance does not address total employee well-being. Our global reach puts an innovative platform that employees love to use in the hands of many more people and allows you to fulfil your purpose of improving business while we go about improving lives.

Your people matter. Make them feel loved.

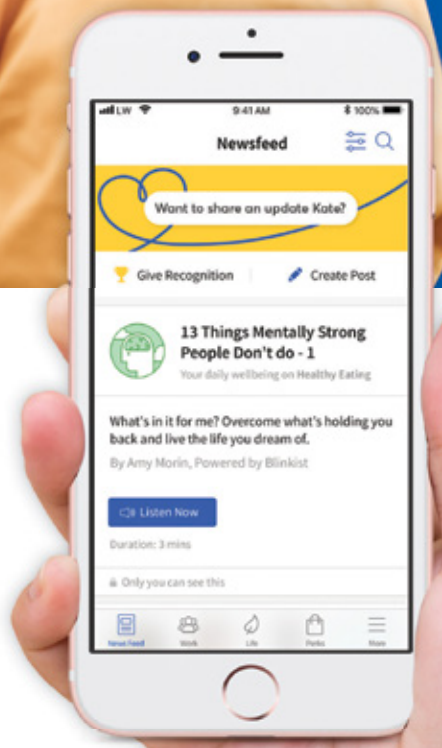
Stephen Liptrap
President and CEO
Morneau Shepell

*World Health Organisation

LifeWorks
by Morneau Shepell

LifeWorks
by Morneau Shepell

100% well-being for 100% of your workforce



Traditional employee assistance programs have a utilization rate of just 5%, only reaching the few employees who raise their hand and ask for help.

With our tech-enabled service, you can empower 100% of your workforce to take control of their mental, physical, social and financial health. Give your employees the total well-being solution that people everywhere love to use.

- Employee assistance
- HR communication and community
- Rewards and recognition
- Perks and savings
- Physical wellness
- Analytics and insights

LifeWorks.com | totalwellbeing@lifeworks.com

FROM THE EDITOR

International Journal of Health & Productivity Further Extends its Thought Leadership Role

Willam B. Bunn III, MD, JD, MPH
Editor-In-Chief

This Special Edition marks IHPM's increased attention to the area of health that has the greatest impact worldwide on functional impairment and, consequently, on lost productivity at work.

This very Special Edition of IHPM's *International Journal of Health & Productivity (IJHP)* stakes out new ground in the Institute's domain of measuring outcomes to know the effectiveness of actions to improve employee health and wellbeing.

IHPM has built a strong connection with the behavioral/mental health community over the years – recognizing the increasingly critical role that providers of these services play in maintaining or improving the total health and wellbeing of the work force. And drawing on its own considerable field research with employers, the Institute long has appreciated the huge impact of health issues such as depression – and psychological factors like stress – on functionality and work performance.

The Workplace Outcomes Suite (WOS) was developed by Chestnut Global Partners (CGP) as a self-report instrument to measure the effectiveness of Employee Assistance Program (EAP) counseling services in addressing these behavioral/mental health issues for employees using those services. In the interest of the entire EAP field, CGP distributed the WOS free of charge to other providers of these services, in exchange for sharing the data from their use of the instrument.

The articles collected here in this special edition of the *IJHP* all report positively on results obtained from use of the WOS with working populations, and include:

- A large global study by Attridge, Sharar, DeLapp and Veder of more than 24,000 counseling cases in 26 countries, though concentrated in the US (79%) and China (15%), reported a drop in monthly missed work time related to the personal issue involved in the counseling from 7.4 to 3.9 hours – or nearly 50 percent;
- A study by Lennox, Sharar, Schmitz and Goehner confirmed that the 5-item version of the original 25-item WOS – with items corresponding to measures of presenteeism, work engagement, life satisfaction and workplace distress as well as a separate measure based on total and partial days absent from work – could be used to approximate the 25-item version without excessive loss of reliability, validity or sensitivity;
- A study of federal government employees by Mintzer, Morrow, Tamburo, Sharar and Herlihy and using the 5-item version of the WOS, reported significant improvements in absenteeism, workplace distress, life satisfaction and workplace presenteeism;
- A study by Lennox, Sharar and Miller reported that the coaching version of the WOS – developed also by Chestnut Global Partners to measure the effectiveness of coaching interventions for a range of workplace behavioral health issues – was employed in disease management services for depression and diabetes, and reliably tested the effectiveness of coaching without creating concern for significant measurement error;
- And a study by Lennox, Sharar, Herlihy and Mollenhauer introduced the Critical Incident Outcome Measurement Scale (CIOM) to measure the success of response to traumatic incidents in the workplace, and beta tested the tool with a pilot sample of respondents; the goal was to select the best single indicator for an abbreviated version of the CIOM to enhance its suitability for applied applications – recognizing that collecting data is not as important as attending to the needs of employees affected by a traumatic incident.

This Special Edition – made possible by financial support from global behavioral health leader Morneau Shepell – marks IHPM's increased attention to the area of health that has the greatest impact worldwide on functional impairment and, consequently, on lost productivity at work. Along with chronic pain, it also is the leading reason for early exit from the work force because of disability – a loss of the experienced and skilled workers that economies can ill afford. The success of EAP services in ameliorating these behavioral/mental health issues is vital to the future performance of the work force, and measurement is key to that success.

William B. Bunn III, MD, JD, MPH
Editor-in-Chief
International Journal of Health & Productivity

EAP Works: Global Results from 24,363 Counseling Cases with Pre-Post Data on the Workplace Outcome Suite© (WOS)

Mark Attridge, PhD, MA; David Sharar, Ph.D;
Gregory DeLapp, MHS, CEAP; Barbara Veder, MSW, RSW

ABSTRACT

The Workplace Outcome Suite© (WOS) is a self-report instrument designed to evaluate the effectiveness of employee assistance program (EAP) counseling services from the perspective of the employee user of the service. More than 30 EAPs collected longitudinal data on all versions of the WOS from 2010 to 2018 and voluntarily submitted their raw data to Chestnut Global Partners for analysis. The 24,363 employees in this aggregated sample represent 26 different countries, but most of the cases were from the United States (79%) and China (15%). The typical EAP case in this data set was a female, age 38, and was a self-referral into an external vendor of EAP services seeking help for a mental health concern. Outcomes were collected at the start of counseling and again approximately three months later. Evidence of the psychometric validity and test-retest reliability for all five WOS measures was found in correlational tests. Other tests of the change in outcomes from before to after use of EAP counseling found large effects on work presenteeism and life satisfaction (ηp^2 = .24 and .19), a medium-size effect on work absenteeism (ηp^2 = .13), and small effects on both workplace distress and work engagement (ηp^2 = .05 and .04). Although most EAP cases had no absence from work either before counseling or at follow-up (58% and 78%, respectively), the average amount per case per month of missed work due to the personal concern was reduced from 7.4 hours before to 3.9 hours after use of the EAP. Weak findings on moderator tests determined EAP counseling was effective to a similar degree on WOS outcomes across contextual factors of client age, sex, country, referral type, clinical concerns, industry of the employer, and delivery models for providing employee assistance counseling (i.e., external vendors, internal staff programs and hybrid models). As an alternative to the fill-in-the-blank response format requiring a specific number of hours, a modified version of the work absenteeism single item is offered that has a 5-point scale with normative levels of absence hours obtained from the Pre EAP use global data that define each of the 1-5 rating options. More details and related findings are presented in the *Workplace Outcomes Suite 2018 Annual Report* from Chestnut Global Partners.

KEY WORDS:
Employee Assistance Program, Counseling, Outcomes, Absenteeism, Presenteeism, Engagement, Life Satisfaction, Workplace, Depression, Stress, Longitudinal

INTRODUCTION

Many working adults suffer from emotional issues, family and home life conflicts, mental health concerns, substance abuse problems, and other health disorders that can interfere with their health and work performance. Recent national epidemiologic survey data indicates nearly one in every five working adults in the United States meets clinical criteria for a behavioral health disorder.[1] StayWell examined data from their health risk appraisal (HRA) surveys and health care claims data from more than 21,000 employees from multiple employers in U.S. and concluded behavioral health issues are common among working adults: 21 percent were at moderate to high risk for depression; 14 percent were at high risk for stress; and 10 percent were at moderate to high risk for alcohol misuse.[2] Thus, there is a need for services to support the behavioral health risks of employees.[3]

One way to respond to at-risk and

distressed employees is to offer an employee assistance program (EAP). These are employer sponsored programs designed to help individuals resolve acute but modifiable behavioral health issues.[4] The EAP is often used for assistance with mild to moderate problems that cause acute stress, such as marital relationship issues, family concerns, work problems, and legal or financial concerns. Individuals in need of treatment of more serious mental health and substance abuse disorders (such as anxiety, depression, alcohol or drug misuse) are provided appropriate referrals to qualified providers and follow-up from the EAP. The general goal of EAPs is to have a positive effect on restoring the health and well-being of the employee, which in turn results in reduced long-term healthcare expenditures and a return to higher productivity. Indeed, what sets Employee Assistance (EA) services apart from other mental health services is its focus on providing brief treatment and practical resources that improve the work performance of employee clients.[5-9] Thus, EAPs are designed to support employees and restore their work performance.

Employee Assistance Programs have provided counseling and speciality support services to employers for many decades in North America and the field continues to expand globally.[10,11] In the 1980s and early 1990s, only a third of employers in the United States offered an EAP.[12] Today, 40+ years later, the vast majority of large and medium size employers in the U.S. now offer an EAP, but having access to employee assistance programs varies by the size of employer. In the public sector in the U.S., 100 percent of federal government employees, 86 percent of state government employees and 71 percent of local government employees have access to an EAP.[13] Also in 2016, according to the U.S. government's national survey of compensation, in the private sector, 85 percent of employers with 500 or more workers had an EAP, with lower prevalence rates as size of the company decreased: 68 percent at employers with 100 to 499 employees; 44 percent at employers with 51 to 99 employees; and only 27 percent of those in small businesses with less than 50 workers.[13] In another example from the private sector, WorldatWork, a human resources organization, surveyed 867 of its member companies in 2017 and found 97 percent of large companies and 88 percent of smaller size companies (under 500 employees) have an EAP.[14] In this same survey, EAP was offered at more employers than any of the 14 different employee health and wellness benefits in the survey. Thus, most employers in the U.S. now sponsor EAP as an employee benefit.

When EAP counseing is provided with adherence to basic quality standards the results are usually positive.[5] There is considerable evidence from reviews of studies conducted in North America[15-18] and in Europe[19-21] that brief counseling provided by EAPs typically reduces stress, improves symptoms of behavioral health problems and restores higher work functioning. This point was further documented in the landmark study by the National Beahvioral Consortium collecting data from 82 different vendors of EAP services with a combined customer base of more than 35,000 client companies and 164 million total covered lives in the United States, Canada, and 10 other countries.[22] Based on the averages of the follow-up surveys of many vendors and representing more than 100,000 individual EAP cases combined, the following facts were obtained about the level of user satisfaction and program impact as an industry: 94 percent of cases were satisfied with EAP services; 86 percent of cases had improved in the issue leading to use of the EAP; 73 percent of cases had improved work productivty (reduced presenteeism); and 64 percent of cases had improved work absenteeism. In general, then, EAPs usually are effective measured by both clinical and work performance outcomes.

The NBC study also found a considerable range of average outcomes obtained by the different EAP vendors for their clients with some EAPs much lower and some higher than the industry average. More importantly, from a comparative standpoint above industry-average results were derived from using many different tools to measure the outcomes. Less than half of EAP vendors in 2011 (42%) were using a standardized research-validated survey outcome tool while the majority of the EAPs instead had developed their own items and survey tools for assesing outcomes of the EAP services.[24] Such "home-grown" tools may be useful for that particular vendor,

but they also have unproven reliability and validity. Moreover, it makes it challenging for the purchasers of EAP services to fairly compare vendors on outcomes when vendors are not using the same metrics.[23]

A lack of common measures and industry benchmarks to assess and compare the effectiveness of counseling on workplace outcomes is a concern when producing such outcomes is highly valued by the purchasers of EAP services. For example, a 2018 survey conducted by the Employee Assistance Society of North America (EASNA) asked 155 senior EAP professionals and purchasers of EAPs which factors were important to the decision of selecting an EAP.[24] The study found that "evidence of user outcomes in improved workplace performance (less absenteeism, presenteeism, turnover)" was rated as either *high* or *very high* in importance as a factor in selecting a vendor by 62 percent of the sample. Thus, finding a way to determine which EAPs are the most effective and have superior outcomes is of keen interest to most purchasers.

THE WORKPLACE OUTCOME SUITE

In this context, the Workplace Outcome Suite© (WOS) was developed in 2010 by the Division of Commercial Science at Chestnut Global Partners (CGP) to provide a scientific, objective measure of these outcomes.[25] The WOS is a self-report instrument designed to evaluate the effectiveness of employee assistance program (EAP) counseling services from the perspective of the employee user of the service. It is completed at two points in time, first at the start of counseling and then again at a longitudinal follow-up several months after the counseling is completed, with a recommended follow-up period of 90 days. The instrument is a measure of change in four key aspects of workplace functioning: a*bsenteeism, presenteeism, work engagement,* and *workplace distress.* As a reflection of general effectiveness on personal issues, the WOS also measures overall level of *life satisfaction* for users of EAP services. These items are answered on a 1-5 rating Likert-type rating scale as follows: 1 = *strongly disagree*; 2 = *agree*; 3 = *neutral*; 4 = *agree*; 5 = *strongly agree*. However, the response to the Absenteeism item is a fill in the blank format with a specific number of hours of absence requested.

Versions of the WOS

The original 25-item WOS has five item entries for each of these constructs. On the 25-item version all scales (except for Absenteeism) have slightly different wording of essentially the same question – a classical psychometric theory called "effect-indicator." In 2012, CGP developed the 5-item version at the request of WOS users to reduce the amount of time it takes for clients to complete the tool and in the process increasing response rates.[26] This was done by selecting the single "best" item from the Presenteeism, Work Engagement, Life Satisfaction, and Work Distress scales of the 25-item version. A different process was done for Absenteeism in which a new single item was created featuring the definitions of absence specified in the first three items of the full 5-item Absenteeism scale. These three items had defined absence as missing work, being late to work or leaving work early. There also is a 9-item version that includes all five of the original absenteeism questions combined with the four single-items from the 5-item brief version. The three WOS measures with full instructions and response options are provided in the Appendix. Today most EAPs have migrated to using the 5-item brief version (see items below).[EN1]

- **Absenteeism** item: *"For the period of the past 30 days, please total the number of hours your personal concern caused you to miss work. Include complete eight-hour days and partial days when you came in late or left early. _____"*
- **Presenteeism** item: *"My personal problems kept me from concentrating on my work."*
- **Workplace Distress** item: *"I dread going into work"*
- **Work Engagement** item: *"I am often eager to get to the worksite to start the day"*
- **Life Satisfaction** item: *"So far, my life seems to be going very well"*

WOS Endorsed as EAP Industry Standard

The WOS is currently the only publicly available instrument psychometrically validated and tested for use in EAP settings. It is available to use at no cost with the signing of a license agreement (go the EAPA website: bit.ly/WOS-License-Agreement).

The WOS presents a *single* tool that can be used across the EAP spectrum for demonstrating effectiveness, and in turn

furthering the field as opposed to a patchwork of measurement tools that have not advanced the EA field. In 2017, the largest industry group – the Employee Assistance Professionals Association (EAPA) – with more than 5,000 members worldwide, endorsed the WOS as an EAP Best Practice for measuring and evaluating work-related outcomes of services provided by EAPs. With access to thousands of EAP professionals across the globe and a deep commitment to the highest standards of EA practice, EAPA believes the WOS, when properly implemented, can bring clarification to the field's value proposition and need for greater evidence of program effectiveness. This collaboration has been successful as evidenced by having more than 600 different EAPs signing license agreements to use the WOS. This interest demonstrates that greater numbers of EAPs are finding the WOS to be highly effective in demonstrating improvement with their EAP counseling clients.

Studies of EAP Outcomes Using the WOS

Since its introduction in 2010, numerous applied studies of EAPs have featured data from the WOS. Several very large organizations in the private sector have used the WOS to assess the effectiveness of their EA programs:

1) Global manufacturing company, Caterpillar has used the WOS to examine EAP outcomes at the overall level, for certain worksites, and to compare EAP counseling provided on-site at the workplace versus counseling provided by off-site locations;[27,28,29]
2) ConocoPhillips, in the oil and gas industry explored the workplace impact of EAP using the WOS;[30]
3) DuPont used the WOS in a year-long demonstration project to document sucessful expansion of EA services to its worksites located in many different countries around the globe;[31,32]
4) Public sector, Federal Occupational Health – the EAP for the federal government that serves more than 1.1 million employees – has recently evaluated the program with WOS data,[33,EN2]
5) The internal staff model EAP at Partners HealthCare System in Boston has included the WOS as part of its ongoing quality improvement initiatives;[34]
6) The WOS also has been used by the Life Solutions is the internal staff model EAP for the University of Pittsburgh Medical Center and also a provider of services to local employers.[35]

There are examples of EAPs operating in countries outside the U.S. using the WOS to measure the impact of their services, including:

1) Benestar EAP in New Zealand;[35]
2) Chestnut Global Partners EAP in Brazil;[36]
3) Chestnut Global Partners EAP in China;[37-39]
4) Chestnut Global Partners EAP in Russia;[40]
5) Hellas EAP in Greece.[41]

The 2018 WOS Annual Report profiles 13 different EAP vendors, internal EA programs and large employers with hybrid EA programs.[38] Each of the profiles shares information on the operational practices for collecting WOS data and case stories of the business impact of having credible workplace outcome data available. In addition to most of the examples noted above, several other EAP vendors in the U.S. market are profiled in the 2018:

1) Cascade Centers;
2) Concern EAP;
3) Empathia EAP;
4) KGA EAP as well as Homewood Health in Canada which is using the WOS to assess outcomes in a depression care specialty program.

All of these applied studies of EAP noted above with WOS data featured a single-group longitudinal research design that included only the users of EA services and no comparison group of non-users of the EAP. These projects were not supported by external research grant funding. In contrast, a rare quasi-experimental study was conducted for a statewide internal EAP program serving government employees in the state of Colorado.[42] This study was funded by a large grant from the Employee Assistance Research Foundation. The study compared experience of users of EAP counseling (n = 156) against a group of employees from the same organization (n = 188) who did not use the EAP but who were matched on level of personal distress, social

support, and demographic characteristics to the EAP cases. The 5-item brief version of the WOS was assessed at the start of the case for EAP users and again as a follow-up that ranged between 2 and 12 months later. The EAP group averaged 4 months at follow-up and the comparison group 8 months at follow-up.

As expected, the two groups did not differ at baseline on any of the WOS measures. Significant differences were found for the extent of change over time between the two groups. Results showed EAP users had reduced their level of presenteeism on the WOS by 21 percent, which was significantly more than the 11 percent reduction found in the comparison group who had not been treated by the EAP. The EAP user group also had a decrease in absenteeism (from 15.0 hours at baseline to 10.7 hours at follow-up), whereas the comparison group *increased* in absenteeism (from 13.0 hours at baseline to 16.9 hours at follow-up). For context, the typical employee at this public sector organization had about 9 hours of absence per month, which is less than both groups of distressed employees at baseline.

Further analysis of the same groups determined that the EAP cases had significantly more improvement over time in mental health outcomes (i.e., symptoms of depression and anxiety) than the matched control employees and these improvements in the mental health clinical outcomes were positively linked to the improvements in both work absenteeism and presenteeism outcomes.[43] Thus, those employees who improved most in their mental health after the use of the EAP counseling also improved the most in their absenteeism hours and presenteeism levels at work. A third study of this project examined the actual timesheets of recorded absenteeism data, which verified the same relatively better outcome of missed time away from work for the EAP group compared to the matched non-EAP user group.[44] Although based on a small sample size, this study from Colorado provides compelling evidence of the superior impact that brief counseling from EAPs can have on workplace performance outcomes over what is normally experienced by employees in distress but do not get assistance from the employee assistance program.

Overview of this Study

The goal was to use this very large aggregated sample to answer the following questions about the impact of EAP counseling and the appropriateness and viability of the Workplace Outcome Suite measurement tools for EAPs:

- What is the psychometric validity and reliability of the five WOS measures?
- What is the extent of improvement in WOS outcomes after use of EAP counseling? And naming the five WOS outcome constructs, which areas have the largest improvement?
- Is the level of improvement in WOS outcomes different for certain factors of the client (sex, age, global location), the clinical experience (referral and clinical problem types), and the employer context (industry and EAP delivery model)?
- Which outcomes on the five constructs featured in the WOS are most relevant to EAP counseling?

METHDOLOGY

Study Design and Sample

Longitudinal Repeated Measures Design. Employee users of the EAP completed the WOS *before* introducing the EAP counseling intervention and then completed the WOS again at several months *after* the intervention. A 90-day follow-up time frame was recommended by CGP for administering the "post" measure rather than doing it immediately after the last EAP visit because it cannot determine if such improvement persists after counseling has ended. The use of a three-month follow-up period was intended to confirm that improvement on WOS constructs experienced at the end of counseling were then maintained over a longer time period. This data measurement approach with the WOS likely represents a more conservative set of results than if outcomes were assessed at the end of treatment.

For example, in a demonstration study conducted in the state of Vermont, U.S., of behavioral health risk screening and enhanced counseling from the EAP staff, users of counseling had a 40 percent reduction in the amount of lost productive time at work in the past month (a combined measure of

absenteeism hours and hours derived from work productivity levels on 0-10 rating scale) from the start of the case to the end of the case (which on average was after four sessions of counseling).[45] The follow-up data (three and six months after the end of the case) showed a 29 percent average reduction from the level at the start of case. Thus, the degree of workplace outcome improvement at the follow-up slightly was less than when assessed at the end of the case, even though both post time period results were a significant change from the level of work performance deficit reported at the start of treatment.

No Comparison Group. Having only the intervention group experiencing EAP counseling with no comparison group of employees equally distressed and not receiving EAP counseling, is known as a "Correlational" or "Before/After" single-group study. This kind of study design can identify *if* employees improved at work after EAP counseling, but it cannot prove EAP counseling was the most important causal factor in this improvement. Although less rigorous than a quasi-experimental or true experimental research study designs with random assignment of participants to treatment and control groups, the single group design is typical of almost all studies of the users of voluntary employee health and wellbeing benefits provided in real-life settings as part of normal service delivery.[46-48]

Data Sources. As of April 2018, more than 30 different EA providers, large employers or EAP industry groups had kindly shared their data to Chestnut Global Partners. Most of these EAPs were from the United States but more than 25 other countries are represented among the cases. Most of these sources are external vendors of EAP services, EAPs the serve hospital systems (and often other employers in the same local community), some internal programs from large corporations and several public sector and government organizations. Almost all of these cases were users of the counseling services from EA providers rather than users of other kinds of non-counselor services provided by the EAP (such as work/life resources or support for financial/legal issues).

Client Anonymity. Although the unique identity of each user of the EAP was tracked from pre to post use of the EAP in order to collect and match up the post use outcome data, clients were guaranteed anonymity and assured their employers would never be allowed to view their individual responses. The aggregated dataset provided for the analysis had only identification numbers and no other client specific personal information.

Sample Size. The sample size used for analysis was **24,363 cases**. This count excludes more than 1,700 other cases that did not have enough data on the WOS at both the Pre and Post time periods or were removed from the final sample for other data integrity issues.[EN3] This criteria included cases that were outliers for work absenteeism, which conceptually was defined if the person had reported more than 160 hours of missed work in the past month (which exceeds the standard full-time work schedule of 40 hours per week for four weeks). These extremes for absence hours could be due to data entry mistakes, people with an abnormally high number of days for their regular employment schedule, or maybe were on a leave from work altogether. Although rare in the total sample (at less than 0.5%), all cases with outlier status for hours of work absence were removed from the dataset in order to have consistent data on the other four WOS scales.

Measurement of Contextual Factors

In addition to the WOS, seven contextual factors of EAP use were also examined. These included the user characteristics of age, sex, and the country where the client lived, the clinical factors of referral source into the EAP and the type of presenting problem or concern, and the contextual factors created by the business sponsor of the EAP related to the industry of the employer and delivery model for the EAP service. All of these factors were taken from measures at the Pre-test period or were added to the dataset later specifically for this analysis by the research team. The specific coding of each these factors was standardized across the various specific formats of the raw data provided by the different EAPs. See **Table 1** for a summary of the counts of cases with data available for each contextual factor. A profile of each factor is also x presented.

TABLE 1: DESCRIPTION OF EAP USER SAMPLE ON VARIOUS CONTEXT FACTORS		
Factor	**Count (*n*)**	**Percentage Valid Cases with Factor**
Client Context		
Sex of EAP User		*n* = 9,219
Male	2,988	32%
Female	6,231	68%
Age of EAP User	M = 38 years	*n* = 8,810
< 30years	2,481	28%
30 - 39 years	3,094	35%
40 - 40 years	1,689	19%
50+ years	1,546	18%
Country of EAP User		*n* = 24,363
United States	19,234	79%
China	3,615	15%
Other Global	1,514	6%
Clinical Context		
Referral Source into EAP		*n* = 5,751
Self	4,950	86%
Family/Other	407	7%
Work Supervisor	274	5%
Work Mandatory	120	2%
Presenting Concern		*n* = 7,428
Mental Health and Stress	3,004	40%
Marital and Family	2,164	29%
Occupational and Work Stress	1,305	18%
Alcohol and Drug	276	4%
Other	679	9%
Employer Context		
Industry of Employer		*n* = 10,461
Healthcare	4,165	40%
Manufacturing	2,589	25%
Government	2,453	23%
Technology	1,254	12%
EAP Delivery Model		*n* = 24,363
External Vendor	15,086	62%
Employer Hybrid with External Vendor	4,760	20%
Employer with Internal EAP Staff	4,517	18%

Sex of EAP Client. For most of the total sample, sex of the client was not reported to CGP (62% missing data). Sex of the EAP client was available for 9,219 cases. Of this group, about twice as many women as men (68% vs 32%) were used of the EAP. Thus, the typical EAP user was a female.

Age of EAP Client. For most of the total sample, age was not provided to CGP (64% missing data), but the age of the EAP client was available for 8,810 cases. This was categorized into four levels: under ager 30; age 30 to 39; age 40 to 49; and age 50 or higher. The profile for age indicates 28 percent of cases were in their 20s or younger, 35 percent in the 30s, 19 percent in the 40s and 18 percent in their 50s or older. Overall, this data indicates users of EAP counseling

are found in all ages, but more so among people were under the age of 40. The typical EAP user in the study was 38 years old.

Country of EAP Client. EAP users were categorized by the country where they lived and received counseling services: *United States* (79% of all cases), followed by *China* (15% - all Chestnut Global Partners EAP China) and a mix of many *Other Global Countries* (6%). The "other global" category included EAP users from New Zealand (305), Brazil (195), Greece (111), Indonesia (17), Thailand (16), Mexico (15), France (13), Russia (13), India (12), Taiwan (11) and 14 other countries with less than ten cases each. The countries with few cases included: Argentina (4), Australia (3), Belgium (7), Chile (3), Columbia (4), Germany (4), Hungary (3), Netherlands (3), South Africa (3), Spain (6), Switzerland (5), Turkey (9), United Kingdom (4), Venezuela (1) and Vietnam (1).

Referral Source Into the EAP. For most of the total sample, type of referral into the EAP of was not reported to CGP (76% missing data), but 5,751 cases were able to be categorized based on referral source. At 86 percent of all cases, a *self-referral* was by far the most common type of referral into the EAP. In contrast, being referred to the EAP by a family member of the employee or other (7%) or by one's supervisor (5%) were much less common. The least common source of referral was a *mandatory* referral from an organization requiring employees to use the EAP as condition for further employment (2%; often by Human Resources for a safety, conduct or substance abuse violation of company policy). This data shows that most employee users of EAPs chose to seek counseling on their own and were not referred by someone else.

Presenting Concern. For most of the total sample, the reason given by the employee prompting use of the EAP was not provided to CGP (70% missing data). However, 7,428 cases had this information. From most to least commonly represented in the sample, the various kinds of presenting problems included: marital relationship (22.7%), work stress (11.4%), depression (10.9%), anxiety (10.6%), behavior and conduct issues (8.5%), occupational or work problems (6.2%), family issues (5.7%), personal stress (5.6%), grief and bereavement (4.7%), substance alcohol abuse (3.6%), violence or abuse (2.7%), medical health problems (2.2%), personal/family financial issues (1.6%), personal/family legal issues (1.3%) and eldercare (< 1%), child care (<1%), alcohol/drug codependency (< 1%), and "Other" at 1.5%. To simplify this data for analysis as a potential moderator factor of outcomes, these 19 categories were recoded into the following five more general types of problems: *Mental Health* = 40 percent (including anxiety, depression, grief, behavioral/conduct problems, personal stress); *Marital & Family Relationships* = 29 percent; *Occupational & Work Stress* = 18 percent; *Alcohol & Drug Use* = 4 percent; *Other* 9 percent.

Industry of the Company Sponsoring the EAP. Nearly half (56%) of the total cases were from EAP vendors having too many different industries among their many customers to classify one dominant industry. However, the remaining 10,461 cases EAPs were categorized into four industries: *Healthcare* (40%), *Manufacturing* (25%), *Government* (23%), and *Technology* (12%). These industries included cases from all three kinds of EAP delivery models.

Delivery Model for EAP. Based on our knowledge of the EAP providers in the study, the delivery model for the EA services was categorized for all cases.[49] The most common type was the *External Model* for a vendor with multiple employer customers (62% of all cases from 15 different EAP vendors). An *Internal Program* with dedicated EAP staff working as the employees of one organization was represented by 10 different employers and accounted for 14 percent of the total cases. One EAP served as a special hybrid type of internal program for a large academic medical center that also provided EAP services as a vendor to other smaller size employers locally (this program accounted for 5% of cases). However, this last EAP was included in the Internal Model type, which brings this type of delivery model to represent 19 percent of the total cases. Also featured was the *Employer Hybrid Model*, with a single large employer that has an external EAP vendor(s) but also has some internal full-time EAP staff who direct the activities of the vendor, provide counseling to employees, consultations to managers, and a wide range of other organization

level support services (19% of cases from 9 different employers).

Measurement of the WOS

WOS Data Collection. Most EAPs conducted the pre-test measure of the WOS telephonically during the client intake process, although other EAPs had the client complete intake paperwork themselves in a waiting area before meeting with an EA professional. The post-test WOS measure typically was collected at roughly 90 days after the pre-test, either by phone or e-mail or a weblink to online survey data collection tool. Most EAP providers adopted a protocol of using up to three follow-up attempts to collect the post-test data, either by e-mail or phone before considering the client as non-responsive. The response rate among the many different EAP providers who contributed raw data to this report was unknown – although it was estimated that approximately 30 percent of clients contacted at follow-up completed the post-test WOS measure.

WOS Measures. All three versions of the WOS are represented in the EAP users included in this sample. The choice of which version of the WOS was used to collect the data was made independently by each EAP. The original 25-item WOS was used by five EAPs and had 629 valid cases (3% of the total cases). The 9-item version of the WOS was used by two EAPs and had 5,847 valid cases (24%). The brief 5-item version of the WOS was used by 30 EAPs and had 17,887 valid cases (73%).

Preparation of WOS Data. A small number of cases (< 1%) had missing data for one or more of the five WOS outcomes at the Pre-test and/or Post-test periods. These few cases had some of their WOS scores estimated in order to preserve a full dataset and not conduct tests with minor variations in the sample sizes depending on very small amounts of missing data on the focal measures. The replacement scores for missing WOS data were estimated in two ways: 1) for subscales of the full 25-item WOS; subscale items were estimated based on matching the individual set of five ratings to scores corresponding to the actual total score for the 5-item scale available from that same specific case (i.e., this option was available because some EAPs shared data with CGP on the total scale score but did not share the individual item scores that added up to the total score); or 2) for single items of the WOS-5 brief scale, scores were estimated based on average rating for the full sample for that same item.

Work Absenteeism Measurement Issues. Because most of the EAP industry uses the brief scale version, this article focuses on data from those single items. Other results involving all of the data from the original 25-item scale are presented in the WOS 2018 Annual Report.[38]

The same single items from the lager five-item measures matching the brief item version were used from the 3 percent sample with data from the original 25-item version. Thus, all cases had the single item WOS measure even if they had originally completed the longer versions of the WOS. This process could not be done, however, for absenteeism, as the full scale and the brief scale have different instructions and questions. Therefore, a new strategy was devised to use only the data from the first three items of the full five-item version of absenteeism. This was done because these three items conceptually match the instructions for the single item on the brief WOS-5 for absenteeism that asks the person to consider absence consisting of missing work altogether, arriving late or taking off early. In contrast, the other two items on the original absenteeism scale of types of absence when being taken away from the workplace and being on phone, email or Internet while at work were excluded, as these are more aligned with the concept of work presenteeism than of missing work.

Exploratory analyses conducted on the approximately one-fourth of the sample with data on the original five item absenteeism measure (n = 6,295 EAP users) revealed an interesting pattern of results for the amount of absence at baseline before starting EAP counseling. The first item on hours of absence from the EAP concern causing the employee to miss work altogether had the highest amount of absence of the five items on the original scale at 7.39 hours, which accounted for 69 percent of the total hours of absence on the scale. The next item: arriving late for work had an average of 0.50 hours. The

third item: taking off early from work had an average of 0.85 hours. The fourth item: being pulled away from normal work location had an average of 0.72 hours. And the last item: being on the phone, e-mail or Internet while at work because of the EAP concern had an average of 1.19 hours. On every one of these five items, the vast majority of cases reported zero hours absent (i.e., 73%, 90%, 85%, 87%, and 80%, respectively). Thus, absenteeism was affecting only a small subgroup of these distressed employees.

When only the first three items were summed together, the average number of hours of absence in the past month at baseline changed from 10.77 hours (*SD* = 22.73) with all five items to 8.74 hours (*SD* = 20.84) when based on only three of the same five items. For comparison, the single-item for absenteeism on the WOS-5 brief scale at baseline had an average number of hours of absence in the past month of 6.79 hours (*SD* = 16.70), based on a different set of 17,579 EAP cases.

In summary, in the month before use of the EAP, the mean number of work absence hours from the adapted full measure (based on the first three items) was about halfway between the other two previously developed measures. Although the person was asked about the same three measures of absence in both instances – missing work, being late to work and leaving early from work – the resulting average number of hours was slightly higher when asking three separate questions than when asking a single question. This finding suggests the WOS-5 brief version slightly underestimates the amount of work absence experienced by the average EAP case by about 2 hours per month at the start of counseling. The additional cognitive effort required to think about three separate questions allows the opportunity for better recall of what happened during the past month and that is why the total absence amount is a bit higher than when the single question is used to collect absence data.

To create a single score of absenteeism hours for the entire sample including data from all three versions of the WOS, the subsample with data from the 25-item or 9-item versions of the WOS was re-coded to use only the responses to the first three items on the full absenteeism scale corresponding to the kinds of absence specified on the instructions for the single item on WOS-5 brief scale. This was done for absenteeism measures at both Pre and Post time periods. The rest of the sample retained their scores on the WOS-5 brief item for absenteeism. The value of having only one absenteeism variable for every case in the sample allowed for the opportunity to include absenteeism outcome in multivariate tests involving the full set of all five WOS constructs with all of the study participants in one sample rather than the more complicated process of conducting multiple tests within the different subsamples according to the two absenteeism measures.

To explore the utility of the new total sample absenteeism measure, results of tests of change over time from before to after use of the EAP with each these different measures of work absenteeism are shown in **Table 2**. All of the measures showed approximately the same relative reduction in the hours of absenteeism of 46 percent to 49 percent. Yet, the net amounts of hours reduced after EAP use were higher for the absenteeism measures involving more items, with 5.2 fewer hours for the full scale, 4.1 hours for the new 3-item adapted version and 3.3 hours for the single item on the WOS-5 version. When the new 3-item and single-item measures were considered together in the total sample (with each measure representing a different part of the total sample), the result was a 48 percent relative reduction in absenteeism and a net change of 3.5 fewer hours of absence in the past month. All of these change over time tests were highly significant and yet represented small size effects.

Recoding of Absenteeism Into New Version of Measure with 1-5 Range. On the WOS, work absenteeism is measured in hours (range from 0 to 160) and usually has a highly skewed distribution of scores as most of the case report either zero absence (58% of cases at Pre EAP) or a very small number of hours. This wide range and skewed distribution of scores is very different from the other four WOS dimensions, which are all measured with agree-disagree ratings on a much smaller response option range of only 1-5. These results for the other WOS measures routinely show a more normal bell-shaped distribution of scores across

Table 2: Reduction in Hours of Work Absenteeism from Pre to Post Use of EAP

WOS Measure of hours of work absenteeism in past 30 days	Sample Size N	Pre M (SD)	Post M (SD)	Percentage Change	Test of Change *F*	Effect Size η_p^2
Sub-Samples						
Work Absenteeism Original 5-item full scale	6,576	10.72 (22.83) 0 = 51%	5.52 (21.09) 0 = 77%	5.20 49% less	235***	.04 small
Work Absenteeism Only use first 3 items of the 5-item full scale	6,356	8.93 (20.97) 0 = 61%	4.82 (20.49) 0 = 83%	4.11 46% less	188***	.03 small
Work Absenteeism Single-item brief measure	17,877	6.80 (16.64) 0 = 59%	3.46 (14.54) 0 = 78%	3.34 49% less	520***	.03 small
Full Sample — Actual Data						
Mix of single item brief and first 3 items of original full scale	24,363	7.36 (17.90) 0 = 58%	3.86 (16.37) 0 = 78%	3.50[a] 48% less	696***	.03 small
Full Sample — Estimated Data						
Work Absenteeism Estimated hours from each 1 - 5 category default number of average hours	24,363	7.36 (15.60) 0 = 58%	3.25 (10.60) 0 = 78%	4.11[a] 56% less	1,684***	.07 medium

*** = $p < .001$

[a] Test of mean difference of mean scores for actual and estimated hours at Post EAP use in full sample was significant ($p < .001$) but the effect size ($\eta_p^2 = .004$) indicated the difference is very small and inconsequential.

the five rating options. From a statistical perspective, hours of absenteeism and ratings of agreement on the other four measures is like comparing apples and oranges. Thus, to more fairly conduct statistical tests using all of the WOS measures, it was important to standardize the range of the rating scales across the five measures. In order to match the 1-5 Likert-type rating scale used for the other four WOS measures and more fairly compare the five outcomes to each other, the absenteeism measure was adapted from the specific hours of work missed (range of 0-160) to a metric with only 5 categories (each with a different range of hours of absence).

This was accomplished in three steps:

Step 1: The distribution of every level of absenteeism hours at the Pre EAP use period (based on the full sample measure described in the above section that used either the WOS-5 single item score or the score from three-item adapted version of the original full scale) were tabulated and sorted from zero to the maximum.

Step 2: The distribution of absence hours then was examined carefully to set the cutoff points needed to break the distribution into five segments to correspond to a 1-5 score range. The first segment was no absence (zero hours) and was the majority of cases in both subsamples. The rest of the distribution that had at least some amount of absence was divided into fourths to evenly balance the remaining cases in the sample into four segments. The specific cut-off points for hours of missed work defining each category are shown in **Table 3**.

Step 3: Each case in the full sample was assigned a new score of 1 to 5 for absenteeism at Pre use of the EAP. The same cutoff levels were used to assign a new score of 1 to 5 for absenteeism at Post use of the EAP.

To support potential future use of this new version on how to measure absenteeism, Table 3 includes modified instructions (adapted from the WOS-5 absenteeism single item) and the new definitions of the range of work absence hours for each rating option. Included also are the default number of hours of absence to assign to cases for each of the new 1-5 ratings (reference bottom of

Table 3: New Version of WOS Work Absenteeism Measure with 1-5 Rating Scale

INSTRUCTIONS:
For the period of the past 30 days, please select the choice below that best represents the number of hours your personal concern caused you to miss work. Include complete eight-hour days and partial days when you came in late or left early.

1	2	3	4	5
No Absence (0 hours)	Less than half a day (< 4 hours)	Less than a full day (< 8 hours)	From one to three days (8 to 24 hours)	More than three days (25 to 160 hours)
Default hours for each rating to use in scoring of change over time in average hours of absenteeism per case for ROI Analyses:				
0 hours	1.58 hours	6.32 hours	15.08 hours	55.07 hours

Note: This is instead of response format of "fill-in-the blank" with estimated total hours used on earlier versions of the WOS.

Table 2). The specific default amounts of absence were derived from the calculating the average amount of reported absence hours in the total sample separately within each of the five segments (i.e., the mean for the subsample representing each absenteeism level rating of 2, 3, 4 and 5). Additionally, the full 24,343 person sample was used to test the validity of estimating the sample average using the default amounts. This process resulted in an estimated mean score at the Post use of EAP of 3.25 which is close to the actual mean score of 3.50 at the Post based on the source raw data of reported specific amounts of absence ranging from 0 to 160 hours (Reference Full Sample Data – Table 2). Compared directly, these two scores at the Post are significantly different, but the effect size of this difference is very small and not of consequence ($\eta p^2 = .004$). Note that because the specific default amounts of absence were derived from the raw absence hours at Pre EAP use within each of the five segments, the mean hours of absence at Pre EAP use were identical for both the raw data and for the new estimated data (both at 7.36 hours).

All of this effort was successful to develop a revised measure of work absenteeism that met the project criteria of:

1) a score with the same 1 to 5 range of the other WOS measures;
2) a score that was available for all of the cases in the full sample (which included source data pooled into one master dataset from the 25-item, 9-item and 5-item versions of the WOS).

Data Analyses

All analysis was conducted using IBM SPSS version 24. The test of improvement over time (Pre to Post) was conducted using a multivariate analysis of variance repeated measures procedure with all five WOS measures included in the same test. A descriptive measure of the percent improvement on each outcome over time was calculated by subtracting the Post EAP mean score from the Pre EAP mean score and then dividing it by the Pre EAP mean score. Other tests of the impact of moderator factors used a general linear model ANOVA approach with repeated measures of time and the other potential moderator factor of interest as an interaction effect with time.

With such an extremely large sample size, the power to detect a particular finding as being statistically significant is very high in this study (power of .99 out of 1.00 maximum to detect a small size effect at $p = .05$ chance level).[50] Thus, a finding too small to have much practical value can nonetheless be declared "significant" from a statistical perspective (i.e., if the test result is $p < .05$). Estimates of statistical effect size offer a way to fairer way to compare the results of the five WOS scales. The size of the *partial eta squared* effect (ηp^2) obtained in SPSS from the GLM repeated measures test results also was examined. The ηp^2 estimate can range

from 0 to more than 1.00 but it is usually a number closer to the zero end of the scale. These effect sizes can be interpreted as follows: small size effect is between .01 to .05; a medium size effect is between .06 to .13; and a large size effect is .14 or greater.[51,52] Effect sizes of less than .01 are considered very small and thus of little practical meaning.

RESULTS

Part 1 – WOS Measurment Psychometrics

The first research question asked: What is the psychometric validity and reliability of the WOS measures? This was answered by conducting a series of correlational tests for WOS measures within only the Pre period, within only the Post period and also over time combining the Pre and the Post period data in the same test.

WOS Validity. The relationships between the five scale dimensions were re-examined in this very large dataset to confirm the pattern of moderately strong associations between the five WOS scales and to rule out redundancy with each other. The findings (see **Table 4**) show moderate size intercorrelations between all five of the WOS measures (all $p < .001$). For the Pre EAP use period, the intercorrelations ranged from a low of $r = -.11$ to a high of $r = -.50$. Similarly, the correlations between the five WOS measures in Post EAP use period ranged from a low of r = -.16 to a high of $r = -.47$.

These findings confirm the shared meaning or overlap of different aspects of the work experience for EAP counseling cases. It also shows that the more general outcome construct life satisfaction is linked somewhat to the four kinds of work outcomes. This pattern is evidence of the convergent validity of these constructs as measured by the questions on the WOS. Also important for establishing the discriminant from of measurement validity is the finding that the shared variance among the WOS measures was not too high (the highest correlation of $r = .50$ when squared reflects 25% shared variance). These findings indicate that although the WOS measures do have some overlap, each measure has its own meaning that is distinct from the others and thus tells a different part of the larger workplace outcomes story.

Other tests (not shown) also revealed only very small size correlations between the client demographic factors of age and sex with the five WOS measures at Pre EAP use ($r = .11$ or less). These findings also offer evidence of the discriminant validity of the WOS, as there was no expectation that men and women or clients of different ages should differ at baseline on levels of workplace outcomes.

WOS Reliability. Measurement reliability is demonstrated when the same measure is positively correlated with itself over time. Having a measure with high temporal reliability indicates consistency or stability over time in the level of responses such that each person in the sample is roughly in the

Table 4: Correlations Between WOS-5 Brief Scale Outcomes at Pre and at Post and Paired Correlation Over Time for Same Measures

		WA	WP	WD	WE	LS
WOS Measure		**Post Use of EAP at 3 months Follow-up**				
Work Absenteeism (WA)	Pre Use of EAP at start of Case	.34	.30	.23	-.16	-.21
Work Presenteeism (WP)		.25	.36	.36	-.23	-.37
Workplace Distress (WD)		.19	.27	.49	-.47	-.31
Work Engagement (WE)		-.11	-.19	-.50	.44	.25
Life Satisfaction (LS)		-.16	-.30	-.22	.21	.36

Note: Total N = 24,363. All 1-5 range of scores. Lower scores indicate better outcomes for work absenteeism, work presenteeism and workplace distress; higher scores indicate better outcome for work engagement and life satisfaction. Correlations below diagonal are from Pre EAP Use; Correlations above diagonal are from Post EAP Use; Correlations on the diagonal are for paired Pre with Post for the same measure. All of the correlations in table are significant at $p < .001$.

same place in the overall distribution scores for the entire sample when ranked from low to high at both of the two time points. In the present study, all five of the WOS measures had significant but moderate size correlations over time from Pre to Post use of the EAP (see diagonal of matrix in Table 1 with r_{paired} = .34 to .49; all $p < .001$).

Part 2 – Improvement Over Time in WOS Outcomes

The second research question asked: What is the extent of improvement in WOS outcomes after EAP use? And which WOS outcome constructs have the largest improvement? These are answered using a multivariate general linear model (GLM) procedure with Time as repeated measures approach with two time points (Pre vs. Post use of EAP counseling) and the 1-5 rating formats for the single items that comprise the WOS-5 (but taken from all three versions of the WOS) and the new adapted 1-5 category rating version of work absenteeism for the full sample that was created for this analysis. The results are presented in **Table 5** and represent the primary findings of the study.

The results found the overall multi-variate effect for Time was highly significant and was a large size effect (η_p^2 = .34). This indicated that as a set, the five WOS measures had a substantial degree of change over time. Although each of the WOS measures individually had an improvement over time that was significant beyond chance levels, the measures differed substantially from each other in the magnitude of the relative size of the change from Pre to Post EAP counseling: work absenteeism (27%) and work presenteeism (26%) had the largest degree of improvement, followed closely by life satisfaction (22%) with workplace distress (14%) and work engagement (8%) both much less relative change. However, the differences between the WOS measures in their statistical effect sizes is the more important finding to discuss. Both work presenteeism and life satisfaction had large effect sizes (η_p^2 = .24 and .19), followed by a medium size effect for work absenteeism (η_p^2 = .13), and then small size effects for both workplace distress and work engagement (η_p^2 = .05 and .04, respectively).

Table 5: Tests of Improvement from Pre to Post Use of EAP in WOS Outcomes

	Pre M (SD)	Post M (SD)	Percentage Change	Effect for Change Over Time	
WOS Measure				Repeated Measures ANOVA F-test	Effect Size η_p^2
All Five Measures as Multivariate Effect	N/A	N/A	N/A	2,543***	.34 large
Work Absenteeism 1 - 5 categories	2.04 (1.40)	1.49 (1.06)	27%	3,543***	.13 medium
Work Presenteeism	3.30 (1.38)	2.43 (1.34)	26%	7,690	.24 large
Workplace Distress	2.25 (1.35)	1.94 (1.18)	14%	1,369***	.05 small
Work Engagement	3.21 (1.32)	3.46 (1.21)	8%	875***	.04 small
Life Satisfaction	3.00 (1.25)	3.66 (1.12)	22%	5,858***	.19 large

Note: Total N = 24,363. All 1-5 range of scores. Lower scores indicate better outcomes for work absenteeism, work presenteeism and workplace distress; higher scores indicate better outcome for work engagement and life satisfaction.

Part 3 – Contextual Factors as Potential Moderators of Improvement Over Time in WOS Outcomes

Is the level of improvement in WOS outcomes different for certain factors of the client (sex, age, global location), the clinical experience (referral and clinical problem types), and the employer context (industry and EAP delivery model)? These factors lack a compelling argument for why they would potentially influence the outcomes of EAP counseling. But as the data was available in such a large sample, it was prudent to conduct some exploratory analyses even though it was not expected to find many differences.

Even though the different sub-groups of each of contextual factor have some slight differences in where they start out on the Pre EAP WOS measures, it is the degree of change from Pre to the Post use of the EAP that of interest. This procedure determines if the degree of change (the relative percentage of improvement from Pre to Post) between the groups is of similar or different magnitude. For instance, do males have a larger reduction in absence hours after use of counseling than females? This question is tested as an interaction effect in the ANOVA statistical models. More specifically, the idea is to test if the *interaction* of Time X Age (or other context factors) was significant beyond chance and if so, how big was the statistical effect size for the interaction term in the repeated measures ANOVA model? Separate tests were conducted using the set of all five WOS-5 outcomes in a repeated measures multivariate analysis of variance (MANOVA) model with a between-subjects factor of the contextual variable and Time as Pre vs Post use of the EAP.

As shown in **Table 6**, the results found that only 6 of 35 tests even had a "small" size effect for the interaction of time and the context factor. Moreover, these six findings of interest were each barely above the partial eta squared effect size cutoff considered the minimum for a "small" effect of $\eta^2 = .01$ (the six results ranged from .010 to .025). To be more specific, these findings were for the presenteeism outcome and context factors of age, country, industry and EAP delivery model and also for the absenteeism outcome and context factors of industry and EAP delivery model. Refer to **Table 7** for the mean scores at Pre and Post and the relative change percentages for these factors and WOS measures.

However, the primary result was the other 29 tests had no effects for the context factors ($\eta^2 < .01$). The WOS outcomes of workplace distress, work engagement and life satisfaction each had similar levels of improvement for the various subgroups within all seven of the context factors examined. A similar degree of improvement on all five WOS outcomes also was found for male and female clients of EAP counseling. A similar degree of improvement on all five WOS outcomes was found for

Table 6: Results of Tests of Context Factors on Improvement Over Time in WOS Outcome Measures

	Workplace Outcome Suite Measure				
Context Factor:	**Work Absenteeism**	**Work Presenteeism**	**Workplace Distress**	**Work Engagement**	**Life Satisfaction**
Sex of Client	No effect	No effect	No effect	No effect	No effect
Age of Client	No effect	η_p^2 = **.012** *small effect*	No effect	No effect	No effect
Country of Client	No effect	η_p^2 = **.010** *small effect*	No effect	No effect	No effect
Referral Type	No effect	No effect	No effect	No effect	No effect
Problem Type	No effect	No effect	No effect	No effect	No effect
Industry of Employer	η_p^2 = **.025** small effect	η_p^2 = **.011** small effect	No effect	No effect	No effect
EAP Delivery Model	η_p^2 = **.016** small effect	η_p^2 = **.012** small effect	No effect	No effect	No effect

Table 7: Details of Key Results for Context Factors on Improvement Over Time in WOS Presenteeism and Absenteeism

Context Factor	N cases	Pre	Post	Percentage Change
Presenteeism Single Item 1 - 5 Rating Scale				
Age of EAP User				
< 30years	2,481	3.20	2.15	**33%**
30 - 39 years	3,094	3.23	2.22	**31%**
40 - 49 years	1,689	3.18	2.58	19%
50+ years	1,546	3.31	2.57	22%
Country of EAP User				
United States	19,234	3.34	2.54	24%
China	3,615	3.09	1.86	**40%**
Other Global	1,514	3.24	2.47	**24%**
Industry of Employer				
Healthcare	4,165	2.99	2.48	**17%**
Manufacturing	2,589	3.22	2.35	27%
Government	2,453	3.44	2.58	25%
Technology	1,254	3.44	2.65	23%
EAP Delivery Model				
External Vendor	15,086	*3.37*	2.40	29%
Employer Hybrid with External Vendor	4,760	*3.30*	2.46	25%
Employer with Internal EAP Staff	4,517	*3.04*	2.51	**17%**
Absenteeism Single Item 1 - 5 Rating Scale				
Industry of Employer				
Healthcare	4,165	2.03	1.59	**22%**
Manufacturing	2,589	2.54	1.76	31%
Government	2,453	2.49	1.48	41%
Technology	1,254	2.16	1.80	**17%**
EAP Delivery Model				
External Vendor	15,086	1.91	1.43	25%
Employer Hybrid with External Vendor	4,760	2.49	1.58	**37%**
Employer with Internal EAP Staff	4,517	2.02	1.61	20%

cases with different kinds of referrals into EAP counseling and for different kinds of presenting problems. In summary, 83 percent of the possible tests conducted indicated no meaningful differences for these context factors and the other positive tests all had very small effect sizes.

When these findings are considered together, the overall conclusion is that age and sex of client, source of referral into the EAP, type of presenting concern, employer industry, EAP delivery model and country of service have almost no practical impact on the level of the effectiveness of EAP counseling as represented in the rates of improvement on the five WOS scales. These contextual factors *do not* appreciably affect the degree to which clients of EAP counseling improve on these outcomes.

Part 4 – Relevance of the WOS Constructs to Employee Who Use EAPs

Finding a big change over time for some WOS measures (work presenteeism and life satisfaction) and yet smaller changes for other WOS measures (workplace distress and work engagement), raises a question of are certain WOS outcomes more relevant than the others to EAP counseling? One possible answer concerns how the starting level of

some of the WOS outcomes simply is not at a high enough level of problem severity (i.e. the negatively valued range of the 1- 5 rating scale, depending on the item for most of the cases in the sample to allow for improvement to occur after treatment. This is referred to as a restricted range problem when the starting score on a particular measure is close to the target level of the rating scale before treatment and thus provides little room for improvement after treatment.

To operationalize this approach involved asking how many cases were at a "problem" level before EAP use and thus had a reasonable opportunity to get better after counseling? This question was answered by using meaning embedded in the labels on the response scales to determine a more clinically relevant sub-portion of the EA user population who score at a "problem level" on a particular WOS outcome. This approach borrows from the wellness field's emphasis on finding employees who are at-risk for a health issue and then trying to reduce those risks through education and lifestyle coaching. This metric is simply the percentage of total cases that are at a "problem level" on each of the WOS measures.

The WOS data was re-coded for problem level status in the following manner. The two WOS scales that are phrased as unhealthy constructs (presenteeism and workplace distress) were considered to be at a "problem level" when a person either agreed or strongly agreed with the item (i.e., ratings of 4 or 5). Conversely, the other two WOS scales phrased as healthy constructs (work engagement and life satisfaction) were considered to be at a "problem level" when cases disagreed with the item (i.e., selected either of the options of disagree or strongly *dis*agree for the ratings of 1 or 2).

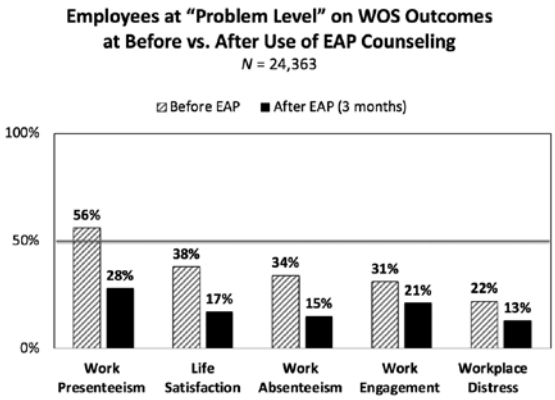
Finally for absenteeism this re-coding process had to be done differently. As the typical employee misses less than half a day of work each month due to health reasons (see review of four national survey studies by Attridge in 2016),[16] a criterion of four hours absence per month was established and an EAP user with 4 or more hours of absence was considered a "problem level" of absenteeism. The WOS scores were re-coded in this manner for all cases with available data.

The problem status variable for absenteeism was based on hours from the modified three-item measure full WOS 5-item version or from the single-item version, depending on which version was used for data collection. The problem status variable for the other four WOS outcomes were based on only the single item featured in the WOS-5 brief version, but this item was taken from all available responses pooled across the full WOS-25, WOS-9 and WOS-5 versions.

The results based on the full sample at baseline reveal that work presenteeism was the most common problem for users of EAP counseling, with 56 percent of cases agreeing that work problems prevented them from concentrating while at work. Next was having a problem with life satisfaction, with 38 percent of all EAP cases feeling their lives was *not* going well. About 1 in 3 EAP cases (34%) had a level of absenteeism from work greater than the typical employee (4 or more hours per month). About 1 in 3 EAP cases were *not* eager to get to the worksite and start their work day (31% had a problem with work engagement). And finally, about 1 in 4 EAP cases started counseling, feeling they dreaded going into work (22% had a problem with workplace distress).

Examining the variables of WOS outcome problem prevalence for the Post EAP use period, there was substantial reduction from Pre to Post use of the EAP for all five of the WOS measures (see **Figure 1**). The number of employees with these kinds of problems in the full sample was cut in half or reduced by one-third from Pre to Post (range from 56% to 32% relative reduction over time depending on the outcome).

Figure 1:



DISCUSSION

This study offers large sample evidence of the psychometric validity and test-retest reliability for all five WOS measures found in correlational tests. It also was discovered that some constructs of the WOS had better results than others concerning the effectiveness of EAP counseling. Of the five WOS measures, work presenteeism was the outcome that improved the most after EAP counseling. About 1 in every 2 EAP users had a work presenteeism problem at the start of counseling and this prevalence rate was cut in half when assessed three months later after counseling. The improvements over time for work presenteeism is the headline in the story of where EAPs make the most difference.

Perhaps most stunning was discovering that absence from work actually is not a significant aspect of the work performance burden among the majority of distressed employees who use an EAP. On average, only 1 in 3 EAP cases had a level of work absence in the past 30 days before they started counseling that exceeded the amount of health-related absence of the typical worker in the U.S. Even though the change in absence hours was good on a relative percentage basis (a 47% reduction), the specific number of hours involved is rather small, with a change from 7.4 hours on average per case at Pre EAP to 3.9 hours at Post EAP. In the big picture, 3.5 hours of restored lost work time is only a small fraction of the total 160 hours in the standard full-time monthly work schedule. Again, this fact points to the importance of the much larger change in work presenteeism among EAP cases as the time lost from being unproductive while at work usually involves far more total hours over the course of a month than does a half day of absence from work. For example, a recent review that data from multiple self-report measures of presenteeism and productivity from EAP cases worldwide estimated the typical case has a loss of 53 hours of unproductive time while at work during the month before seeking assistance from the EAP.[16] This is about six and a half total days of work loss from on-the-job presenteeism beyond the one day for absence found in the present study of WOS absenteeism data.

The large size statistical effect for the outcome of life satisfaction was somewhat of a surprise. This result can be interpreted as the life satisfaction item functioning as a more general indicator of the clinical relief and restored personal well-being that was experienced after getting personalized support and referral to needed additional resources from the EAP counselor.

Although EAP use does move work engagement and workplace distress in positive directions, these outcomes both had smaller effect sizes than the other three WOS measures. It may be that EAP interventions delivered at the individual level by counselors in private sessions cannot directly impact the larger workplace and managerial conditions operating at the organizational level that strongly influence work engagement and workplace distress.[53] More meaningful improvements in work engagement and work distress likely require other kinds of EAP services such as workplace training, managerial coaching, and work culture interventions. Even so, EAP counseling still has a measurable – if smaller – impact on work engagement and workplace distress.

Concerning the results of the exploratory tests of moderating factors, it was found that client demographic factors, clinical factors and employer contextual factors had either no effect or very small differences on degree of change from Pre to Post EAP use on the WOS outcomes.

The new analytical approach of splitting the cases into two groups of those with or without a "problem" with each kind of WOS outcomes revealed new insights about what appears to be the rather healthy status of the typical user of EAP counseling. Other than a slight majority of the cases with a work presenteeism problem, the majority of EAP cases were *not* at a problem level on all four of the other WOS scales when starting counseling. The rather low prevalence rates of having problems on multiple kinds of work outcomes can also be interpreted as an indication of the mostly non-clinical nature of employees who seek help from the EAP. It is self-evident that even though something happens in the personal or work life that prompts the need for seeking immediate assistance from employer-sponsored employee assistance programs, all of the individuals using EAP services are healthy enough to be gainfully employed (rather

than unemployed or on disability).

What is perhaps most interesting is the wide range between the different WOS measures in the percentage of cases at baseline with a "problem" on the measure. The biggest difference was that more than twice as many users of EAP counseling had a work presenteeism problem than had a workplace distress problem at the start of counseling. Such differences in the levels of baseline problem prevalence may explain the differences that were found between the WOS measures in their statistical effect sizes as described in Part 2 of the Results. Finding small size effects for workplace distress and work engagement makes more sense when considering these outcomes both had the smallest percentage of cases at a problem level when beginning counseling.

Putting the WOS Findings In Context

A global data review of EAP outcomes using a variety of measures and methods and based on 122,755 cases from 9 studies (excluding the other past studies in the review using the WOS measures) found that 27 percent of EAP cases, on average, have a problem with work absence in the month prior to counseling.[16] Thus, the problem status level of 34 percent in the present study of WOS aggegated data is on par with what has often been found in the past studies of absence among EAP users.

The same review of other global EAP outcomes data for work presenteeism examined eight studies (excluding the other past studies in the review using the WOS measures) with a variety of measures and methods and represented a combined total of 121,273 cases.[16] It found that 55 percent of EAP cases, on average, had a problem with being productive during the month prior to counseling. The finding that 56 percent of EAP cases had a problem with work presenteeism in the present study of WOS data is very consistent with what is typically found in past studies of presenteeism and lack of productivity among EAP cases.

The findings with the WOS emphasize that work presenteeism is much more of a significant issue than absenteeism for employees. Indeed, of the five WOS outcomes, work presenteeism is the most common problem at the start of counseling and also has the largest effect size for the magnitude of improvement after counseling from the EAP. This emphasis on presenteeism over absenteeism also has been found in many studies of worker health.[54-56]

Limitations

There are numerous variables missing from this study that are potentially responsible for differences in WOS results. Some of these factors are in the overall health or well-being status of the client (clinical risk factors), the counselor rated level of clinical severity of the case (seriousness of the risks), the number of counseling sessions experienced (clinical dosage delivered), the fidelity of the counseling interventions provided to meeting best practices for EAP (quality), whether or not the case was referred out after the EAP for more serious treatment (clinical referral), if the sessions were provided in-person or telephone or via e-health technology tools (clinical modality). All of these important factors are not in the present global dataset of WOS data. Thus, further research is needed to tease apart which of these factors are most strongly impacting how much clients improve (or do not improve) resulting from use of EAP counseling. Even so, the present study clearly indicates the "typical" kinds of EAP counseling delivered all over the world has a substantial impact on work presenteeism and overall life satisfaction of employees and has smaller effect on reducing problems of work absence, distress over conditions at the workplace and lack of engagement in one's work.

Conclusion

The workplace outcomes approach represents a departure from conventional measures by objectively identifying when employee assistance services demonstrably work in the context of the workplace. No self-report survey instrument is perfect, but the WOS is the best tool to date that EA professionals have for refuting the age-old question skeptical employers have of whether EAPs actually can contribute to improving the well-being and work performance of distressed employees.

The Workplace Outcome Suite also was developed to make the case for whether investing in EAP makes business sense.[57-59] There's a broad misconception that all EA

services and service providers are equal except in price.[60] Since the typical cost for EAP is remarkably low even when selecting the highest cost vendor, organizations would benefit from focusing on identifying an EAP partner that is a good fit for the organization, and not just one that is offered at the lowest price. These issues are reviewed in several other papers.[61-63] Indeed, employers and brokers focusing on finding the lowest cost provider or selecting an EAP vendor only on the level of program utilization should expand their selection criteria to include a more rigorous analysis of the EAP's ability to deliver direct and indirect business value to the organization using a validated outcomes tool with a credible methodology.

The Workplace Outcome Suite can fill this need and now that worldwide normative results are available, the different EAPs collecting WOS data can be compared with each other on the same outcome metrics. Such a scenario could usher in a new era that offers EAPs the chance to compete for business on program outcomes as well as on program price.

AUTHORS

Mark Attridge, PhD, MA, is president of Attridge Consulting, Minneapolis, MN, U.S.A. and a research scholar in independent practice focusing on workplace mental health and employee assistance programs.

David Sharar, PhD, is CEO of Chestnut Health Systems and Director of Chestnut Global Partners, (a Morneau Shepell company) Division of Commercial Science, Bloomington, IL, U.S.A., is co-creator of the Workplace Outcome Suite© (WOS) and a leading provider of employee assistance services worldwide.

Gregory DeLapp, MHS, CEAP is CEO of the Employee Assistance Professionals Association (EAPA), an international organization representing more than 4,000 EAP professionals based in Arlington, VA, U.S.A.

Barbara Veder, MSW, RSW, is Vice President, Global Clinical Services, Research Lead, and Chief Clinician at Morneau Shepell.

The authors wish to thank Kathryn Modisette of Chestnut Global Partners for her assistance in preparing the master file of WOS data for the study.

ENDNOTES

EN[1] For more information on the history of the WOS, see the paper by Lennox, Sharar, Schmitz, and Goehner in this same special issue that compares the three versions and examines the psychometric properties of the 5-item short form.

EN[2] For more information on the case study of the WOS data at Federal Occupational Health, see the paper by Mintzer, Morrow, Tamburo, Sharar and Herlihy in this same special issue.

EN[3] Excluded from this total was one internal EAP with 354 cases. This was excluded because it had not collected data on two of the five WOS measures (i.e., had missing data for all cases at Pre and Post on absenteeism and engagement) and no data on the case level client demographic or clinical experience factors.

REFERENCES

1. Karg RS, Bose J, Batts KR, Forman-Hoffman VL, Liao D, Hirsch E, Pemberton MR, Colpe LJ, Hedden SL. Past year mental disorders among adults in the United States: Results from the 2008–2012 Mental Health Surveillance Study. *CBHSQ Data Review.* Published by the Center for Behavioral Health Statistics and Quality. Washington, DC: SAMHSA. 2014, October.
2. Nyce S, Grossmeier J, Anderson DR, Terry PE, Kelley B. Association between changes in health risk status and changes in future health care costs: A multiemployer study. *Journal of Occupational and Environmental Medicine.* 2012;54(11):1364-1373.
3. Bennett J, Bray J, Hughes D, Hunter J, Frey JJ, Roman P, Sharar DA. Bridging Public Health with Workplace Behavioral Health Services. Unpublished independent paper jointly sponsored by the Employee Assistance Professionals Association, The Employee Assistance Research Foundation, and the Employee Assistance Society of North America. 2015. Available from: http://hdl.handle.net/10713/4876
4. Attridge M. Employee Assistance Programs: Evidence and current trends. In: Gatchel RJ, Schultz IA, eds. *Handbook of Occupational Health and Wellness.* Boston, MA: Springer. 2012: 441-467.
5. Employee Assistance Professionals Association. EAPA Standards and Professional Guidelines for Employee Assistance Programs. Arlington, VA: Author. 2010.
6. Employee Association Professionals Association. Definitions of employee assistance program and core technology. 2011. Available from: http://www.eapassn.org/i4a/pages/index.cfm?pageid=521
7. Roman PM, Blum TC. The Core Technology of Employee Assistance Programs. *The ALMACAN.* 1985;15(3):8-9,16-19.
8. Roman PM, Blum TC. Reaffirmation of the Core Technology of Employee Assistance Programs. *The ALMACAN.* 1988;19(8):17-22.
9. Sharar DA. General mental health practitioners as EAP network affiliates: Does EAP short-term counseling overlap with general practice psychotherapy? *Brief Treatment and Crisis Intervention.* 2008;8:358-369.
10. Shakespeare I, Ichikawa K, Antonissen D, Ozeisei A, Wigger H, Espinoza, J. A global look at EAP. *Journal of Employee Assistance.* 2014;44(4).
11. Attridge M. Future Trends in Organizational Risk Management: Survey Results and Implications for EAPs in Europe. [discussion leaders Antonissen D, Sommer M.). Keynote presentation at the 16th annual conference of the Employee Assistance European Forum, Milan, Italy; 2017, June. Available from: http://hdl.handle.net/10713/6721
12. Hartwell T, Steele P, French

M, Potter F, Rodman N, Zarkin G. Aiding troubled employees: The prevalence, cost, and characteristics of employee assistance programs in the United States. *American Journal of Public Health.* 1996;86(6):804-808.
13. Bureau of Labor Statistics, U.S. Department of Labor. The Economics Daily, Employer-provided quality-of-life benefits, March 2016. 2017. Available from URL: https://www.bls.gov/opub/ted/2016/employer-provided-quality-of-life-benefits-march-2016.htm
14. WorldatWork. 2017 Inventory of Total Rewards Programs & Practices. White paper. Washington, DC: Author. (2018).
15. Attridge M. The business case bibliography: 100 review papers on the workplace value of mental health, addiction and EAP services. EASNA Research Notes. 2011;2(4):1-10. Available from: http://hdl.handle.net/10713/5129
16. Attridge M. EAP Industry Outcomes for Employee Absenteeism and Presenteeism: A Global Research Analysis. Presented the annual conference of the Employee Assistance Professionals Association, Chicago, IL; 2016, October. Available from: http://hdl.handle.net/10713/7203
17. Csiernik R. The glass is filling: An examination of employee assistance program evaluations in the first decade of the new millennium. *Journal of Workplace Behavioral Health.* 2011;26(4): 334-355.
18. Blum TC, Roman PM. Cost-Effectiveness and Preventive Implications of Employee Assistance Programs. U.S. Department of Health and Human Services, Public Health Service, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Prevention. 1995.
19. McLeod J. The effectiveness of workplace counselling: A systematic review. *Counselling and Psychotherapy Research,* 2010;10(4):232-248.
20. Joseph B, Walker A, Fuller-Tyszkiewicz M. Evaluating the effectiveness of employee assistance programmes: A systematic review. *European Journal of Work and Organizational Psychology.* 2018;27(1):1-15.
21. Mellor-Clark J, Twigg E, Farrell E, Kinder A. Benchmarking key service quality indicators in UK Employee Assistance Programme Counselling: A CORE System data profile. *Counselling and Psychotherapy Research.* 2013;13(1):14-23.
22. Attridge M, Cahill C, Granberry GW, Herlihy PA. The National Behavioral Consortium industry profile of external EAP vendors. *Journal of Workplace Behavioral Health.* 2013;28(4):251-324. Available from: http://hdl.handle.net/10713/3687
23. Jacobson JM, Jones AL, Bowers N. Using existing employee assistance program case files to demonstrate outcomes. *Journal of Workplace Behavioral Health.* 2011;26(1):44-58.
24. Attridge M. Evolution of the Employee Assistance Program industry in North America: Historical changes and what purchasers want today. Unpublished report. Alexandria, VA: Employee Assistance Society of North America, 2018. Available from: http://hdl.handle.net/10713/8061
25. Lennox RD, Sharar DA, Schmitz E, Goehner DB. Development and validation of the Chestnut Global Partners Workplace Outcome Suite. *Journal of Workplace Behavioral Health.* 2010;25(2):107-131.
26. Sharar DA, Lennox R. The workplace effects of EAP Services: "Pooled" results from 20 different EAPs with before and after WOS 5-item data. EASNA Research Notes. 2014;4(1). Available from: http://www.easna.org/publications
27. Pompe JC, Lennox R. Case Example: Measuring EAP Success Using Workplace Outcomes. Presented at the 22nd Annual Institute of the Employee Assistance Society of North America, Montreal, QB, Canada. 2010, May.
28. Sharar DA, Pompe JC, Lennox R. Evaluating the workplace effects of EAP counseling. *IHPM Journal of Health & Productivity.* 2012;6(2):5-14.
29. Sharar DA, Pompe JC, Attridge M. On-site versus off-site EAPs: A comparison of workplace outcomes. *Journal of Employee Assistance.* 2013;43(2):14-28.
30. Sharar DA, Eertmans A. *Happiness, Life Satisfaction and Work Engagement: A Case Study with the WOS Demonstrating Effects of EAP (ConocoPhillips).* Presented at the annual conference of the Employee Assistance European Forum, Dublin, Ireland. 2013, June.
31. Lennox R, Mollenhauer M. The Workplace Outcome Suite: Results from an EAP Research Network. Presented at the Annual Conference of the Employee Assistance Professionals Association, San Diego, CA. 2015, October.
32. Sharar DA, Mollenhauer M, Heck P. EAP Works Across Cultures and Borders. White paper on EAPA website. 2016. Available from: http://www.eapassn.org/Study-EAP-Works
33. Mintzer J, Back Tamburo M. Demonstrating Value: Measuring Outcomes and Mitigating Risks with the Workplace Outcome Suite within the Federal government. Presented at the annual institute of the Employee Assistance Society of North America, Atlanta, GA. 2017, May.
34. Attridge M, Menco H, Stidsen A. Fewer employees have workplace problems after counseling: WOS-5 results at Partners HealthCare System EAP. *Journal of Employee Assistance.* 2018, in-press.
35. Chestnut Global Partners. Workplace Outcome Suite 2018 Annual Report: Global Results for Counseling Outcomes, EAP Profiles and Rethinking Absenteeism. Bloomington, IL: Author. 2018.
36. Servizio L. The universal language of ROI: Converting your EAP benefit into an investment. Keynote presentation at the 5th Forum on Employee Assistance Programs, Athens, Greece. 2017, May.
37. Li P, Mollenhauer M, Zhang C. Examining EAP effectiveness in China. *Journal of Employee Assistance.* 2015;45(3):24-27.
38. Li P, Sharar DA, Lennox R, Zhuang W. Evaluating EAP counseling in the Chinese workplace: A study with a brief instrument. *Journal of Workplace Behavioral Health.* 2015;30(1-2):66-78.
39. Li P. EAP Research Presentations from Around the Globe. Presented at the Annual Conference of the Employee Assistance Professionals Association, Chicago, IL. 2016, November.
40. Burgess K, Lennox R, Sharar DA, Shtoulman A. A substance abuse intervention program at a large Russian manufacturing worksite. *Journal of Workplace Behavioral Health.* 2015;30(1-2):138-153.
41. Mazouropoulou C. WOS (Workplace Outcome Suite): Longitudinal Outcomes of the Face to Face EAP Counseling Services in Greece. Presentatied at the 5th Forum on Employee Assistance Programs, Athens, Greece. 2017, May. Available from: http://hdl.handle.net/10713/7232
42. Richmond MK, Pampel FC, Wood RC, Nunes AP. The impact of employee assistance services on workplace outcomes: Results of a prospective, quasi-experimental study. *Journal of Occupational Health Psychology.* 2017;22(2):170-179.
43. Richmond MK, Pampel FC, Wood RC, Nunes AP. Impact of employee assistance services on anxiety, depression and risky alcohol use: A quasi-experimental study. *Journal of Occupational and Environmental Medicine.* 2016;58(7):641-650.
44. Nunes AP, Richmond MK, Pampel FC, Wood RC. The effect of employee assistance services on reductions in employee absenteeism. *Journal of Business and Psychology.* 2017:1-11. Available from: https://doi.org/10.1007/s10869-017-9518-5
45. Attridge M, Dickens, S. Health and Emotional Wellbeing in the Workplace: Behavioral Screening and Intervention as an Effective Approach to Wellness. Presented at the Vermont Worksite Wellness Conference, Burlington, VT. 2018, March. Available from: http://www.healthvermont.gov/wellness/physical-activity-nutrition/2018-worksite-wellness-conference
46. Attridge M. Can EAPs experiment in the real world. *EAPA Exchange.* 2001(2);26-27.
47. Lerner D, Rodday AM, Cohen JT, Rogers WH. A systematic review of the evidence concerning the economic impact of employee-focused health promotion and wellness programs. *Journal of Occupational and Environmental Medicine.* 2013;55(2):209-222.
48. Pompe J, Sharar DA. Preparing for the challenges of the Employee Assistance Research Foundation: A response to Carl Tisone's 'Call to Action'. *Journal of Workplace Behavioral Health.* 2008;23(3), 217—227.
49. Attridge M, Amaral T, Bjornson T, Goplerud E, Herlihy P, McPherson T, Paul R, Routledge S, Sharar DA, Stephenson D, Teems L. EAP services, programs and delivery channels. EASNA Research Notes. 2009;1(4). Available at: http://www.easna.org/publications
50. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, New York: Routledge. 1998.
51. Richardson JTE. Eta squared and partial eta squared as measurements of effect size in educational research. *Educational Research Review.* 2011;6:135-147.
52. Wuensch KL. Standardized Effect Size Estimation: Why and How? Unpublished Report. 2015. Available from: http://core.ecu.edu/psyc/wuenschk/StatHelp/Effect%20Size%20Estimation.pdf
53. Attridge M. Measuring and managing employee work engagement: A review of the research and business literature. *Journal of Workplace Behavioral Health.* 2009;24(4):383-398.
54. Johns G. Presenteeism in the workplace: A review and research agenda. *Journal of Organizational Behavior.* 2010;31(4):519-542.
55. Goetzel RZ, Long SR, Ronald MS, Ozminkowski J, Hawkins K, Wang P, et al. Health, absence, disability and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers. *Journal of Occupational and Environmental Medicine.* 2004;46(4):398-412.
56. Pauly MV, Nicholson S, Polsky D, Berger ML, Sharda C. Valuing reductions in on-the-job illness: 'Presenteeism' from managerial and economic perspectives. *Health Economics.* 2008;17(4):469-486.
57. Attridge M, Servizio L, Sharar DA, Mollenhauer M. EAP ROI Calculator - Conceptual approach and default data inputs: Research review appendix. White Paper. Bloomington, IL: Chestnut Global Partners. 2015.
58. Servizio LA. Why CFOs should learn to love EAPs. *WorkSpan.* 2017;April:51-57.
59. Servizio LA, Mollenhauer M, Shjerven T. Taking the mystery out of ROI calculations. *Journal of Employee Assistance.* 2016;46(4):26-27.
60. Pompe JC, Sharar DA, Jacobson-Frey J. EAP options for today's workforce: Maximizing value by design. Benefits Pro. 2017, April. Available from: https://www.benefitspro.com/2017/04/06/eap-options-for-todays-workforce-maximizing-value/?slreturn=20180308175439
61. Sharar DA, Hertenstein E. Perspectives on commodity pricing of Employee Assistance Programs: A survey of the EAP Field. *WorldatWork Journal.* 2006;Q1:34-41.
62. Attridge M. Integration Insights Column (#7): Implications of Pricing for EAP integration. *Journal of Employee Assistance.* 2017;47(1):26-27.
63. Frey JJ, Pompe J, Sharar DA, Imboden R, Bloom L. Experiences of internal and hybrid employee assistance program managers: Factors associated with successful, at-risk, and eliminated programs. *Journal of Workplace Behavioral Health.* 2018;33(1):1-23.

Demonstrating Value: Measuring Outcome & Mitigating Risk: FOH EAP Study Utilizing the Workplace Outcome Suite

Jeffrey Mintzer LICSW, CEA; Veronica Y. Morrow , MSW, LCSW-C, CEAP; Melissa Tamburo, Ph.D, LCSW-C; David Sharar Ph.D; Patricia Herlihy Ph.D., RN

ABSTRACT

KEY WORDS: Employee Assistance Program , workplace outcomes, FOH, WOS, measurement tools

Despite the popularity and prevalence of Employee Assistance Programs (EAPs), and the historical emphasis on how EAP can improve work performance, there has been very little rigorous evaluation of the workplace effects of EAP counseling. The aim of this Federal Occupational Health (FOH) outcome study was to examine if and to what degree EAP counseling is associated with improved workplace effectiveness with this particular population.

Federal Occupational Health (FOH) is the largest provider of occupational health services in the Federal Government, serving more than 360 federal agencies and reaching 1.8 million federal employees. FOH began providing Employee Assistance Program (EAP) services in 1980 and is Health and Human Services' recognized expert in this key area of employee health programs, delivering specialized EAP services exclusively to over 905,624 federal employees. In 2004 Selvick, Stephenson, Plaza and Sugden published one of the few studies that demonstrated statistically and practically significant outcomes from the FOH's EAP. Their work showed significant improvement from pre- to post-EAP intervention on measures of productivity; work and social relationships; perceived health status; attendance and tardiness; and global assessment of functioning. In an effort to revitalize the findings with more current outcomes, FOH engaged an industry gold-standard tool, the Workplace Outcome Suite (WOS). This measurement tool consists of a 5-item measure, that has been psychometrically tested and is also easy to administer telephonically. It consists of five scales that measures absenteeism, presenteeism, work engagement, life satisfaction, and workplace distress. In October of 2015 FOH began to collect data on specific outcomes for clients who accessed the EAP. This study reports findings from 2016 and 2017 data that indicates a significant decrease in absenteeism and workplace distress as well as increases in life satisfaction and workplace presenteeism.

INTRODUCTION

William Huddock of SAMSHA (2017) stated that: "articulating a clear value proposition is a necessary, overarching goal for the EAP field. Employers and government leaders increasingly are demanding evidence-based programs that produce demonstrable results. Outcome measurement is the single most powerful tool in revamping the EAP proposition Since 2009 Chestnut Global Partners has been constructing a measurement tool, the Workplace Outcome Suite (Lennox, 2009) to provide a means for all EAPs to collect outcome data from their services. Many EAPs from across the globe have begun using the WOS tool as it evolves into a reliable instrument for gaining solid outcome results.

The Federal Occupational Health Program began using the WOS tool in October 2015 and immediately began collection of outcome data. This paper presents the data collected in both 2016 and 2017 and its implications for the field, as well as some recommendations for further related research (WOS Annual 2017 Report).

BACKGROUND

The Federal Occupational Health (FOH) is a non appropriated agency within the Program Support Center (PSC) of the U.S. Department of Health and Human Services and works in partnership with other federal agencies to deliver comprehensive occupational health solutions that "improve the health, safety and productivity of the federal employee." Created in 1946 by an amendment to the Public Health Service Act (42U.S.C.), Federal Occupational Health has been providing services exclusively to federal agencies for almost seven decades and is the largest provider of occupational health services in the Federal Government, serving more than 360 federal agencies and reaching 1.8 million federal employees.

FOH began offering Employee Assistance Program (EAP) services in 1980. It delivers specialized EAP services exclusively to over 986,390 federal employees. It's unique leadership role with Health and Human Services enables Federal Occupational Health to leverage the expertise of Health and Human Service agencies such as the Substance Abuse and Mental Health Services Administration (SAMHSA) and the Centers for Disease Control and Prevention (CDC) to incorporate the latest trends and best practices within the evolving field of employee wellness. More than ten years ago, the EAP was integrated with the Work/Life Program, an integrated restructuring practice that had become popular in the field at that point in time. (Herlihy, 2011)

FOH fully understands that preserving sound psychological and physical health while also strengthening the coping abilities of personnel, is essential to maintaining a fully capable work force ready to serve our nation. As the self-supporting independent agency that employees 600,000 and is responsible for "providing a reliable, efficient, trusted and affordable universal delivery service that connects people and helps businesses grow", FOH understands the unique challenges of the diverse agencies they serve.

FOH provides access to on and/or off-site assessment, counseling and problem solving for a wide range of emotional and Work/Life issues. FOH provides EAP services in cities and towns all across the country including some of the most remote corners of the United States. In addition to staff counselors located in more than 50 counseling offices in federal buildings, FOH provides services through a vast network of "affiliate" counselors in approximately 21,000 locations across the country.

EAP Services are also available to federal employees and their families stationed in more than 100 countries overseas. Federal Occupational Health offers critical incident response, performance coaching, management consultation, educational activities, referral, monitoring, behavioral health consultation, and follow-up services 24 hours a day, 7 days a week, and 365 days a year.

STUDY DESIGN

This study deployed a group of "pre-post" or "before-after" design to examine the relationship between EAP counseling and specific workplace effects. This design is frequently used when access to a matched comparison group is not available or permitted. It was selected because it was not disruptive to the subject's normal help-seeking process and the investigators were not in a position to remove or manipulate the intervention. Its purpose in this study was to test the strength of the association between EAP counseling and work effectiveness, so it can identify if employees are improving at work. Our intent was not to authoritatively explain why, or prove that EAP services caused the improvements at work. Rather to simply observe if there were any differences in workplace functioning post counseling services provided by the FOH EAP.

METHODOLOGY

To choose an appropriate number for the sample with this particular client company, a power analysis was used. The FOH program averages approximately 8,100 requests for EAP services per month from federal government employees. With this high volume of requests, the team decided a probability sample whereby a smaller number of participants can represent the total population of EAP clients, as long as each participant had an equal probability of

being selected.

In 2015 FOH began collecting WOS data by asking the five additional tool questions after the initial intake questions related to the employee's reason for calling for EAP assistance. A random sample of the clients seeking services were asked to answer the 5-item WOS questionnaire. Approximately three months (90 days) later these same respondents were re-contacted for follow up and asked the same 5-item WOS questions. The study sample size consisted of 2,603 in 2016 and 2,197 in 2017 for a total of 4800 respondents who completed both the pre and post tests and for a response rate of 28 percent. A paired sample t-test was used to examine changes in average scale scores from before respondents used EAP to after services were rendered.

WORKPLACE OUTCOME SUITE

The 25-item WOS measure was specifically developed for use with EAPs by the Division of Commercial Science at Chestnut Global Partners in 2010 (Lennox, 2010). It has demonstrated a robust degree of statistical sensitivity to change in EAP evaluations with as few as fifty (50) respondents. Although this measurement tool is comparably short, many EAPs consider the 25-item WOS too long for regular use in routine outcome monitoring. As a response to these comments, Chestnut Global Partners developed a 5-item version of the WOS (WOS-5) (Lennox, 2018) in 2015 (Consult Appendix). This shorter version takes one question from each of the original 25-item WOS scales of Presenteeism, Work Engagement, Life Satisfaction, and Workplace Distress. Traditional scaling techniques allowed the selection of the best representation question from each latent variable using the same confirmatory factor analysis reported in the original 25-item WOS validation (Lennox, et al., 2010). For a description of the psychometrics of the WOS-5 scaling down from the original WOS-25 tool please consult Lennox (2018).

FOH'S RATIONALE FOR USE OF WOS TOOL:

1. Workplace focused (not a clinical measure)
2. Validated with demonstrated psychometrics
3. Short but precise (sensitive to change)
4. FREE with signed license agreement
5. EAPA has endorsed the short 5-item version.

ANALYSIS:

Descriptive Statistics: Table 1 shows the respondents with their average pre and post-test scores reflecting the single-item of the scale for the fiscal year 2016. The range for the single absenteeism item is relatively similar with a slight decrease in range (0 – 176 hours of missed work). The rest of the items are relatively similar although there does seem to be a significant difference in Life Satisfaction between the 2016 and 2017 numbers which is hard to analyze due to the variety of intervening variables during that time period.

A paired samples t-test was conducted on the 2016 sample as well, to compare pre-test to post-test scores in the five measures. There was a significant difference found in the scores for Absenteeism at pre-test (M = 10.27, SD = 20.97) and post-test (M = 3.16, SD = 10.68); t(1,523) = 12,77, p <.001. A significant difference was found in the scores for Presenteeism at pre-test (M = 3.56, SD = 1.45) and post-test (M = 2.75, SD = 1.63); t(1,523) = 18.86, p <.001. A significant difference was found in Work Engagement scores for the 2016 pre-test (M = 3.21, SD = 1.5) and post-test (M = 3.3, SD = 1.43); t(1,523) = -2.26, p =.024. A significant difference was found in the scores for Life Satisfaction at pre-test (M = 3.22, SD = 1.37) and post-test (M = 4, SD = 1.16); t(1,523) = -20.3, p <.001. As well as a significant difference found in Workplace Distress from pre-test (M = 2.39, SD = 1.54) to post-test (M = 2.15, SD = 1.42); t(1,523) = 6.3, p <.001.

Descriptive Statistics: Table 2 shows the respondents with their average scores on pre- and post-tests from the year 2017 reflecting the single-item structure of the scale. The range on the single absenteeism items is from 0 hours to 160 hours of missed work. The remaining four items have a range of 1 to 5 adhering to the Likert scale responses. The means and standard deviations for the four Likert scales show the means to

Table 1. Results for 5-item Workplace Outcome Suite Pre-Use and Post-Use Scores - FOH FY2016

WOS Scale	Pre-Score	Post-Score	Numbers	Raw Difference Score	*p*-value	Differnce Percentage	Effect Size *d*	Effect Size Interpreted
Absenteeism*	**10.27**	**3.16**	**1,524**	**-7.11**	**0.000**	**-69.2%**	-0.34	Small
Presenteeism*	**3.56**	**2.75**	**1,524**	**-0.81**	**0.000**	**-22.8%**	-0.56	Medium
Work Engagement**	**3.21**	**3.30**	**1,524**	**0.09**	**0.024**	**2.8%**	0.06	None
Life Satisfaction**	**3.32**	**4.00**	**1,524**	**0.78**	**0.000**	**24.2%**	0.57	Medium
Workplace Distress*	**2.39**	**2.15**	**1,524**	**-0.24**	**0.000**	**-10.0%**	-0.16	None

approximate the center of the distributions and the standard deviations to reflect some amount of variability around the measures.

A paired samples t-test was conducted on the 2017 sample to compare pre-test to post-test scores for the five measures. There was a significant difference in the scores for Absenteeism at pre-test (M = 10.67, SD = 21.36) and post-test (M = 3.33, SD = 11.44); t(2,196) = 15.44, p <.001. A significant difference was found in the scores for Presenteeism at pre-test (M = 3.48, SD = 1.52) and post-test (M = 2.6, SD = 1.55); t(2,196) = 23.68, p <.001. A significant difference was not found in Work Engagement scores. A significant difference was found in the scores for Life Satisfaction at pre-test (M = 3.23, SD = 1.41) and post-test (M = 3.85, SD = 1.06); t(2,196) = -19.6, p <.001. As well as a significant difference found in Workplace Distress from pre-test (M = 2.26, SD = 1.5) to post-test (M = 2, SD = 1.31); t(2,196) = 8.28, p <.001.

Table 2. Results for 5-item Workplace Outcome Suite Pre-Use and Post-Use Scores - FOH FY2017

WOS Scale	Pre-Score	Post-Score	Numbers	Raw Difference Score	*p*-value	Differnce Percentage	Effect Size *d*	Effect Size Interpreted
Absenteeism*	**10.67**	**3.33**	**2,197**	**-7.336**	**0.000**	**-68.8%**	-0.34	Small
Presenteeism*	**3.48**	**2.60**	**2,197**	**-0.88**	**0.000**	**-25.3%**	-0.58	Medium
Work Engagement**	3.40	3.45	2,197	0.05	0.126	1.5%	0.03	None
Life Satisfaction**	**3.23**	**3.85**	**2,197**	**0.62**	**0.000**	**19.2%**	0.44	Small
Workplace Distress*	**2.26**	**2.00**	**2,197**	**-0.26**	**0.000**	**-11.5%**	-0.17	None

Notes. *Lower scores are a better outcome; **Higher scores are a better outcome. Significant results are bolded.

Table 3: Comparison of 2016 and 2017 Final Results		
WOS Constructs	**FOH 2016 Results**	**FOH 2017 Results**
Absenteeism	69.2%	68.8%
Work Presenteeism	22.8%	25.3%
Work Engagement	2.8%	1.5%
Life Satisfaction	24.2%	19.2%
Workplace Distress	10.0%	11.5%

GENERAL RESULTS AND DISCUSSION

All five items were found to be statistically significant which indicates that they demonstrate a positive response to EAP counseling. Please consult **Table 3**, above, for a comparison of the key results of the 2016 and 2017 data.

Return on Investment analysis is forthcoming and expected to demonstrate significant value to the federal government. Together the five measures make a compelling case for the importance the FOH EAP has in influencing employee performance. The remarkable nearly 70 percent reduction in both the 2016 and 2017 data regarding significant decrease in absenteeism alone highlights the value of offering support to employees that are challenged with personal concerns.

The percentage of increase in presenteeism initially in the 2016 data and slightly more in 2017 data appears lower than the dramatic decrease in absenteeism but perhaps the implications of the variable "presenteeism" are slightly less obvious at this point. Further research is needed to understand fully what this construct translates to in terms of overall employee productivity. The construct presenteeism focuses on "functional impairment," or the ability to attend to work tasks while physically at the job. Already early research has demonstrated that behavioral health concerns (depression, anxiety, stress) are the primary driver of lost productivity, with absenteeism following closely behind (Sullivan, 2017). Similar research is needed to explore this issue more thoroughly. Improvement of at least 23 percent and 25 percent indicates that presenteeism is affected by FOH EAP services which will hopefully result in successfully working with clients to address their concerns and allow them to focus more effectively on the job.

Comparatively, the nearly 3 percent improvement in both the 2016 and 2017 data in work engagement appears to be a seemingly small impact. However, these findings are consistent with other studies that utilize the WOS as an outcome measure (Pompe, 2015). EAP Services may not be a huge contributor towards the concept of increased workforce engagement, whereas the increase in Life Satisfaction demonstrates the impact on their work and life issues or one's general wellbeing. The Life Satisfaction item captures a perceived improvement in one's quality of life or sense of wellbeing (Sharar, Pompe & Lennox, 2012) which bodes well for retention and job satisfaction. Finally, findings show a 10 percent improvement in both 2016 and 2017 in workplace distress combined with the nearly 25 percent improvement in Life Satisfaction in 2016 hopefully will translate to less turnover for these employees. The researchers were unable to obtain turnover rates for FOH employees so hard to completely understand the effect of EAP services on this issue and is more of a hopeful impression. It is noted that the results for effect on Life Satisfaction decreased during the 2017 data collection period. Even though these results were not statistically significant it does make one wonder what other sociological factors may have affected this particular variable during this time period.

CONCLUDING THOUGHTS

In summary, FOH EAP demonstrates significant impact on improving employees' work performance in a variety of dimensions. These findings show promise as one looks to validate the positive effects the FOH EAP has on people's lives. In addition to the services available for individual employees, FOH EAP is a broad-brush program that includes services for supervisors and workplace leaders.

In addition to individual services for employees the EAP provides guidance to Agency leadership in managing the organizational impact of change building a resilient workforce; unlimited consultations with supervisors/managers on performance management concerns and consultation and on-site response to critical incidents. In addition to EAP services, FOH offers comprehensive Work/Life programs that integrate with existing EAP services and further improve the supportive resources for an Agency's employees.

Return on Investment – early results indicate that for every $1.00 that is spent another $1.78 is saved. The national average for ROI for EAPs is 3:1 (Attridge, 2009). Although currently there is a debate about what items should be included in the determination of a Return on Investment (Attridge, 2018), so it is unclear how to evaluate this results at this point in time. Impact on retention of employees who were seen in the FOH EAP might be related to the 25 percent improvement in Life Satisfaction 10 percent reduction in Workplace Distress, although the researchers admit this is simply a guess especially since turnover data at FOH was not made available to the research team. But the question remains whether these results translate into meaningful bottom line issues.

Absenteeism can be monetized but at the moment there is still debate whether the other constructs can be reliably turned into economic outcomes. Although some might argue there are more important factors than just economic ones: work-life satisfaction; workforce engagement; less workplace distress and overall belief that their company cares about their employees, all are areas that most likely effect an employee's work experience. But all these issues need further research to provide evidence-based data on their real impact on workplace productivity.

A final comment is that in some unpublished EAP outcome studies, changes in the WOS scale scores from before to after use of an EAP have been even larger than these reported above in the FOH Study. These findings have led FOH and other companies to ponder whether these differences are due to the characteristics of the particular company's EAP or its delivery model (hybrid, internal, external), the way in which the outcome studies were conducted, or possibly the unique culture of the company or institution being studied.

RECOMMENDATIONS

1) To make outcome measurement ubiquitous and integral in EAP Delivery.
2) WOS Data is one type of evidence to demonstrate the effectiveness and value of an EAP.
3) Results from WOS data can be used to set programmatic goals and targets for process improvement.
4) WOS Data can be utilized to examine effects for specific subpopulations as well as the employee population at large.
5) WOS data can be used to evaluate On-site or Off-site services and how they differ depending on the culture of that particular employee population.

AUTHORS

Jeffrey Mintzer, LICSW, CEA, is Deputy Director, Behavioral Health Sciences for Federal Occupational Health (FOH), headquartered in Bethesda, MD, U.S.A.

Veronica Y. Morrow, MSW, LCSW-C, CEAP, is Associate Director for Federal Occupational Health (FOH), headquartered in Bethesda MD, U.S.A.

Melissa Tamburo, Ph.D, LCSW-C, is Director for Optima EAP, serving more than 350,000 members based in Virginia Beach, VA, U.S.A. Prior to Optima EAP, Melissa was Acting Director, EAP and WorkLife Programs for FOH.

David Sharar, PhD, is CEO of Chestnut Health Systems and Director of Chestnut Global Partners, (a Morneau Shepell company) Division of Commercial Science, Bloomington, IL, U.S.A., is co-creator of the Workplace Outcome Suite© (WOS) and a leading provider of employee assistance services worldwide.

Patricia A. Herlihy Ph.D, RN, is CEO & Founder of Rocky Mountain Research headquartered in Ellicott City, Maryland and Co-Founder of the International Employee Assistance Digital Archive at the University of Maryland.

BIBLIOGRAPHY

Attridge, Mark; DeLapp, Gregory P.; Herlihy, Patricia A.; Ihnes, Pamela; Jacquart, Mike; Lennox, Richard; London, Marina; Servizio, Lou; Sharar, David. (2017). Comparing improvement after EAP counseling for different outcomes and clinical context factors in over 16,000 EAP cases worldwide. (2017). WOS Annual Report 2017. Chestnut Global Partners. Retrieved from: http://hdl.handle.net/10713/7171

Attridge, Mark. (2009). Employee Assistance Programs: A research based primer. Chapter in Cooper, Quick and Schabracq's Book: The Handbook of Work and Health Psychology – 3rd Edition. Wiley-Blackwell.

DeLapp, Greg; Sharar, David and Attridge, Mark. (2017). Workplace impact of EAPs. Presentation: October 4, 2017. EAPA International Conference, LA.

Herlihy, A. Patricia. (2011). The story of integration: Reflections on four research studies. Research Brief for EASNA's KTR Publications. Retrieved from: http://hdl.handle.net/10713/4476

Huddock, William. (2017). Personal Communication.

Lennox, R., Sharar, D., Schmitz, E., and Goehner, D. and Herlihy, P. (2018). Validation; of the 5-item form version of the Workplace Outcome Suite (In press).

Pompe, John; Sharar and Tatcliff, Monica. (2015). Caterpillar's Employee Assistance Program effects of EAP services. *Mental Health Works* – First Quarter. pp.5-8.

Tamburo, Melissa Back and Mintzer, Jeffrey. (2017). Measuring outcomes and mitigating risk with the Workplace Outcome Suite in the federal workplace. White Paper. Retrieved from: http://hdl.handle.net/10713/6658

Selvik, R., Stephenson, D., Plaza, C., & Sugden, B. (2004). EAP impact on work, relationship, and health outcomes. *Journal of Employee Assistance,* 34(2), 18-22.

Sharar, David A., Pompe, John, & Lennox, Richard. (2012). Evaluating the workplace effects of EAP counseling. *Journal of Health & Productivity*. Volume 6 (2). Pages 5-14.

Sullivan, Sean. (2017). EAP Workplace Outcome Suite. From the *Employee Assistance Practice-Based Research Network Series*. Jan. 17.

APPENDIX

Description of the 5 Constructs/Scales:

- **Absenteeism** (looks at the number of hours absent due to a personal problem taking the employee away from work). *"For the period of the past 30 days, please total the number of hours your personal concern caused you to miss work. Include complete eight-hour days and partial days when you came in late or left early."*
- **Presenteeism** (measures decreases in productivity even though the employee is not absent per se but not working at his or her optimum due to unresolved personal problems). *"My personal problems kept me from concentrating on my work."*
- **Workplace Distress** (examines the degree of anxiety or stress at work). *"I dread going in to work."*
- **Work Engagement** (refers to the extent to which the employee is invested in or passionate about his or her job). *"I am often eager to get to the work site to start the day."*
- **Life Satisfaction** (addresses one's general sense of well-being). *"So far, my life seems to be going very well."*

WORKPLACE OUTCOME SUITE - 5 ITEM VERSION

GENERAL INSTRUCTIONS

Below is a series of statements that refer to aspects of your work and life experience that may be affected by the personal problems you want to address at the EAP during the past 30 days. Please read each item carefully and answer as accurately as you can.

	1.	For the period of the past 30 days, please total the number of hours your personal concern caused you to miss work. Include complete eight-hour days and partial days when you came in late or left early.					
	2.	My personal problems kept me from concentrating on my work.	1	2	3	4	5
	3.	I am often eager to get to the work site to start the day.	1	2	3	4	5
	4.	So far, my life seems to be going very well.	1	2	3	4	5
	5.	I dread going into work.	1	2	3	4	5

Copyright © Chestnut Global Partners 2013.07.02

Development and Validation of a Critical Incident Outcome Measure

Richard D. Lennox Ph.D; Dave Sharar Ph.D;
Patricia A. Herlihy Ph.D, RN; Matthew Mollenhauer, MS

ABSTRACT

This paper briefly reviews the current literature in the critical incident response field (CIR) as it intersects with the employee assistance field (EAP). A specific study is presented which illuminates the lack of evidence-based research on the organizational effects arising from traumatic incidents in the workplace. This makes the case for further professional collaboration to seek consensus on uniformity of definitions, intervention and more rigorous measures to establish the degree of effectiveness and efficiency of these interventions. The goal of this paper is to present the initial theoretical construction and statistical validity of the Critical Incident Outcome Measurement Scale (CIOM). The tool was beta tested with a pilot sample of two-hundred and fifty responses. The psychometric approach utilized a latent variable approach to first define the critical incident outcome space with multiple indicators. Next the model capacity was tested to recover most of reliable variance in the item set. Finally, confirmatory factor analysis was used to select the best single indicator for an abbreviated version of this measurement tool to enhance its suitability for applied applications. The specific psychometric results are provided in the conclusion. As important as the authors believe it is to collect this data, it is not nearly as important as tending to the needs of the affected employees and employers of any traumatic event.

KEY WORDS:
EAP, Employee Assistance, Outcome Measures, Measurement, Emotional Distress, Resiliency, Life Satisfaction, Return to Work

INTRODUCTION

This paper introduces the concept of a Critical Outcome Incident Measurement tool (CIOM) for use in the Employee Assistance (EA) Field. To better understand the context and timing of this project, it is important to have some background information both about the field of CIR and how it intersects with the EAP field through discussion of a few key studies. The studies underscore a need to measure the outcomes of efforts to support both individuals and organizations in responding to traumatic events.

BACKGROUND

Critical Incident Response (CIR) is a set of services aimed at ameliorating the effects of traumatic and emotionally charged adverse events that occur, albeit rarely, in the workplace. Examples of such events include bank robberies, sudden onset life threatening behaviors such as physical attacks or suicides, natural disasters, or work-related accidents. Each of these events pose challenges to supporting employees to quickly return to their pre-event level of functioning. Over time CIR services have migrated from an ad hoc set of externally contracted consulting assignments to emerge as an integrated component of Employee Assistance Program (EAP) services. Masi (1994) concluded, *"Irrespective of how CIR emerged as an EAP service, by the early 1990s CIR was considered one of the many functions of the new "broad brush" model of EAP services."* Although the effectiveness of various CIR approaches has been debated over the years (Bisson et. al., 1997; Everly & Mitchell, 2000, Rose et. al. 2003; Ruzek et. al. 2007; Bonanno et. al. 2011, Devilly et. al. 2017), no study has focused directly on the workplace outcomes of EAP services associated with CIR. For a detailed history of the CIR field and related research projects over the last 25 years consult Herlihy's *CIR Literature Review via an EAP*

Lens (2015).

In the current practice environment of evidenced-based science, it is imperative that data demonstrate that an intervention performs beneficially and as intended. Unfortunately, there are two main challenges evaluating EAP-based CIR. The first challenge emerges from the infrequency which such events arise in any one company. In all but the largest EAPs, CIR events are by definition rare in any 12-month period. This means that assessment of the effectiveness of the services may have to cross organizational boundaries and calendar years to gather sufficient data and statistical power to evaluate efficacy. This situation calls for EAP's to use a standardized instrument where responses can be pooled across organizations and over time.

In addition, a growing interest in the topic of resilience among psychological researchers over the past decade has resulted in an expanded set of research methods and intervention approaches. Many of these new approaches focus on the possibility of positive outcomes arising from adverse events which has led to potential trainings to increase individual's resilience. Chan (2012) published a paper on "Improving Resistance and Resiliency through Crisis Intervention Training." It includes a brief survey of students taking the Individual Crisis Intervention Stress Foundation (ICISF) course and found students' levels of resistance and resiliency improved. These results led the authors to ask whether organizations should consider training their employees in life coping skills and crisis intervention practices as a way to strengthen the overall resiliency of their organization.

Following the response to the U.S. terrorist attacks in September 2001, requests for workplace critical incident response (CIR) services have continued to increase (Jacobson, 2006; Jacobson & Attridge, 2011). Employers often rely on their Employee Assistance Programs (EAPs) to provide and coordinate on-site and other CIR services following a workplace crisis or traumatic event. In response to the increased number of service requests, EAPs have also begun to collaborate with private companies that provide workplace CIR services such as Crisis Care Network which is now under the aegis of R3 Continuum.

In 2013 CCN, in collaboration with Jodi Jacobson Frey from the University of Maryland School of Social Work, conducted data mining of EAP case management notes from 132 EAP organizations in the U.S. offering approximately 32,000 CIR Services over a period of five years. There were two main findings. First, three types of traumatic events occurred when CCN was called to provide intervention services:

1) employee death – both natural or accidental
2) bank robberies – both with and without weapon
3) layoffs or downsizing announcements

The second significant finding was that "a successful CIR, as defined by the end user, has as much to do with business objectives as it does with clinical efficacy" (Gorter, 2015, p.12). These CCN results shift the focus from a purely clinical intervention designed to address pathology toward one that facilitates natural resilience, reducing dysfunction, and returning both employees and employers to pre-event functioning (Gorter, 2015).

Attridge and Vanderpol (2010) point out a lack of research studies that concentrate on the effectiveness of CIR on organizational performance. In the early 1990s CISM units were established within some external EAP programs to address this discrepancy to support both individuals and organizations experiencing traumatic events. In 2013, DeFraia conducted an exploratory study of an EAP based CISM unit's data over a twenty-year period (1990 – 2010). The unit responded to approximately 3,000 critical incidents per year and 60,000 incidents over the two decades. Due to an inability to find a scale to accurately assess the severity level of an isolated incident, they developed the Critical Incident Severity Index Scale (CrISIS – R). DeFraia's research contribution was a reminder of how important it is that EAPs CIR response address both individual and organizational issues.

There are many questions facing practitioners and researchers today about CIR services. Is there a model that is more effective for particular events and

demographic populations? Should bank and retail robberies be handled differently than deaths in the workplace? Is there a more nuanced approach that could be flexible and effective in responding to individuals, organizations, cities, and across international borders? To adequately address some of these questions, more data on the outcome effect of current EAP related CIR services is needed. Unfortunately, due to the nature of unpredictability of these events, it is nearly impossible to use a pre-post respondent intervention approach. This paper's authors have developed a Critical Incident Outcome Measurement tool (CIOM) focused on how EAP-CIR services are currently provided and have conducted a beta test to insure the tool's validity and reliability. This paper describes the scale, how its validity has been authenticated, and raises suggestions for next steps. The goal is that with accumulating new data, the field can move forward and refine models of service delivery and document the efficacy of EAP – CIR related services.

RATIONALE FOR THE DEVELOPMENT OF CIOM

A review of the empirical justification of critical incidence intervention is somewhat fragmented, especially measuring the workplace outcomes of services provided. Thus, having a set of more scientifically designed outcome measures at this juncture would be beneficial. Given these variables, a psychometric measurement approach provides the most defensible methodological construction and validation approach. The beta study is designed to provide the basic psychometric foundation for the responses to a short list of critical incidence outcome measure questions. A two-staged approach was utilized. The measures were defined with multiple measures of a latent variable to provide stable estimates of the constructs involved that can then be used to create single item measures of the latent variable based on their factor loading on the confirmatory factor analysis. This approach provided a basis for establishing the empirical lineage of the individual measures in terms of their reliability as measures of the larger latent variables.

BETA TEST

Chestnut Global Partners (CGP) is building on their earlier success with the Workplace Outcome Suite (WOS; Lennox, et al., 2010) by developing an empirically based tool called the CIOM. This new tool expands the original design of WOS to meet the needs of credibly demonstrating and quantifying the positive workplace effects of CIR services offered by EAPs. Similar to the WOS, the CIOM tool explores the five constructs of emotional distress; presenteeism; resiliency; return to work time, and perception of leadership's role (**Exhibit 1**). Presentations

Exhibit 1: Item pool for the Critical Incident Outcome Measure
Absenteeism
1. For the period of the past thirty (30) days, please total the number of hours the incident has caused you to miss work including, complete 8-hour days, and partial days when you came in late or left early. _____
Emotional Distress
2. I feel sad most of the time.
3. I have trouble getting interested in things around me.
4. I don't feel like talking to anyone.
5. I feel like crying a lot.
6. I don't have any energy.
Presenteeism
7. I have a hard time dong my work because of the incident.
8. The incident keeps me from concentrating on my work.
9. I am not able to enjoy my work because of the incident.
10. The incident makes me worried about completing my normal duties.
11. I cannot do my job because of the incident.
Resiliency
12. I actively look for ways to replace the losses I encounter in life.
13. I look for creative ways to alter difficult situations.
14. Regardless of what happens to me, I believe I can control my reactions to it.
15. I believe I can grow in positive ways by dealing with difficult situations.
16. I can usually find ways to deal with most problems I encounter.
Return to Work
17. I believe that I can return to my duties without any interference from the incident.
18. I feel that I can perform my work without any problems from the incident.
19. I feel competent to return to my normal duties.
20. There shouldn't be any problem with doing my regular work.
21. At this point the incident does not affect my ability to work.
Perception of Leadership
22. On a scale of 1 to 10 with 1 being inadequate and 10 being superior, how would you rate the effectiveness of your leadership's reaction to the incident?

and focus groups have been used to help refine this tool as well as determine the level of interest in utilizing this type of tool in the field. Several companies, particularly in the financial and healthcare industries, have expressed interest in this model while others remain hesitant about how the implementation would be conducted while maintaining the confidentiality needed for such services.

In the spring of 2017 a beta test was conducted with a sample of over 250 individual responses in order to demonstrate the scientific validity of the CIOM tool. What follows is the statistical analysis and results from the initial beta test supporting the tool's validity.

BETA TEST METHODOLOGY

Subjects:

Individuals were recruited from a broad range of employees and future employees. In total there were 239 useable and completed responses. The initial psychometric analysis of the CI scale was directed to testing the ability of the items to define a latent variable space that corresponded to the theoretical structure. The definition of the relevant constructs with multi-item scales and two with single item scales are defined below. Our initial analysis is directed at a broad sample space of personnel who experienced a critical incident and those who have not but may in the future. Including respondents who are considering hypothesized incidents assures that the scale was not biased by leaving out important respondents for future analysis. The scale was administered in a single sitting taking anywhere from 5 to 20 minutes depending on the respondent's comfort with the English language due to a diverse subject pool. Subjects were not remunerated for their participation yet, surprisingly, very interested in participating in this process.

STATISTICAL ANALYSIS:

Item Analysis:

The initial analysis was directed at the frequency distribution of the individual items. This is referred to as a distributional sensitivity, and it reflects the ability for items to be consistent with the theoretical distribution. In this case, the items were expected to demonstrate a relatively normally spread bell shape, reflecting two tails of extreme responses at the lower and higher ends of the distribution and occupied by a very small number of extreme scorers. This assumption was based on the sampling frame of normal respondents and that most of the respondents would be in the middle of the distribution. In choosing the test sample, the goal was to purposely select from a wide range of respondents so as not to bias the psychometric analysis that is critical to the initial stage of the scale development.

Scale Analysis:

The second stage of the analysis focused on the reliability of the individual multiple item scales. Tests of internal consistency of the each of the following items: Emotional Distress, Presenteeism, Return-To-Work, and Resiliency were conducted. Coefficient Alpha was used to quantify the internal consistency of the items and to assess the degree of random error found within the item sets.

Construct Validity:

To test the structural validity of the initial scales, maximum likelihood confirmatory factor analysis was conducted on the responses to the candidate question items. The initial hypothesis was that four factors would underlie the covariance matrix of the items comprising the multi-item scales: Emotional Distress, Presenteeism, Return-To-Work, and Resiliency. Single item measures such as absenteeism and perception of leadership were not constructed as multi-item scales and therefore were not included in the confirmatory factor analysis. The assumption being that the four measures are part of a superordinate construct measuring the outcomes of a CI intervention. Therefore, it is assumed that the four measures are inter-correlated with one another but they are not necessarily part of a second order factor. Neither is it expected that the inter-correlations are particularly high, but given the direction of scoring the items, they are expected to be positive. The confirmatory factor analysis thus consists of four latent variables reflecting the covariance among the items from their respective variable.

The scales will therefore be correlated but distinct from one another. Hopefully, the different items will have the ability to identify different factors to offer its use as the basis for asserting discriminant validity (Campbell & Fiske, 1957)

RESULTS

The specific results are identified in the tables below and illustrate the basic psychometric process validating the new CIOM tool. In **Table 1** the reader will find the basic means and rates of standard deviation of each of the items in the critical incident pool. Tables 2 – 5 address the question of reliability of the multi-item scales, including the constructs of

1) emotional distress
2) presenteeism
3) resiliency
4) returning to work time.

The psychometric analysis aims to assure that the items are not confusing or otherwise foreign to the respondents, such as might be expected from respondents for which English is not their primary language or if

Table 1: Critical Incident Item Descriptive		**Numbers**	**Mean**	**Standard Deviation**
Absenteeism				
1.	For the period of the past thirty (30) days, please total the number of hours the incident has caused you to miss work including, complete 8-hour days, and partial days when you came in late or left early. _____	169	4.72	11.16
Emotional Distress				
2.	I feel sad most of the time.	235	2.37	1.31
3.	I have trouble getting interested in things around me.	235	2.54	1.37
4.	I don't feel like talking to anyone.	235	2.87	1.38
5.	I feel like crying a lot.	236	2.07	1.16
6.	I don't have any energy.	235	2.54	1.20
Presenteeism				
7.	I have a hard time doing my work because of the incident.	236	2.24	1.24
8.	The incident keeps me from concentrating on my work.	236	2.34	1.27
9.	I am not able to enjoy my work because of the incident.	236	2.23	1.28
10.	The incident makes me worried about completing my normal duties.	236	2.20	1.26
11.	I cannot do my job because of the incident.	234	1.68	0.94
Resiliency				
12.	I actively look for ways to replace the losses I encounter in life.	233	3.17	1.37
13.	I look for creative ways to alter difficult situations.	235	3.76	1.12
14.	Regardless of what happens to me, I believe I can control my reactions to it.	234	3.83	1.00
15.	I believe I can grow in positive ways by dealing with difficult situations.	234	4.28	0.84
16.	I can usually find ways to deal with most problems in encounter.	236	4.18	0.87
Return to Work				
17.	I believe that I can return to my duties without any interference from the incident.	232	3.80	1.14
18.	I feel that I can perform my work without any problems from the incident.	235	3.82	1.12
19.	I feel competent to return to my normal duties.	235	4.12	1.05
20.	There shouldn't be any problem with doing my regular work.	236	4.07	1.11
21.	At this point the incident does not affect my ability to work.	235	4.01	1.23
Perception of Leadership				
22.	On a scale of 1 to 5 with 1 being inadequate and 5 being superior, how would you rate the effectiveness of your leadership's reaction to the incident?	223	3.68	1.12

Table 2: Reliability Analysis Emotional Distress

Reliability Statistics					
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Items			
0.870	0.869	5			
Inter-Item Correlation Matrix					
	Feel sad	Feel anxious	Worry a lot	Feel like crying	Don't have any energy
Feel sad	1.000	0.712	0.618	0.640	0.424
Feel anxious	0.712	1.000	0.755	0.581	0.472
Worry a lot	0.618	0.755	1.000	0.611	0.446
Feel like crying	0.640	0.581	0.611	1.000	0.441
Don't have any energy	0.424	0.472	0.446	0.441	1.000
Item-Total Statistics					
	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha If Item Deleted
Feel sad	10.030	17.569	0.737	0.586	0.833
Feel anxious	9.840	16.631	0.790	0.675	0.818
Worry a lot	9.519	16.816	0.753	0.619	0.828
Feel like crying	10.312	18.929	0.691	0.499	0.845
Don't have any energy	9.857	20.332	0.518	0.272	0.883

the language was too sophisticated for, say a younger respondent. When used with the assumption that the items measure the same thing, one can assume that they should be highly inter-correlated with one another. If the situation arises that they are not highly inter-correlated, one might consider a random measurement error.

The statistics in **Table 2** represent the results for the Emotional Distress portion of the scale. The top of the table shows that the scale as a whole contains a relatively high internal consistency with a coefficient alpha of .87. This is based on the assessment on the inter-correlations among the entire item set. As shown in the inter-correlations matrix in the middle of the table, all items are positively correlated with one another. This is due to the fact that the questions were designed to essentially measure the same things but with slightly different words for the specific reason of offsetting the random error. The bottom of the table shows the scales characteristic if any one item of the set is removed. The far right of the table shows that some minor improvement might be achieved by removing the *"don't have energy"* phrase. However, this improvement is small and it might be a mistake to make this adjustment based on a single beta test sample. As it is, the .87 it very good level of internal consistency, especially for a small set of items (Nunnally 1978). The results as a whole show the Emotional Distress scale to be relatively unaffected by random measurement error. This is not particularly surprising given the long history of self-assessment constructs such as depression and anxiety that respondents are very familiar with as assessed in the Emotional Distress scale. The analysis suggests that the emotional distress items should prove a strong basis for selecting a single item to represent the latent construct in the final abbreviated scale.

Table 3: Reliability Analysis Presenteeism

Reliability Statistics					
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Items			
0.925	0.926	5			
Inter-Item Correlation Matrix					
	Hard time doing work	Can't concentrate	Not able to enjoy work	Worried about normal duties	Cannot do my job
Hard time doing work	1.000	0.829	0.778	0.671	0.646
Can't concentrate	0.829	1.000	0.805	0.707	0.633
Not able to enjoy work	0.778	0.805	1.000	0.753	0.659
Worried about normal duties	0.671	0.707	0.753	1.000	0.674
Cannot do my job	0.646	0.633	0.659	0.674	1.000
Item-Total Statistics					
	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha If Item Deleted
Hard time doing work	8.459	17.758	0.832	0.733	0.903
Can't concentrate	8.356	17.360	0.851	0.761	0.899
Not able to enjoy work	8.459	17.258	0.857	0.741	0.898
Worried about normal duties	8.506	18.061	0.785	0.637	0.913
Cannot do my job	9.017	20.905	0.722	0.533	0.926

Table 3 contains the reliability analysis for the Presenteeism Scale which is the same scale used in the Workplace Outcome Suite (WOS). The top portion of the table shows the Presenteeism Scale to have high internal consistency with coefficient alpha of .925. This a very high level of internal consistency and typical of the levels seen in the WOS. Here again, the statistics show that the items sets are not heavily affected by random measurement error. The correlations shown in the inter-item correlations are even higher than the one seen in the Emotional Distress Scale. Finally, the bottom of Table 3 shows that none of the items can be removed to create a more internally consistent scale. Here again, the high internal consistency suggests that the item set will provide a good basis for selecting a single item to represent the latent construct in the final abbreviated scale.

The statistics shown in **Table 4** which represent the analysis of the Resiliency Scale are not as impressive as the Emotional Distress and/or the Presenteeism Scales. The .76 alpha coefficient is acceptable, but not particularly strong evidence of freedom from random measurement error. Still, the inter-item correlations matrix shows all items to be positively inter-correlated with one another. The item scale analysis shows that one item *"Replace-the-losses-encountered"* may be the source of some confusion in the items set. The alpha coefficient does increase if it is removed, but here again the improvement is not large and would be a mistake to remove it based on a single sample. The lower level of internal consistency suggests that care will be needed in selecting the best representation of the Resiliency Scale with these items. On the other hand, the marginal performance of

Table 4: Reliability Analysis Resiliency

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Items			
0.761	0.785	5			
Inter-Item Correlation Matrix					
	Replace the losses encountered	Creative solutions for difficult situations	I can control my reactions	Grow by dealing with situations	Deal with most problems
Replace the losses encountered	1.000	0.547	0.219	0.260	0.200
Creative solutions for difficult situations	0.547	1.000	0.384	0.551	0.441
I can control my reactions	0.219	0.384	1.000	0.453	0.439
Grow by dealing with situations	0.260	0.551	0.453	1.000	0.724
Deal with most problems	0.200	0.441	0.439	0.724	1.000
Item-Total Statistics					
	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha If Item Deleted
Replace the losses encountered	16.05	9.116	0.406	0.303	0.787
Creative solutions for difficult situations	15.47	8.682	0.675	0.490	0.661
I can control my reactions	15.41	10.358	0.466	0.255	0.739
Grow by dealing with situations	14.97	10.144	0.647	0.599	0.690
Deal with most problems	15.07	10.404	0.564	0.541	0.712

the *"Replace-the-losses-encountered"* suggests that it would be a mistake to select this item for the single-item measure of resiliency.

Table 5 contains the reliability of the Return-To-Work Scale which produces the highest measure of internal consistency of the critical incident measures as a whole. The .94 coefficient alpha shows the item set to be clear and unambiguous. This is borne out again in the extremely high inter-correlations seen in the inter-item correlation matrix. Finally, the item-total statistics suggests that none of the items can be removed in order to create a more internally consistent item set. The statistics show the Return-To-Work item to be extremely internally consistent, indicative within this particular measurement model of very low random measurement error. In addition, this result provides an excellent rationale for this item being selected for the single item in the final abbreviated scale.

Together, the four sets of reliability analyses, show the item sets to contain a useful mix of items from which to draw the final single item version of the CI outcome measures. In the next phase of the analyses

Table 5: Reliability Analysis Resiliency Return to Work					
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Items			
0.941	0.942	5			
Inter-Item Correlation Matrix					
	Return to duties	Perform my work	Feel competent to return	No problems with regular work	Does not affect ability to work
Return to duties	1.000	0.847	0.720	0.713	0.659
Perform my work	0.847	1.000	0.792	0.810	0.726
Feel competent to return	0.720	0.792	1.000	0.847	0.739
No problems with regular work	0.713	0.810	0.847	1.000	0.797
Does not affect ability to work	0.659	0.726	0.739	0.797	1.000
Item-Total Statistics					
	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha If Item Deleted
Return to duties	16.193	16.245	0.801	0.725	0.935
Perform my work	16.140	15.971	0.882	0.812	0.920
Feel competent to return	15.855	16.591	0.855	0.757	0.925
No problems with regular work	15.895	16.095	0.879	0.805	0.920
Does not affect ability to work	15.934	15.965	0.795	0.662	0.936

we turn to the ability of the four sets of items to define separate but related sub-constructs of the CI behavioral constructs.

CONFIRMATORY FACTOR ANALYSIS:

A confirmatory factor analysis (CFA) was conducted to examine the patterns of loading of four theoretical latent constructs. In this analysis it was hypothesized that the patterns of factor loading should be such that the items should load high on their respective factor and zero on their non-related constructs. For example, the five emotional distress items, should load on one and only one factor that is defined by their covariance with the other five emotional distress items. The same is true of the items for presenteeism, resiliency and Return-To-Work. We establish sets of a theoretical structure with these patterns and model them using Maximum Likelihood and test it against the observed covariance matrix. A goodness-of-fit statistic quantifies our ability to recapture the observed covariance as modeled by the CFA. At the same time, it provides the opportunity to examine how strongly the factor load is on each respective factor. This pattern of loading on the respective factor is indicative of how well each item correlates with the latent factor compared to the other items in the set. The loading provides a good basis for selecting the best representation of the latent factor and the selection of a single item for the final abbreviated CI scale.

Table 6 contains some summary statistics of the CFA including the standardized estimate of the factor loading on the far right of the table. The second column shows which scale each item is expected to load on,

Table 6: Standardized Estimates for the Confirmatory Factor Analysis

Item Labels		Scale	Estimate	*P*	Standardized Estimate
2.	Feel sad	ED	1.572	.000	0.816
3.	Trouble getting interested	ED	1.730	.000	0.861
4.	Don't feel like talking	ED	1.637	.000	0.804
5.	Feel like crying	ED	1.257	.000	0.734
6.	Don't have energy	ED	1.000		0.570
7.	Hard time doing work	PR	1.561	.000	0.877
8.	Keeps me from concentrating	PR	1.648	.000	0.904
9.	Not able to enjoy work	PR	1.640	.000	0.896
10.	Worried about completing duties	PR	1.461	.000	0.810
11.	Cannot do my job	PR	1.000		0.740
12.	Actively look to replace losses	RES	0.694	.000	0.356
13.	Look for creative ways	RES	0.997	.000	0.628
14.	I believe I can control	RES	0.786	.000	0.556
15.	I can grow positive	RES	1.014	.000	0.851
16.	I can find ways to deal	RES	1.000		0.812
17.	I can return to normal duties	RTW	0.937	.000	0.834
18.	I can perform my work	RTW	0.990	.000	0.906
19.	I feel competent to return	RTW	0.935	.000	0.902
20.	Shouldn't be any problems	RTW	0.997	.000	0.922
21.	Doesn't affect my ability to work	RTW	1.000		0.834
Chi-sq. (164) = 396.26, p = .000; TL= 0.911, CFI = 0.931; RMSEA = 0.077					

as a definer of the latent factor. Given the design of the scale, each of the items should have a significant and strong correlation with their respective scale. In this analysis the loading on the non-relevant latent factor is set to zero. For example, the *"feel sad"* item is expected to load strongly on the emotional distress factor but it is fixed to zero on the remaining four factors. As seen in the table this item does load significantly and substantially on the emotional distress factor with a loading of 0.816 (<.001). In fact all of the remaining items in the set show the same pattern of significance with the lone exception of the *"I actively look to replace loss"* items from the resiliency scale, and even this item loads significantly just not as strongly as the others. Four of the items did yield a significance test value due to the fact that they are used to "set the scale" of the latent factor. The lack of "*p*" value does not reduce the value of the analysis due to the resulting high estimate on the factor loading.

The table as a whole shows very high factor loading for all items with the exception of the item mentioned above. The test of the items zero loading can be seen in the goodness-of-fit statistics which captures the validity of "fixed" parameters. As shown in the notes at the bottom, the Chi Square test shows a significant difference between the actual and modeled covariance matrix, the RMSEA which indexes this difference, shows it to be very small (0.077), and yet the fit was still impressive with the Tucker Lewis Index of 0.911 and the comparative fit index of 0.931.

The confirmatory factor analysis supports the discriminant validity of the items in their ability to define four sets of distinct latent factors. So even though some of the factors in the different sets may be highly correlated with one another they still have enough distinctiveness to define the respective factors as different from the others. In summary this portion of the analysis indicates that one can conclude that the item sets are appropriately

Table 7: Intercorrealtions Among Latent Factors

Pairs of Sub-scales	*r*	*p*-value
Emotional distress with presenteeism	0.755	< .001
Presenteeism with resiliency	-0.129	0.083
Emotional distress with resiliency	-0.155	0.044
Return-to-work with resiliency	0.581	< .001
Return-to-work with presenteeism	-0.405	< .001
Return-to-work with emotional distress	-0.245	0.001

classifying into their respective scale, and the virtual items strongly load on their respective latent factors. The lone exception, resiliency, does as well just not as strongly.

This next section deals with the concept of Inter-correlations. **Table 7** displays the inter-correlations of the latent factor pairs. Not surprisingly, presenteeism and emotional distress are strongly correlated with high scores on presenteeism as reflected by higher impairment and greater difficulty with work performance. Similarly, Return-To-Work also correlated strongly with resiliency, indicating some underlying association with one another that is not part of the emotional distress and presenteeism association. In fact, Return-To-Work is negatively related to presenteeism whereas resiliency is uncorrelated with presenteeism. This is also part of a broader patterns of low and negative inter-correlations with resiliency and Return-To-Work. Taken together there seems to be a lack of cohesiveness among the latent variables that suggest a need for consideration of the four constructs as distinct rather than as part of a superordinate construct.

THE FINAL ABBREVIATED CI SCALE: The CFA provides us with the last pieces of information for collapsing this overall scale down to six items or questions. The WOS absenteeism item will again be used here in the simplified scale. These two items are not modeled in a latent variable format and are expected to stand as they were in the questionnaires.

For the four remaining constructs, the best single measure from the CFA will be selected. For example, in the Emotional Distress scale, the best single indicator is number 3 *"I have trouble getting interested in things around me"* with a loading of 0.861. Several other items also do well. The item *"I feel like crying"* doesn't do well, largely because it often doesn't apply to men. Men can be very distressed but not moved to tears because of social stereotypes and an unwillingness to even admit that they might feel like crying. Regardless, the *"trouble getting interested in the things around me"* is the best empirical indicator, and the fact that it is not obviously directed at emotional distress makes it less affected by any emotional distress stigma that may be present and possibly threaten the veracity of self-reports in the workplace. The CFA provides good empirical support for the selection of the more indirect measure over items such as *"I feel sad."*

For the Presenteeism Scale, the best single indicator is item 8 *"The incident keeps me from concentrating on my work"* with a loading of 0.904. Again, there are other close measures in the presenteeism set, which is of course not surprising given its high level of internal consistency. However, the empirical data show that the *"concentrating on my work"* item is the best reflection of the presenteeism item set.

As expected from the reliability analysis, the Resiliency Scale poses some challenges for selecting the best indicator for the single item measures. Most of the factor loading is in the low to moderate range. However, item 15, *"I believe I can grow in positive ways by dealing with difficult situations"* produces a respectable loading of 0.851 and should reflect the variance in the item set adequately.

The five items in the Return-To-Work set also provides an excellent basis for selecting a good single item measure. Unlike Resiliency, the Return-To-Work set

produces three loadings over 0.90. Item 20, *"There shouldn't be any problem with my doing my regular work"* gets right to the heart of the Return-To-Work latent construct with a loading of 0.922.

With the selection of these "best" indicators we have constructed the abbreviated version of the scale (**Exhibit 2**). Although many people prefer multiple items to offset item unreliability, this approach allows us to estimate the amount of random error there is in an item based upon it's loading on the factor. While this information can be used as a formal estimate of reliability in a structural equation model, the value is largely used to justify the use of the single item in a test and/or testing. Another example for using the single item measures of presenteeism allows us to say that the reliability of the item is 0.90 as a measure of the latent construct model in this paper. That statement should be enough to address any serious concern about a Type II error caused by measurement error, especially in a small sample. There are also ways to use the information in addressing the power of a specific test. For example, knowing that the reliability of an item allows us to conduct a power analysis of any study responses knowing the sample size and the reliability of the measure. The analysis provides a rational for reducing the size of the measures to make it manageable in an applied setting.

RETROSPECTIVE REPORTS:

In collecting self-report data it is not always practical or even possible to gain access at the precise time it's desired. Sometimes, research studies have no other alternative but to resort to look-back or use of retrospective reports whereby the subject is asked to recall his/her experience from a time in the past. Retrospective reports are a reasonable alternative when real-time reports are not possible. For example, Collins, et. al. (1985) point out that this approach has been used in the substance abuse field and found to be valid. What's more, Sobell, et. al. (1988) and (Simpura and Poikolainen (1983) found an acceptable level of validity and reliability even for reports in the distant past. Taken together, the drug abuse literature provides sufficient evidence that retrospective reports can be effectively used as a reasonable alternative to concurrent reports.

SUGGESTED PRE-TREATMENT/ POST-TREATMENT ASSESSMENT METHODOLOGY:

There are some inherent problems with conducting pre-treatment/post-treatment critical incident research studies. The most difficult issue arises from the inherent risk of exacerbating the emotional trauma for the employee from the incident. In addition there is an assumption when making an assessment that the reports received are valid and reliable. There is a high probability

Exhibit 2: The Abbreviated Critical Incident Scale

1.	For the period of the past thirty (30) days, please total the number of hours the incident has caused you to miss work including, complete 8-hour days, and partial days when you came in late or left early. _____
22.	On a scale of 1 to 10 with 1 being inadequate and 10 being superior, how would you rate the effectiveness of your leadership's reaction to the incident?
3.	I have trouble getting interested in things around me.
8.	The incident keeps me from concentrating on my work.
15.	I believe I can grow in positive ways by dealing with difficult situations.
20.	There shouldn't be any problem with doing my regular work.

and risk that the trauma will adversely impact what is reported from the event. To address this concern, a recommendation would be to use retrospective reports for the period 30 days prior and one day after the incident. The method thus relies on two retrospective reports cited above. Another survey performed 30 days following the end of the CI treatment would also be useful. This approach contributes a very rich set of measures to assess several dimensions of an event's impact and the CI intervention as follows:

- A retrospective assessment 30 days prior to the incident a normal baseline comparison
- A retrospective report immediately following the event yields the near-term impact
- A concurrent report administered 30 days following the end of treatment provides data on the follow-up for pre-treatment – post-treatment change

CONCLUDING THOUGHTS

This paper presents the initial construction and validation of the Critical Incident Scale to be used as a Critical Incident Outcome Measure (CIOM). A latent variable approach was used to first define the critical incident outcome space with multiple indicators, test the model capacity to recover most of the reliable variance in the item set, and then provide for selecting the best single indicator for an abbreviated version more suitable for applied applications. Such short measures are essential for collecting data related to a critical incident. This requirement is particularly relevant following a severe incident or one which occurs in a setting that is especially unusual for any occurrence of a critical incidents. For example, a bank robbery in a satellite branch located in a rural setting. In a situation like this where there is a great deal of emotional trauma among workers, it is essential to gather the information and support helping employees return to their usual level of functioning as soon as possible. Collecting outcome data is important, but not nearly as important as tending to the needs of affected workers.

PRACTICAL CONSIDERATIONS FOR CIR RELATED OUTCOMES ASSESSMENTS

One of the major challenges for outcome assessment among those impacted by a critical incident is when there is an adverse reaction to the incident that is so traumatic it gets in the way of a rational, and/or valid self-report. This traumatic impact may simply make immediate assessment impossible, especially in those cases of severe trauma where a clinical intervention is recommended. In addition, there is the issue of when it is sensible to collect valid and reliable data from someone experiencing psychological trauma?

Retrospective reports are often used in psychological research when concurrent reports are not available. For example, in some applied research the study team only gains access to respondents after the event has occurred. Retrospective reports are usually viewed as secondary in value to concurrent reports. However, they are used and accepted as a reasonable measurement approach when circumstances warrant. Research on critical incidents clearly falls into this category. There will be times and circumstances that make it necessary to contact respondents a few weeks after the trauma has subsided to ask them to recall their experience. These reports can be contrasted with outcome responses obtained soon after the intervention event and compared as a change score.

A problem for critical incident research is when a large group of subjects have been effected by an incident. These experiences are rare by their nature, yet there is still a need to conduct research and acquire outcome data to understand how services vary amongst CIR events which impact smaller numbers of employees/employers. Collecting and collating results across companies performing CIR services and studying the variety of individual critical events experienced would be potentially valuable. While combining results runs the risk of treating subtle difference events and outcomes as similar, use of a standardized tool will certainly help. On the other hand, pooled response analysis may yield important advantages for generalizing results to a broader population. As with most applied research, results must be interpreted cautiously.

IMPLICATIONS FOR EMPLOYEE ASSISTANCE PROGRAMS

Employee Assistance Programs have experienced severe downward pressure on the price of their services for many years. This has led many companies to give EAP services away free of charge when bundled with other offerings. In part, this is due to a lack of empirical evidence that these programs are cost effective or even capable of producing any change in clients. The Workplace Outcome Suite was designed to provide the EAP community with a standardized psychometrically validated method for evaluating EAP services. There has also been some movement toward demonstrating the effectiveness of a variety of counseling programs and although there are still some challenges to providing empirical proof of efficacy, the field is clearly heading in this direction.

EAP based Critical Incident interventions using the methods proposed in the Critical Incident Outcome Measure tool offers a strong approach for counseling programs as well. Unlike counseling interventions, CI interventions are seen as valuable precisely when used close in time following traumatic situations. Being able to show similar effectiveness results as the counseling programs would provide a strong basis for demonstrating the value for the entire EAP field to the organizations they serve.

AUTHORS

Richard D. Lennox PhD, is founder and President of Work Matters based in New York, U.S.A. Prior to Work Matters, Dr. Lennox was Chief Scientist and Vice President of Commercial Science at Chestnut Global Partners (CGP).

David Sharar, PhD, is CEO of Chestnut Health Systems and Director of Chestnut Global Partners, (a Morneau Shepell company) Division of Commercial Science, Bloomington, IL, U.S.A., is co-creator of the Workplace Outcome Suite© (WOS) and a leading provider of employee assistance services worldwide.

Patricia A. Herlihy Ph.D, RN, is CEO & Founder of Rocky Mountain Research headquartered in Ellicott City, Maryland, U.S.A. and Co-Founder of the International Employee Assistance Digital Archive at the University of Maryland.

Matthew Mollenhauer, MS is Chief Clinical Officer for Chestnut Health Systems in Bloomington, IL. U.S.A., and Managing Director of Chestnut Global Partners (a Morneau Shepell company).

BIBLIOGRAPHY

Attridge, Mark and VanderPol, Bob. (2010). The Business Case for Workplace Critical Incident Response: A Literature Review and Some Employer Examples, *Journal of Workplace Behavioral Health.* 25:132–145. Retrieved from: https://www.researchgate.net/publication/233475988

Bisson, J., Jenkins, P., Alexander, J., and Bannister, C. (1997). Randomized control trial of psychological debriefing for victims of acute burn trauma. *British Journal of Psychiatry.* 171: 78-81

Bonanno, George, Westphal, Maren, and Mancini, Anthony. (2011). Resilience to loss and potential trauma. *Annual Review Clinical Psychology.* 7(1) 1-25. Retrieved from: https://www.researchgate.net/publication/49622474

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.

Chan, Angelina, Yiong Huak, and Kee, Jass. (2012). Improving resistance and resiliency through crisis intervention training. *International Journal of Emergency Mental Health.* 14(2): 77-86. Retrieved from: https://www.science.gov/topicpages/r/resistance+training+intervention.html

Collins, L.M., J.W. Graham, W.B. Hansen, and C.A. Johnson. 1985. Agreement between retrospective accounts of substance use and earlier reported substance use. *Applied Psychological Measurement* 9:301-09. Google Scholar, Link

DeVilly, Grant and Annab, Rachid. (2017). A Randomized controlled trial of group debriefing. *Journal of Behavior Therapy & Experimental Psychiatry.* (In Press)

DeFraia, Gary. (2015). Psychological trauma in the workplace: Variation of incident severity among industry settings and between recurring vs. isolated incidents. *The International Journal of Occupational and Environmental Health.* 6(3), 155-168. Retrieved from: http://hdl.handle.net/10713/4676

DeFraia, Gary. (2013). EAP-Based critical incident stress management: Utilization of a practice–based assessment of incident severity level in responding to workplace trauma. *International Journal of Emergency Mental Health and Human Resilience.* 15(2): 105 – 122. Retrieved from: http://hdl.handle.net/10713/4569

Everly, George and Mitchell, Jeffrey. (2000). The debriefing controversy and crisis intervention: A review of the lexical and substantive issues. *International Journal of Emergency Mental Health.* Special Edition: Pp.211-225. Retrieved from: https://www.scribd.com/document/261721017/The-Debriefindfyhdrtg-Controversy

Gorter, Jeff, Frey, Jodi and O'Brien, Sharon. (2015). Broaden the value of critical incident response. *Journal of Employee Assistance.* 45(3), 10-13. Retrieved from: http://hdl.handle.net/10713/5142

Herlihy, Patricia A. (2015). Critical Incident Response via an EAP Lens: Literature Review. White Paper: Retrieved from: http://hdl.handle.net/10713/5156

Masi, D. A. (1994). Evaluating your employee assistance and managed behavioral care program. Troy, MI: Performance Resource Press. Retrieved from: http://hdl.handle.net/10713/3363

Nunnally, J. C. (1978). *McGraw-Hill series in psychology. Psychometric theory.* New York: McGraw-Hill.

Rose, Susanna, Bisson, Jonathan and Wessely. (2003). A systematic review of single-session psychological interventions (debriefing) following trauma *Psychotherapy Psychosomatic.* 72:176– 184. Retrieved from: https://www.researchgate.net/publication/10720965

Ruzek, J., Brymer, M., Jacobs, A.K., Layne, C., Vernberg, E.M. & Watson, P.J. (2007). Psychological First Aid. *Journal of Mental Health Counseling* 29 (1): 17 – 49.

Simpura, J., and K. Poikolainen. 1983. Accuracy of retrospective measurement of individual alcohol consumption in men: A reinterview after 18 years. *Journal of Studies on Alcohol.* 44(5): 911-17. Google Scholar, Crossref, Medline.

Sobell, L., M. Sobell, D. Riley, R. Schuller, D. Pavan, A. Cancilla, F. Klajner, and G. Leo. 1988. The reliability of alcohol abusers' self-reports of drinking and life events that occurred in the distant past. *Journal of Studies on Alcohol.* 49:225-32. Google Scholar, Crossref, Medline.

Validation of the 5-item Short Form Version of the Workplace Outcome Suite©

Richard D. Lennox, Ph.D, David Sharar, Ph.D,
Eileen Schmitz, David B. Goehner, LCSW

ABSTRACT

This article reports on the validation of the WOS-5 an abbreviated version of the Workplace Outcome Suite© (WOS). For reasons of efficiency and ease of delivery the field was looking for an abbreviated version of the original WOS tool. In this new abbreviated version four of the 5-items correspond to latent variable measures of presenteeism, work engagement, life-satisfaction and workplace distress. These items were selected based on highest factor loading from the original confirmatory factor analysis in the 25-item WOS development study. The fifth item is the single measure of absenteeism created using a formative measures model to count total hours missed by collapsing the total and partial days absent from work. Correlation evidence indicates the 5-item WOS to be a good measurement representation of the 25-item version. Test of sensitivity for three versions of the WOS (WOS-5, WOS-9, and WOS-25) showed the 5-item version to provide comparable sensitivity to change from various EAP service interventions from our pooled dataset. The newly constructed single absenteeism measure for the 5-item scale was shown to be the most sensitive of the various measures, even outperforming the 25-item version. The advantage, while small, was statistically reliable. Together, these results suggest that the 5-item WOS can be used to approximate the 25-item version without excessive loss of reliability, validity or sensitivity.

KEY WORDS:
EAP, Employee Assistance, Workplace Outcomes, WOS, Measurement, Absenteeism, Presenteeism, Workplace Distress, Life Satisfaction

INTRODUCTION

The Workplace Outcome Suite (WOS) was originally designed to provide standardized outcome measures for evaluating the efficacy and effectiveness of Employee Assistance Programs on the following dimensions: absenteeism, presenteeism, work engagement, life satisfaction and workplace distress (Lennox, Sharar, Schmidt, Goehner, 2010). Such an outcomes measure must provide the "sharpest pencil" possible, reliably detecting small effect sizes given the variety in both degree of severity and breath of different problems presenting to current EAP providers. There is also an expected variance in the types of services offered and the quality of the intervention provided by counselors, especially when call center affiliates are involved.

The 25-item WOS has demonstrated a robust degree of statistical sensitivity to change in such EAP evaluations with as few as 50 observations. Although short when compared to other measurement tools, many EAPs consider the 25-item WOS as too long for regular use in routine outcome monitoring. As a response to these comments, we developed a 5-item version of the WOS that takes one question from each of the original 25-item WOS scales of Presenteeism, Work Engagement, Life Satisfaction and Workplace Distress. Traditional scaling techniques allowed us to select the best representation of each latent variable using the confirmatory factor analysis reported in the original 25-item WOS validation (Lennox, et al., 2010).

Since factor loading cannot be used to reduce the Absenteeism scale, the WOS-5 includes a new single item based more on the semantic meanings of language than on earlier empirical findings. For that reason, the original validation analysis could not be confidently extended to the WOS-5 so entirely new data was collected for validation purposes. Psychometric theory leads one to expect a loss of reliability and subsequent

predictability when the Presenteeism, Work Engagement, Life Satisfaction, and Workplace Distress scales are reduced from 5 items down to a representative one. There is also reason to expect a loss of validity in the original five Absenteeism scales collapsing into a single new item. This article examines the validity and sensitivity of the WOS-5 5-item version of the tool as compared with the validation findings of the original 25-item WOS scale.

ORIGINAL WOS MEASUREMENT MODELS

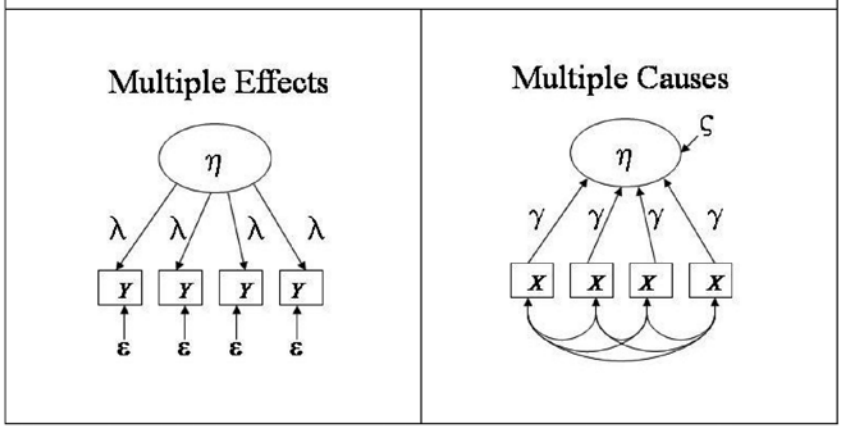
The original measurement models for the 25-item WOS (Lennox, et al. 2010) were derived from traditional psychometric theory and practice. Bollen & Lennox (1991) provide two illustrations of the models in terms of the items in the measures and their expected directional relation to the underlying latent variables. As seen in the left section of the figure below, the multiple effect model assumes that it is the underlying construct that causes the variance in the items *(Y)*. In psychology these items would be referred to as symptoms and manifestations of the latent construct. On the right side of the figure, the direction of cause goes the other way where formative items *(X)* combine to create or cause the latent variables. They are not manifestations or symptoms of the latent construct but the building block for a broader construct. For example, in our absenteeism scale we use five different ways in which an employee may be away from his or her job – missing full days, being late, leaving early, and being actively engaged in activities at the workplace related to his or her problem rather than on their job. The measure combines these different responses in a summed assessment of total absenteeism. In this way, the total absenteeism score is more comprehensive than the particular response given.

Figure 1 depicts the creation of the 5-item WOS scale from the original 25-item version and sets up the soundness of the short version.

While the pictures look similar, it is the directionality of the arrows linking the items to the underlying construct that differ, as well as the scoring mechanics used. Connecting multiple effects to their underlying constructs involves adding together similar items that vary in their random measurement error. The statements used in the tool have just slightly different wording aimed at the same key point. This kind of model approach provides a basis for selecting the best single item to represent the latent construct from what are essentially highly similar measures.

The approach followed in abbreviating the 25 items came in two distinct steps. The first involved creating single measures for the four effect-indicator models: presenteeism, work engagement, life satisfaction and workplace distress. This involved the relatively straightforward process of selecting the item from each sub-scale that had the

Figure 1: Original Measurement Models for the Workplace Outcome Suite



highest loading on their respective latent variables from the two samples in the original factor analysis (Lennox, et al., 2010). This is a common approach to selecting single indicators and can be shown empirically to be the best representation chosen from the item set. They can also be evaluated in terms of item-total correlations in a reliability analysis.

The second step followed in the abbreviation effort involved selecting one of the five absenteeism scale items. Selecting a single item from a cause indicator model can be challenging since the items are different and selecting the wrong one could lead to a serious bias in the final measure. Creating a single item absenteeism scale from the original 5-item scale required forming a completely different type of interim measurement tool to test.

A 9-item version of the WOS was created that retained the original five items of the absenteeism scale along with the single items for presenteeism, work engagement, life satisfaction and workplace distress. The 9-item WOS instrument detected change results similar to the full 25-item version using a pre-treatment/90-day post-treatment EAP interventions follow-up evaluation.

ABBREVIATING THE 5-ITEMS ABSENTEEISM SCALE

The original construction of absenteeism was based on a cause-indicator measurement model that could not be abbreviated by selecting the best single indicator due to the fact that the items were not designed to parallel one another. The original 5-item absenteeism scale is presented below:

Please report for the period of the past thirty (30) days the total number of hours your personal problem:

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.

The first step involved removing from consideration the last two items that referred to actually being at work, but away from one's regular job-duties while attending to one's personal problems. We reasoned that these two ideas were more closely aligned with presenteeism than absenteeism and whatever variance there was in these items was probably picked up by the presenteeism scale. One can certainly argue that the two items removed may be considered "bridge items" that span the absenteeism and presenteeism constructs and, as such, are part of both. The positive correlation between absenteeism and presenteeism reported in Lennox, et al. (2010) supports this view.

That left three items each of which measures time away from the job in full days, hours leaving early, or hours arriving late. This effectively defines "absent from work" as "not physically at the workplace". This definition collapses the first three items to simply asking about the total hours absent.

Regarding the two items removed, these may be considered "bridge items" that span the absenteeism and presenteeism constructs and as such are part of both. Yet, to define them as measures of presenteeism, then they would be the only presenteeism items measured in terms of "hours" instead of a Likert format.

CHALLENGES:

As stated previously, the purpose of this article is to investigate the validity of the 5-item version of the WOS as a time-saving and reasonable alternative to the full 25-item scale. Classical psychometrics predicts that the 5-item version will be less reliable than the full scale and more likely to be affected by greater random measurement error. The study began by examining the extent to which the 5-item construct captured the variance found in the 25-item version.

An additional goal of the investigation is to explore the degree to which any decrease in instrument reliability poses threats to the outcomes of certain types of research studies. For example, the effect of an increase in random measurement error may be offset by an inferential test of the WOS hypotheses by simply increasing the sample size. Prudence would then dictate that the 5-item version, with an expected decrease in reliability, might not be a good candidate for use in efficacy tests of small sample studies. Such research efforts would be better served by using the fine grain details from the original 25-item scale or be cautious when interpreting any non-significant finding due to the potential lack of statistical power.

METHOD

The validation investigation method followed is based on a quasi-experimental or correlational approach. Unlike a randomized clinical trial, no attempt is made to manipulate an independent variable to observe its effect on a dependent variable. The approach used here measures the constructs and assesses the covariance among scores from the same subject's responses. The 5-item WOS was evaluated for its correspondence to the 25-item and 9-items versions in terms of simple correlations and the similarity with which the various scale versions correlate with external criteria.

DATA SOURCES

Sample 1: Data from the paper-and-pencil study came from 200 clients served by Personal Assistance Services of St. Louis Missouri. Personal Assistance Services (PAS) began providing Employee Assistance Program (EAP) services in 1982 and has been a pioneer in transforming the traditional EAP model into a progressive risk management strategy to help employees negotiate and face many life events. PAS specializes in customizing innovative service designs that help keep employees healthy and productive on the job. Subjects were recruited by PAS as they arrived for routine in-clinic visits. To capture the broadest range of employee responses, no exclusion criteria were used. Subjects were simply asked to provide answers to the 25-item WOS tool on the questionnaire form. Responses were then transferred to an excel spreadsheet for compilation and analysis.

Sample 2: Data for the telephone interview modality was provided by 210 clients of Empathia, Inc. Empathia provides behavioral health solutions aimed at improving the well-being, safety and productivity of organizations and individuals. The company collaborates with private and public-sector entities ranging from Fortune 500

corporations to small businesses addressing a range of services. This range of services includes: employee assistance, disaster response and planning, workforce well-being, employee relations, leadership development, training, and benefits support.

Subjects were recruited as they participated in routine telephone contact with an Empathia clinician. The interviewer read the instructions and then asked the 25-item WOS questions in the order they appeared on the paper-and-pencil version, recording each response on the hard-copy questionnaire. Completed questionnaires were mailed to Chestnut Global Partners and entered into an excel spread sheet for analysis. During the pilot stage of the data collection at Empathia, respondents expressed frustration at being asked to recall their absenteeism for the previous 30 days. They claimed to simply be unable to remember the exact details. A modification was then made to switch to the Food Drug Administration's (FDA) standard recall period of 7 days for this site while leaving the paper-and-pencil sample site at 30 days to provide a basis for comparison.

RESULTS AND DISCUSSION

Although the methods and analysis included several WOS versions to capture employee responses, the paper's results focus on what was learned from exploring the validity of the 5-item WOS. First, the paper addresses the performance results using the single absenteeism item in the 5-item scale constructed by collapsing "away-from-work" to asking if the employee was absent a full day, arrived late or left early. The remaining analysis then focuses on the cross-validation results derived from the original 25-item WOS validation study (Lennox, et al., 2010).

Descriptive Statistics: Table 1 presents the descriptive statistics for the 5-item version of the WOS. The top portion of

Table 1: Descriptive Statistics

Empathia Sample					
Pre-test	N	Minimum	Maximum	Mean	Standard Deviation
Absenteeism	210	0	160	10.91	20.423
Presenteeism	210	1	5	3.48	1.349
Work Engagement	210	1	5	2.89	1.360
Life Satisfaction	210	1	5	2.48	1.223
Workplace Distress	210	1	5	2.51	1.439
Valid N	210				
PAS Sample					
Pre-test	N	Minimum	Maximum	Mean	Standard Deviation
Absenteeism	204	0	128	5.03	14.531
Presenteeism	200	1	5	2.79	1.340
Work Engagement	201	1	5	3.11	1.242
Life Satisfaction	201	1	5	2.91	1.156
Workplace Distress	199	1	5	2.18	1.298
Valid N	210				

Table 1 presents the Empathia sample.

Table 1 shows the 210 Empathia respondents with their minimum and maximum scores reflecting the single-item structure of the scale. The range on the singe absenteeism items is from 0 to 160 hours. The remaining four items have a range of 1 to 5 adhering to the Likert scale responses. The means and standard deviations for the four Likert scales show the means to approximate the center of the distributions and the standard deviations to reflect some amount of variability around the measures. The mean and standard deviation of the absenteeism question suggests the typical skewed distribution toward the lower end of hours missed. The maximum 160 hours is an extreme case. This skewed distribution is typical of some health behavior measures that capture data on workplace absenteeism and health care utilization.

Bivariate Correlations: The bottom portion of Table 1 show the same results for the PAS sample. The sample size of between 199 to 204 reflects a small amount of data missing from the responses. The means and standard deviation show a pattern similar to the Empathia sample on the four Likert items, with a slightly lower mean and standard deviation for the absenteeism measure. Taken together, the descriptive statistics show the expected central tendency and dispersions levels, including the skewed distribution on the absenteeism measure found in both samples.

The top of **Table 2** presents the bivariate correlations between and among the 5-item short scale, and the 25-item full scales of the WOS. It serves as the basic test of the capacity of the 5-item scales to actually stand-in for the 25-item version when a short scale is used. The correlations presented on the main diagonal of the matrix present the association between each of the sub-scales measures and the two different methods of measuring responses. As expected all of the correlations are statistically significant beyond the .000 level as they measure the same thing. Even so, they are expected to be affected by some degree of random measurement error. In Sample 1, all correlations exceed .80 except work engagement which is greater than .6. This strong correspondence shows the measures can be used interchangeably without losing too much precision. Even the work-engagement question in the 5-item scale does not appear to lose significant precision. For the four effect-indicator measures, the correlation results are not surprising. Each of the four measures come from the parent measure so the correlation is actually the correlations of the specific items and the respective total of the items in the 25-items sub-scale.

The absenteeism item for the 5-item scale is entirely new and is not affected by common item variance. Therefore, the resulting correlation of .927 is a bit surprising given that some of the items were removed and two items collapsed. The high correlation suggests the single collapsed and truncated absenteeism items are an adequate proxy for the entire 5 items contained in the absenteeism section of the full WOS.

The bottom of Table 2 presents the bivariate correlations between the full WOS and the individual items of the 5-item WOS for the PAS sample. In general the pattern of correlations in the PAS sample parallel those in the Empathia sample. Although there are some differences in strength of correlation, the pattern of statistical significance in the two samples is virtually identical. Not only did the two samples produce parallel results of statistical significance, they also produced parallel results of non-significance. The results for the single absenteeism items are also parallel across the two samples. The one slight difference is in the correlations between the 5-item absenteeism scale from the 25-item WOS and the newly constructed single absenteeism item of the 5-item WOS scale. Specifically, the correlations for the Empathia sample was a very strong at .927 ($p>.000$) while slightly less at .626 ($p<.000$) in the PAS sample. While the absenteeism item showed some shrinkage from the Empathia to, the PAS sample, it still produced a significant and strong correlation. The difference does not pose a significant threat to the validity of the newly constructed absenteeism items.

The last line of the sample reports the efficiency measures for each of the five-item scales relative to their respective longer 25-item scale. Efficiency = (the number of short-form items, in this case the number of items in the full scale) divided by the diagonal

Table 2: Correlations between the 24- and 5-item version of WOS

Empathia (N = 210)					
5-item WOS Scale					
Full WOS	Absenteeism	Presenteeism	Work Engagement	Life Satisfaction	Workplace Distress
Absenteeism	.927***	.287***	-.178**	-.174**	.182**
Presenteeism	.286***	.863***	-.389***	-.380***	.422***
Work Engagement	-0.85	-.260***	.683***	.214**	-.391***
Life Satisfaction	-.123	-.431***	.159*	.823***	-.212**
Workplace Distress	.169*	.405***	-.594***	-.262***	.895***
Efficiency	0.22	0.23	0.29	0.24	0.22
PAS (N = 200)					
5-item WOS Scale					
Full WOS	Absenteeism	Presenteeism	Work Engagement	Life Satisfaction	Workplace Distress
Absenteeism	.626***	.312***	-.059	-.124*	.037
Presenteeism	.199**	.897***	-.346***	-.348***	.371***
Work Engagement	.019	-.219**	.754***	.259***	-.339***
Life Satisfaction	-.042	-.413***	.310***	.724***	-.246***
Workplace Distress	.114*	.373***	-.519***	-.185**	.929***
Efficiency	0.32	0.22	0.27	0.28	0.22

Note: *** indicates p <.000; ** indicates p <.01; * indicates p <.05

correlation between the two measures. The lower the result number, the better the efficiency since fewer items are associated with a higher correlation. The efficiency of the measures also supported the 5-item scale with all measures exceeding .20 suggesting that the items capture the meaning of the individual construct. As can be seen on the last rows of the individual matrices in Table 2, all efficiency values are greater than .20 indicating a good level of efficacy for all single-item scales. The one exception is in the absenteeism scale which exceeds .30 in the PAS sample. This efficiency measure is not particularly high and it does not cross-validate to the other sample.

Table 3 presents the correlations between the individual scales of the 5-item WOS with several external criterion measures. The criterion measures represent behaviors and feelings that are expected to be differentially related to the five single-item measures. The Empathia sample shows significant correlations with all five WOS measures for trouble getting out of bed, feeling sad and falling behind at work. These three

Table 3: Pearson Correlations of the 5-item WOS with External Criterion Variables

Empathia (N = 210)					
Criterion Variables	Absenteeism	Presenteeism	Work Engagement	Life Satisfaction	Workplace Distress
1. Getting out of bed	**.283*** **	**.312*** **	**-.235*** **	**-.287*** **	**.327*** **
2. Feel sad	**.204*** **	**.369*** **	**-.241*** **	**-.456*** **	**.246*** **
3. Falling behind	**.232*** **	**.372*** **	**-.376*** **	**-.197** **	**.373*** **
4. Rarely late for work	**-.231*** **	**-.199** **	.099	.127	-.049
5. Working after hours	0.78	.055	**-.132** **	.058	.107
PAS (N = 200)					
	Absenteeism	Presenteeism	Work Engagement	Life Satisfaction	Workplace Distress
1. Getting out of bed	.063	**.333*** **	**-.250*** **	-.092	**.438*** **
2. Feel sad	**.183** **	**.506*** **	**-.196** **	**-.354*** **	**.374*** **
3. Falling behind	.026	**.417*** **	**-.311*** **	-.126	**.506*** **
4. Rarely late for work	-.121	-.137	-.001	.012	-.109
5. Working after hours	.099	**.157** **	.082	-.048	**.152** **

Note: *** indicates p <.000; ** indicates p <.01; * indicates p <.05

behaviors are measures of constructs that may well translate into problems at work. All are positively associated with absenteeism, presenteeism, and workplace distress and are negatively associated with work engagement and life satisfaction.

The rarely-late-for-work and working-after-hours responses show significance for the absenteeism and presenteeism scale but fail to correlate with work engagement, life satisfaction or workplace distress. The working-after-hours questions only correlated significantly with work engagement. The fact that these two items do not correlate with all measures offers some support for the discriminant validity of the scale.

An attenuated pattern of significant correlations with these three criterion measures is found in the PAS sample. All correlations are in the expected direction, but some fail to reach statistical significance. This pattern offers some limited support for the construct validity for all of the single item measures in the 5-item WOS. It is important to point out that all items are self-reported and subject to some level of measurement bias. The research reported thus far focuses on the correlations of the 5-item WOS scales to other forms of the WOS and some self-reported criteria. In the next section of the report we shift to considering the relative sensitivity of the measures to change across types of EAP interventions.

Table 4 presents the results of pooled comparison of the WOS version across EAP interventions.

Tables 4a – 4d presents the results of a sensitivity analysis of the three versions of the WOS. The analysis is based on a pooled collection of the results from several pre-treatment and post-treatment comparisons of

EAP interventions. The quasi-experimental analysis does not use randomized assignment or control groups. The analysis simply points to each version of the WOS to detect change regardless of it actual cause.

Table 4a shows the sensitivity analysis results for the 25-item WOS compiled from several studies. However, not all studies reported chose to use all the scales which helps explain the different sample sizes noted in the analysis. The samples approximate 1,000 study responses which should provide substantial statistical power for detecting the comparison between pre-treatment post-treatment change. In fact, for the 25-item version, all WOS scales produced statistical significance change from the baseline except for work engagement. It did produce change however in the proper direction and the difference reached the traditional level of statistical significance. All other scales showed statistical change at the .000 level.

Table 4b presents the results for the 9-item WOS version. Again, the 9-items version contains all five of the original absenteeism items and a single item from each of the remaining four Likert scales selected on their factor loading in the original validation study (Lennox, et al., 2010). The 9-item scale results were based on an analysis of approximately 3,300 responses. This large sample produced statistically significant change for the five-item absenteeism scales and for the single item measure for the remaining scales. Work engagement was statistically significant for all measures at the .000 level.

Table 4c presents the results for the 5-item WOS scale based on a study sample of approximately 4,400 responses. The rationale for the modification of the absenteeism scale into a single item was presented earlier. They used this new collapsed measure and the same four single items of presenteeism, work engagement, life satisfaction and workplace distress as used in the 9-item version. The results mirrored the 9-item WOS scale, producing statistically significance change scores for all scales at the .000 level.

Finally, **Table 4d** shows the various definitions combined into a single pool with analysis extracting items from their different scale scores where possible. Results show the measures to be able to detect statistically significant change at .000 levels for all WOS scales.

It is also useful to consider the effect size of the intervention in addition to looking at the statistical significance of the various comparisons by examining the size of test statistic itself. As described earlier, the single absenteeism item used in the 5-item WOS performed well.

In summary Tables 4a – 4d shows the expected difference at pre-score, post-score, and the raw difference score. The percentage difference flips around to show a bigger percent improvement for the single item a difference that remains after using the *z* score. The data suggest that the 1-item version is more sensitive to change than

Table 4a. Results for Workplace Outcome Suite 25-Item Pre and Post-Test Scores

Wos Scale	Pre Score	Post Score	N	Raw Difference Score	t^a	*p*-value	Difference Percentage
Absenteeism*[b]	**12.89**	**6.81**	**950**	**-6.08**	**-8.83**	**0.000**	**-47%**
Presenteeism*	**14.13**	**11.20**	**1,292**	**-2.93**	**-16.99**	**0.000**	**-21%**
Work Engagement**[b]	17.94	17.78	932	-0.16	-1.19	0.235	-1%
Life Satisfaction**	**12.27**	**13.64**	**1,288**	**1.37**	**-12.92**	**0.000**	**11%**
Workplace Distress*	**13.16**	**11.96**	**1,287**	**-1.20**	**-8.45**	**0.000**	**-9%**

Notes: *Lower scores are a better outcome; **Higher scores are a better outcome. Significant results are bold. Presenteeism, work engagement, life satisfaction and workplace distress are the single item scores across all 3 versions of the WOS. [a]Wilcoxon signed rank test used to test change in absenteeism. The Z statistic is reported.

Table 4b. Results for Workplace Outcome Suite 9-Item Pre and Post Test Scores

Wos Scale	Pre Score	Post Score	N	Raw Difference Score	t^a	*p*-value	Difference Percentage
Absenteeism*	**11.70**	**6.30**	**3,316**	**-5.4**	**-22.34**	**0.000**	**-46%**
Presenteeism*	**3.51**	**2.53**	**3,312**	**-0.98**	**-41.21**	**0.000**	**-28%**
Work Engagement**	**3.19**	**3.50**	**3,312**	**0.31**	**16.34**	**0.000**	**10%**
Life Satisfaction**	**2.73**	**3.56**	**3,312**	**0.83**	**39.76**	**0.000**	**30%**
Workplace Distress*	**2.40**	**2.03**	**3,309**	**-0.37**	**-18.38**	**0.000**	**-15%**

Table 4c. Results for Workplace Outcome Suite 5-Item Pre and Post Test Scores

Wos Scale	Pre Score	Post Score	N	Raw Difference Score	t^a	*p*-value	Difference Percentage
Absenteeism*	**5.30**	**2.38**	**4,333**	**-2.92**	**-25.43**	**0.000**	**-55%**
Presenteeism*	**3.25**	**2.31**	**4,453**	**-0.94**	**-40.37**	**0.000**	**-29%**
Work Engagement**	**3.25**	**3.45**	**4,448**	**0.20**	**9.43**	**0.000**	**6%**
Life Satisfaction**	**3.05**	**3.75**	**4,449**	**0.70**	**33.92**	**0.000**	**23%**
Workplace Distress*	**2.05**	**1.75**	**4,441**	**-0.30**	**-16.26**	**0.000**	**-15%**

Table 4d. Results for Workplace Outcome Suite Pre and Post Test Scores Pooled Across Versions

Wos Scale	Pre Score	Post Score	N	Raw Difference Score	t^a	*p*-value	Difference Percentage
Absenteeism 9 and 25-item WOS versions*	**11.97**	**6.42**	**4,266**	**-5.55**	**-23.87**	**0.000**	**-46%**
Absenteeism 5-item WOS versions*	**5.30**	**2.38**	**4,333**	**-2.92**	**-25.43**	**0.000**	**-55%**
Presenteeism*	**3.31**	**2.40**	**9,056**	**-0.91**	**-58.78**	**0.000**	**-27%**
Work Engagement**	**3.22**	**3.44**	**8,689**	**0.22**	**16.14**	**0.000**	**7%**
Life Satisfaction**	**2.94**	**3.64**	**9,041**	**0.70**	**51.70**	**0.000**	**24%**

Notes: *Lower scores are a better outcome; **Higher scores are a better outcome. Significant results are bold. Presenteeism, work engagement, life satisfaction and workplace distress are the single item scores across all 3 versions of the WOS. [a]Wilcoxon signed rank test used to test change in absenteeism. The Z statistic is reported.

the 5-items, and thus probably more valid. However, it would be a mistake to rely on the either the raw pre or post scores alone as a basis for effectiveness for the single item measure because of the differences in scale or the difference in the hours missed. These differences can be corrected by using a change score standardized as a t statistic.

CONCLUSIONS

This paper reports on the validation results for an abbreviated and modified version of the Workplace Outcomes Suite. Measures of presenteeism, work engagement, life satisfaction, and workplace distress were selected from the initial confirmatory factor analysis that best represented their respective factor with the effect-indicator measurement model. Absenteeism was redesigned as a collapsed version of the original three items and now captures data on full and partial days taken off from work due to personal problems. The WOS-5 also removed two other scale items regarded as more in line with presenteeism than absenteeism.

Results from the correlational analysis show that the single absenteeism scale correlates highly with the other five absenteeism scales. The pattern of correlations is also consistent with the 25-item and 9-item versions of the scale. Tests of sensitivity of the various measures for evaluating EAP intervention effects show the 5-item WOS to parallel results from the longer tools. The new collapsed absenteeism items works slightly better than other absenteeism measures, especially when considering the effect size and statistical significance.

Finally for a straightforward test of the effectiveness of an EAP program, either the WOS-5 or the full 25-item WOS work well with the latter being more sensitive for use with smaller sample groups. Although the shortened scales would be suggested by psychometric theory to be less reliable, comparisons of the pooled data suggest they work very well. Taken together the correlational analysis and the pretreatment-post-treatment comparison of the pooled data show the 5-item WOS to be a reliable and valid measure when testing for the outcome constructs.

AUTHORS

Richard D. Lennox PhD, is founder and President of Work Matters based in New York, U.S.A. Prior to Work Matters, Dr. Lennox was Chief Scientist and Vice President of Commercial Science at Chestnut Global Partners (CGP).

David Sharar, PhD, is CEO of Chestnut Health Systems and Director of Chestnut Global Partners, (a Morneau Shepell company) Division of Commercial Science, Bloomington, IL, U.S.A., is co-creator of the Workplace Outcome Suite© (WOS) and a leading provider of employee assistance services worldwide.

Eileen Schmitz is Vice President at Personal Assistance Services (PAS), an Employee Assistance Program serving organizations globally located in St Louis, Missouri, U.S.A.

David B. Goehner, LCSW, is Director of Privacy & Compliance for Empathia, Inc. headquartered in Waukesha, WI, U.S.A.

BIBLIOGRAPHY

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

Burgess, K., Lennox, RD., Sharar, D, & Shtoulman, A (2015). A Substance Abuse Intervention Program at a Large Russian Manufacturing Worksite (2015)., *Journal of Workplace Behavioral Health,* 30:138–153.

Sharar, D., Pompe, J. & Lennox, R. (2012). Evaluating the workplace effects of EAP counseling: *Journal of Health and Productivity.* 6:2 5-14.

Lennox, R., Sharar, D., and Burke, J. (2010). Conducting an EAP evaluation using the Workplace Outcome Suite. *Journal of Employee Assistance,* Volume 40, No. 4, 24-27.

Lennox, R., Sharar, D., Schmitz, E., and Goehner, D. (2010). Development and validation of the Chestnut Workplace Outcome Suite. *Journal of Workplace Behavioral Health*, 25, 107-131.

Lennox, R., Sharar, D., Burke, J. (2009). Measuring workplace outcomes in EAPs. *Journal of Employee Assistance,* Volume 39, Number 4, 18-19.

Sharar, D., Lennox, R. (2009). A new measure of EAP success: A workplace outcomes-based survey model provides validation. *Society for Human Resource Management Online,* 9/24/09.

APPENDIX

CGP WORKPLACE OUTCOME SUITE (WOS)							
	GENERAL INSTRUCTIONS Below is a series of statements that refer to aspects of your work and life experience that may be affected by the personal problems you want to address at the EAP during the past 30 days. Please read each item carefully and answer as accurately as you can.						
	INSTRUCTIONS FOR ITEMS 1 - 5 Please report for the period of the last 30 days the total number of hours your personal problems.		NUMBER OF HOURS				
ABSENTEEISM	1.	Caused you to miss work altogether.	1	2	3	4	5
	2.	Made you late for work.	1	2	3	4	5
	3.	Caused you to take off early.	1	2	3	4	5
	4.	Pulled you away from your normal work location.	1	2	3	4	5
	5.	Required you to be on the phone, e-mail or internet while at work.	1	2	4	4	5
.	**INSTRUCTIONS FOR ITEMS 6 - 25** The following statements reflect what you may do or feel on he job or at home. Please indicate the degree to which you agree with each of the statements for the past 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 personal problems.	1	2	3	4	5
	7.	My personal problems kept me from concentrating on my work.	1	2	3	4	5
	8.	Because of my personal problems I was not able to enjoy my work.	1	2	3	4	5
	9.	My personal problems made me worry about completing my tasks.	1	2	3	4	5
	10.	I could not do my job well because of my personal problems.	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 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 is going very well.	1	2	3	4	5
	19.	There isn't anything I would change about my life 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 the 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

APPENDIX

WORKPLACE OUTCOME SUITE - 5 ITEM VERSION							
	GENERAL INSTRUCTIONS Below is a series of statements that refer to aspects of your work and life experience that may be affected by the personal problems you want to address at the EAP during the past 30 days. Please read each item carefully and answer as accurately as you can.						
			NUMBER OF HOURS				
AB	1.	For the period of the last 30 days, please total the number of hours your personal concern caused you to miss work, Include complete eight-hour days and partial days when you came in late or left early.					
	INSTRUCTIONS FOR ITEMS 2 - 5 The following statements reflect what you may do or feel on he job or at home. Please indicate the degree to which you agree with each of the statements for the past 30 days. Use the 1 - 5 response key to the right.		STRONGLY DISAGREE	SOMEWHAT DISAGREE	NEUTRAL	SOMEWHAT AGREE	STRONGLY AGREE
PR	2.	My personal problems kept me from concentrating on my work.	1	2	3	4	5
WE	3.	I am often eager to get to the work site to start the day.	1	2	3	4	5
LS	4.	So far, my life seems to be going very well.	1	2	3	4	5
WD	5.	I dread going to work.	1	2	3	4	5
							Copyright © Chestnut Global Partners 2013.07.02

Measuring Coaching Effectiveness: Validation of the Workplace Outcome Suite for Coaching

Richard D. Lennox, Ph.D; David Sharar, Ph.D; Francine Miller

KEY WORDS:
Workplace Outcomes, WOS, Coaching Call Centers, Disease Management, Depression, Work Engagement, Absenteeism, Presenteeism, Workplace Distress, Life Satisfaction

ABSTRACT

Chestnut Global Partners developed the coaching version of the Workplace Outcome Suite (WOS) to measure the effectiveness of coaching interventions in helping to address a range of workplace issues. For this purpose the original 25-item version of the WOS, developed in 2009, was slightly modified to provide outcome information on the coaching process. While the changes were modest, there was some concern that they might adversely affect the psychometric characteristics of the 25-item tool compromising its validity and reliability. Prudence requires the coaching version of the WOS be evaluated to assure results are equivalent with those of the original WOS. Data from 309 clients, provided with disease management coaching services for depression and diabetes, were analyzed. Various methods, including reliability analysis, confirmatory factor analysis, and correlational studies, were included in the study. The results of these analyses showed that the coaching tool performed slightly better than the original WOS. The instrument was shown to be extremely reliable, particularly for a short scale. The analysis revealed that the 25-item coaching version of the WOS can be used to test the efficacy and effectiveness of a coaching program without concern for significant measurement error. Results suggest that psychometric studies of the original 25-item WOS can also reflect clinical change from the workplace health programs coaching version even with samples as small as 50 clients.

INTRODUCTION

A growing literature offers promising support for the efficacy of coaching-based disease management (DM) programs for many chronic diseases, including heart disease, hypertension, depression, and diabetes.[1-2] This research demonstrates that organizing EAP call centers for a case-management strategy can produce statistically positive clinical changes over time. Moreover, the use of randomized clinical trial (RCT) designs in many of these studies has provided convincing evidence that the structured intervention contributes to improvement in employees' clinical status. However, this type of research does have its critics. Lewis and Khanna assert these studies to be methodologically flawed and unable to claim that the DM approach is effective.[3]

Despite these concerns, recent corporate-based disease management programs do support coaching to help mitigate employer costs of such chronic conditions such as depression, diabetes, and heart disease.[4] While focusing on absenteeism, productivity and company-based health insurance costs, these programs use 24-hour call centers and case management behavioral support services to encourage medication compliance, treatment adherences, and other wellness activities. Studies have found that workplace disease management improves clinical status while also improving presenteeism.[5] In additon long-term outcome studies find there are benefits in the area of reduced health care utilization.[1-2]

Impact studies of coaching-based DM programs have explored specific effects within the health services systems.[1] For example, research has sought to establish the efficacy of DM programs in terms of their ability to improve specific clinical disease markers with less attention given to the

complimentary impact these programs have on workplace functioning and general health.[1] Often, this emphasis has been accompanied by the use of Randomized Controlled Trial (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. With the proliferations of coaching in EAP programs, it would be prudent to examine various coaching models within the context of broader long-term effectiveness research. Thus far, only short-term efficacy models have been designed to isolate evidence that an intervention is capable of producing change in clinical markers.

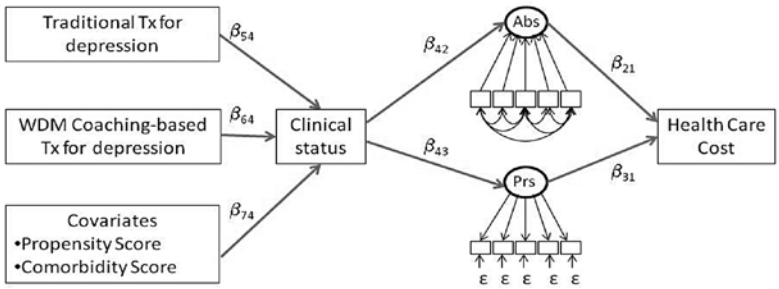
There has been a growing interest in DM programs to treat depression and other mental health problems.[2] Recent RCT studies have reported that DM programs targeting depression resulted in reduced symptoms, improved job retention, increased work productivity, and reduced absenteeism.[6-8] Although these results are very encouraging, little is known about the inter-relationships among improvements in health and improvements in workplace functioning. That is, are changes in depression symptoms associated with reductions in absenteeism and presenteeism, and are those changes associated with reductions in future health care costs?

Most existing evaluations of coaching programs were not designed to also assess comprehensive workplace outcomes. Schneider et al.,[9] conducted an evaluation of a coaching program for managing diabetes but the evaluation was limited to satisfaction with the program and clinical goal attainment. Unfortunately, it did not include a control group. Merrell et al.,[10] designed a wellness program clinical outcome study, but it did not include a coaching component, workplace oriented functional measures, or a control group. Hawkes et al.,[11] conducted a randomized clinical trial of a scripted telephone coaching program for treating cardiovascular disorders. The study included measures of quality of life from the Short Form 36-item (SF36) Health Survey and level of physical activity but did not ask about workplace functioning or life satisfaction. Linden et al.,[12] employed an interview-based coaching approach comparing participants to non-participants. They found higher levels of self-efficacy, patient activity, lifestyle change, perceived health status, and lower levels of risk over time. Another researcher tested the effectiveness of a Body Mass Index (BMI) coaching program, finding improvement in weight loss, dietary habits, and quality of life as measured with the SF36 and life satisfaction.[13] Again, none of these studies examined the impact of these coaching changes on employees' workplace outcomes. Patja et al.,[14] also tested the effectiveness of telephonic coaching for the self-care of hypertension disease in a randomized clinical trial. Results, however, failed to support the hypothesis when testing for primary clinical outcomes, quality of life, or work functioning measures.

While some studies were unable to find consistent support for coaching, the Margolis et al. study[15] of hypertension treatment found an interesting frequency response effect indicating that the more coaching sessions a patient received, the better the outcomes. These studies suggest that the DM coaching field could benefit from standardized outcome measures with which to test for the efficacy and effectiveness of various health promotional activities.

Figure 1 illustrates a model for using structural equation modeling with latent variables to examine the impact of coaching for depression and its impact on future health care costs.[16] The model shows how three exogenous variables related to coaching sessions and propensity scores capture pre-existing differences between a treatment group and a control group on clinical measures of depression and its associated effect on absenteeism and presenteeism. The various β terms in the figure illustrate the structural coefficients that can be isolated by

Figure 1: A Path Model of Coaching-Based Depression DM



the model.

For example, β_{21} and β_{31}, depict how absenteeism and presenteeism contribute to future health care costs and how mitigating the effects of depression by coaching assignments can impact health care costs through the other variables in the model. The primary objective of this type of model would be to test the effectiveness of DM on three classes of healthcare cost outcome measures:

(1) ***proximal*** clinical outcomes,
(2) ***medial*** workplace outcomes, and
(3) ***distal*** future workplace outcomes.

The Structural Equation Model (SEM) presented in Figure 1 also illustrates the systemic impact that depression and diabetes DM programs can have on the workplace. The path model shows that treatment for depression contributes to reducing depression symptoms and impacting workplace functioning in terms of absenteeism and presenteeism.[17]

The Workplace Outcome Suite was designed to evaluate the effectiveness of EAP services. It was validated from studies on the effectiveness of traditional call center and other therapeutic intervention programs but not on coaching interventions per se.[18]

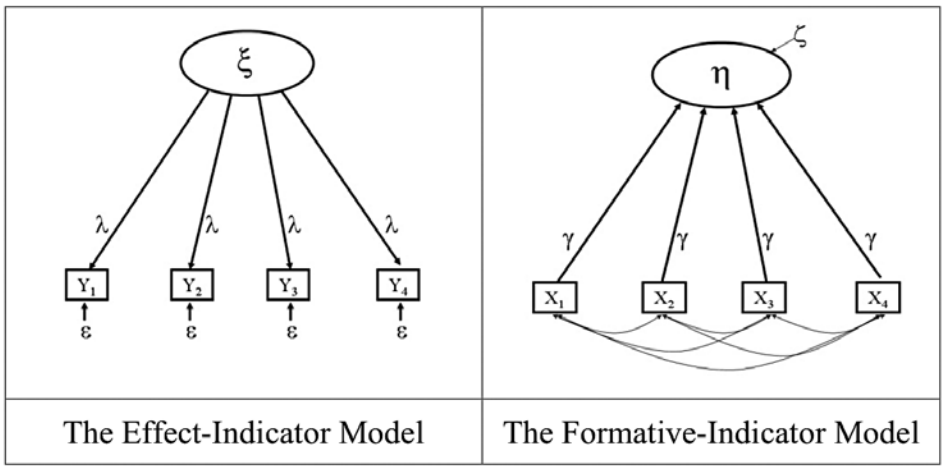
WHAT IS THE WORKPLACE OUTCOME SUITE?

The original Workplace Outcome Suite (WOS) is an easy-to-administer tool developed by Chestnut Global Partners' (CGP) Division of Commercial Science. It uses a short, precise, and easy-to-administer survey that collects EAP specific outcome data both before (pre – at start of the counseling) and after (post – usually after three months) EAP services. Thus, the WOS is a measure of change that examines five key aspects of workplace functioning: absenteeism, presenteeism, work engagement, workplace distress, and life satisfaction. Consult the 2017 WOS Report for a discussion of how it is administered, basic calculation methods for change over time, and the history of the development of the full 25-item, 9-item, and brief 5-item versions.[19]

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 25-item, 9-item, and brief 5-item versions designed around the same measurement theory and validated using the same set of psychometric principles.[20-21] Starting with the explicit measurement model that Bollen prescribed for scales, a 5-item version was constructed to facilitate obtaining results on services provided in a short amount of time.[16]

The original WOS was constructed on the basis of two formal measurement models: an effect-indicator model and a formative measurement model.[22] As shown in **Figure 2** in the effect-indicator model, the measured indicators are thought to emerge from a single underlying latent variable (illustrated by the ellipse). The small

Figure 2: Formal Measurement Models



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: $Y_i = \lambda_{i1}\xi_1 + \varepsilon_i$ [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 and provides the basis for offsetting random errors. Errors 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.

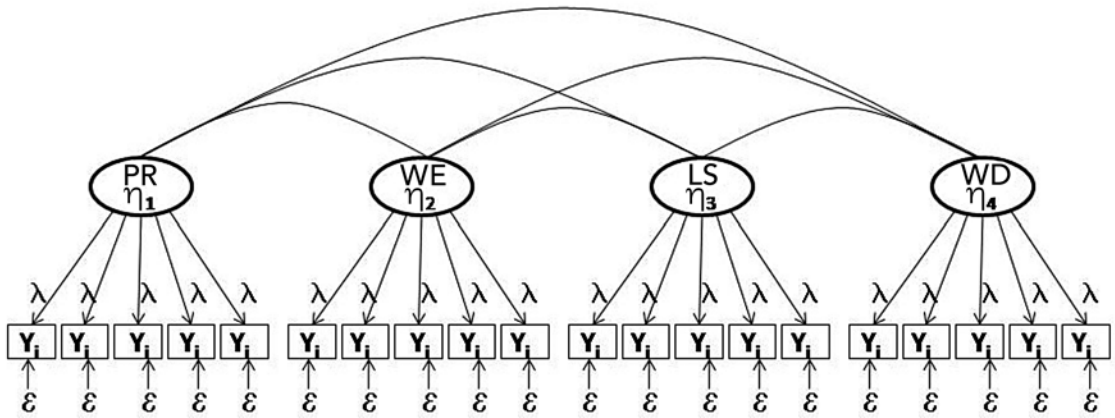
The formative measurement model uses different items that combine linearly to create a broad construct. The right side of Figure 3 illustrates the manner in which the items combine to form the latent construct. The formative model can be expressed as:

$$\eta_1 = \gamma_{11}X1 + \gamma_{21}X2 + \gamma_{31}X3 \ldots \gamma_{i1}X4 + \zeta_1 \quad [2]$$

Where η_1 is the latent construct and $\gamma_{11}X1 + \gamma_{21}X2 + \gamma_{31}X3 \ldots \gamma_{i1}X4$ are coefficients that indicate the contribution 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 considered appropriate for assessing reliability in this type of model.[16,22,23]

Results of the psychometric analysis of the WOS provide initial support for its reliability, structural validity, and construct validity. Two separate validation studies of the suite, one completed using paper-and-pencil (N = 220) and the other conducted via a telephone interview (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. 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 use of the WOS to evaluate EAP services and interventions. The coefficient alphas for the telephone interview study were similar to those of the Paper and Pencil Study. Item-total correlations for all

Figure 3: Theoretical Structure of the Effect-Indicator Measurement Models of the Presenteeism, Work Engagement, Life Satisfaction and Workplace Distress Scales



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 Work Engagement scale that suggest some room for improvement. This point will be better explained 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.

In a previous study 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 original Workplace Outcome Suite.[18] after being 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 sub-scales 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.

This article tests the validity of using the previously validated Workplace Outcome Suite (WOS) slightly modified, for use with coaching interventions.[18]

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 a coaching service as part of the disease management program for depression or diabetes. In Study II, the Well Call Center provided data from 311 clients also using a coaching approach.

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 distributions 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 alphas will be used to assess the internal consistency of the respective item-sets in the individual scale.[24] 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 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 the four latent constructs are 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 is 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 are used to fit a four-factor model to the covariance matrix, with a X^2 test of goodness-of-fit used to test the residual covariance matrix against the model matrix. It is hypothesized that 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 correlated 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 upset employee 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 1** shows the coaching version of the 25-item Workplace Outcome Suite.

Table 2 presents the means and standard deviation for the 25 items in the WOS across the board for the coaching version of the WOS.

Reliability analysis. **Table 3** presents the internal consistency for the four effect-indicator 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 item level internal consistency and the 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

Table 1: The Coaching Version of the 25-item 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 6-15. 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 have 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

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.

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 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 2: Descriptive Statistics of the CGP Outcome Suite Candidate Items

	Abbreviated Item Text	M	SD	*N*
1.	…caused you to miss 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 at work.	.29	3.44	616
5.	…required you to be on the phone, e-mail, or internet while at work.	1.36	8.84	612
6.	I had a hard time doing my work because of my personal problems.	1.52	0.96	609
7.	My personal problems kept me from concentrating on my work.	1.54	0.96	610
8.	Because of my personal problems I was not able to enjoy my work.	1.58	1.00	610
9.	My personal problems made me worry about completing my tasks.	1.52	094	607
10.	I could not do my job well because of my personal problems.	1.40	0.85	605
11.	I feel stimulated by my work.	4.04	0.97	602
12.	I often think about work on my way to the work site.	4.03	1.07	599
13.	I feel passionate about my job.	4.15	0.98	603
14.	I am often eager to get to the work site to start the day.	3.70	1.09	601
15.	I often find myself thinking about my work at home.	3.76	1.14	603
16.	My life is nearly perfect.	3.44	1.05	610
17.	I am not very satisfied with my life as a whole.	1.96	1.07	609
18.	So far, my life seems to be going very well.	4.05	0.90	612
19.	There isn't anything about my life that I would change if I could.	2.64	1.23	606
20.	I am very disappointed about the way my life has turned out.	1.60	0.93	612
21.	I often feel anxious at work.	2.66	1.25	601
22.	Thinking about being at work makes me upset.	2.00	1.18	601
23.	I am unhappy most of time at work.	1.83	1.11	600
24.	I dread going into work.	1.85	1.15	596
25.	I can't wait to get away from work.	2.28	1.25	601
V1	I have a hard time getting out of bed.	2.04	1.16	290
V2	I often feel sad.	1.76	1.03	292
V3	I keep falling behind schedule at work.	1.97	1.20	291
V4	I am rarely late for work.	4.24	1.19	293
V5	I often get home late from work.	3.41	1.29	292

Table 4 contains the standardized loading for the items on their respective latent factor. 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 and may instead be 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 fit 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 confirmatory factor analysis. These factors are estimated by using the precise weights

Table 3: Reliability Analysis of the CGP Outcome Suite Coaching Items

	Abbreviated Item Text	Item-Total *r*	α if Deleted
1.	caused you to miss 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 work.	NA[1]	NA[1]
5.	required you to be on the phone, e-mail, or internet while at work.	NA[1]	NA[1]
6.	I had a hard time doing my work because of my personal problems.	0.84	0.92
7.	My personal problems kept me from concentrating on my work.	0.84	0.92
8.	Because of my personal problems I was not able to enjoy my work.	0.80	0.93
9.	My personal problems made me worry about completing my tasks.	0.85	0.92
10.	I could not do my job well because of my personal problems.	0.86	0.92
11.	I feel stimulated by my work.	0.64	0.77
12.	I often think about work on my way to the work site.	0.60	0.78
13.	I feel passionate about my job.	0.71	0.75
14.	I am often eager to get to the work site to start the day.	0.66	0.77
15.	I often find myself thinking about my work at home.	0.46	0.83
16.	My life is nearly perfect.	0.69	0.74
17.	I am not very satisfied with my life as a whole.	0.56	0.78
18.	So far, my life seems to be going very well.	0.68	0.75
19.	There isn't anything about my life that I would change if I could.	0.51	0.81
20.	I am very disappointed about the way my life has turned out.	0.58	0.78
21.	I often feel anxious at work.	0.58	0.90
22.	Thinking about being at work makes me upset.	0.79	0.85
23.	I am unhappy most of time at work.	0.76	0.86
24.	I dread going into work.	0.82	0.85
25.	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

from 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) [**Consult Table 1**]. 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 WOS 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.

DISCUSSION

These studies support the basic psychometric properties of the Workplace Outcome Suite as a valid measurement tool of EAP outcomes in the workplace and for coaching support for disease management. The formal measurement models provide a transparent description of the manner in which the individual items impact the latent variable. The two potential problems

Table 4: Standardized Coefficients from the Four-factor Confirmatory Factor Analysis of the CGP Outcome Suite Coaching Items

	Abbreviated Item Text	Standardized Loading
1.	caused you to miss 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 work.	NA[1]
5.	required you to be on the phone, e-mail, or internet while at work.	NA[1]
6.	I had a hard time doing my work because of my personal problems.	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 work.	0.83
9.	My personal problems made me worry about completing my tasks.	0.88
10.	I could not do my job well because of my personal problems.	0.90
11.	I feel stimulated by my work.	0.81
12.	I often think about work on my way to the work site.	0.48
13.	I feel passionate about my job.	0.90
14.	I am often eager to get to the work site to start the day.	0.90
15.	I often find myself thinking about my work at home.	0.35
16.	My life is nearly perfect.	0.77
17.	I am not very satisfied with my life as a whole.	0.65
18.	So far, my life seems to be going very well.	0.77
19.	There isn't anything about my life that I would change if I could.	0.58
20.	I am very disappointed about the way my life has turned out.	0.66
21.	I often feel anxious at work.	0.60
22.	Thinking about being at work makes me upset.	0.82
23.	I am unhappy most of time at work.	0.83
24.	I dread going into work.	0.90
25.	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

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 WOS in its current form for coaching interventions. Alpha coefficient for the Presenteeism and the Workplace Distress scales are in the .90 range, showing them to be extremely reliable, especially for short scales. The mid-70s 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, suggesting that the scale has acceptable reliability for a short scale in its early development.[20] The basic psychometric characteristics of the WOS show it capable

of measuring individual difference relevant to EAP interventions. 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 WOS was to provide EAP researchers with a set of short measures useful 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 their 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. The standardized approach to the development and validation of the five constructs facilitate evaluation of the measures for specific use. Finally, Chestnut Global Partners, recently acquired by Morneau Shepell, makes the measures available for specific intervention testing at no cost to use the scientific method for evaluation, comparison, and learning about the impact of EAP services.

AUTHORS

Richard D. Lennox PhD, is founder and President of Work Matters based in New York, U.S.A. Prior to Work Matters, Dr. Lennox was Chief Scientist and Vice President of Commercial Science at Chestnut Global Partners (CGP).

David Sharar, PhD, is CEO of Chestnut Health Systems and Director of Chestnut Global Partners, (a Morneau Shepell company) Division of Commercial Science, Bloomington, IL, U.S.A., is co-creator of the Workplace Outcome Suite© (WOS) and a leading provider of employee assistance services worldwide.

Francine Miller is Vice President of Wellness Divisions & Director of Clinical Services for Health Advocate headquartered in San Francisco, CA, U.S.A. Francine was Co-Founder of WellCall, which was acquired by Health Advocate.

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 community. 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

1. 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.
2. 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.
3. Lewis, A., & Khanna, V. (2014) Surviving Workplace Wellness With Your Dignity, Finances and Major Organs Intact, *The Health Care Blog;* 1 edition
4. Petrak & Herpertz, (2009). Treatment of Depression in Diabetes: an update, *Current Opinion Psychiatry,* 22 (2) 211-217.
5. Burton, Chen, Conti; Schultz & Edington, 2006; Johns, 2008; Burton, Chen, Conti, Schultz, Pransky, & Edington, 2005. The Association between Health Risk Change and Presenteeism Change. Wayne N. Burton, Chin-Yu Chen, DanielJ. Conti, Alyssa B. Schultz, Dee W.Edington, *Journal of Occupational and Environmental Medicine.*48 (3):252-263, 2006.
6. 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.
7. 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.
8. 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.
9. 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.
10. Merrill RM, Aldana SG, Vyhlidal TP, Howe G, Anderson DR, Whitmer RW. *J Occup Environ Med.* 2011 Feb:53(2):127-31. Joj:10.1097/JOM.0b03e318209c18b.
11. 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.
12. 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.
13. Aoun, S., Osseiran-Moisson, R., Shahid, S., Howat, P., & O'Connor, M. (2011). Telephone Lifestyle Coaching: Is it Feasible as a Behavioral Change Intervention for men? *J. Health Psychol.* Retrieved from http://hpq.sagepub.com/content/early/2011/07/08/1359105311413480
14. 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.
15. 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.
16. Bollen, KA (1989) Structural Equations with Latent Varialbes. New York: Wiley.
17. Sanderson, Kristy, Tilse, Elizabeth, Nicholson, Jan, Oldenburg, Brian and Graves, Nick (2006). Which presenteeism measures are more sensitive to depression and anxiety?. *Journal of Affective Disorders* 101 (1-3) 65-74. https://doi.org/10.1016/j.jad.2006.10.024
18. Lennox, R., Sharar, D., Schmitz, E., and Goehner, D. (2010). Development and validation of the Chestnut Workplace Outcome Suite. *Journal of Workplace Behavioral Health,* 25, 107-131.
19. CGP Annual Report 2017 Link: http://chestnutglobalpartners.org/Portals/cgp/Publications/Trends-Report-April2017.pdf
20. Nunnally, JC (1978). *Psychometric Theory.* (2nd Edition) New York: McGraw-Hill.
21. Nunnally, JC, & Bernstein, IH., (1994) *Psychometric Theory,* (3rd Edition). New York: McGraw-Hill.
22. Bollen, K, A., & Lennox, R, D. (1991). Conventional Wisdom on Measurement: A Structural Equations Perspective. *Psychological Bulletin*, 110:2, 305-314.
23. Bollen, K. A. (1984). Multiple Indicators: Internal Consistency or no necessary relationship? *Quality and Quantity,* 18, 377-385.
24. Cronbach, LJ. (1951). Coefficient alpha and the internal consistency of tests. *Psychometricka,* 16, 297-334.

Is your EAP Digital?

wayForward powers your EAP to stand out helping deliver results

- Increase Engagement 2-4x
- Anytime, anywhere access to care
- Identify early & prevent behavioral health problems
- ROI - cut organizational & healthcare costs

Covering over 5 MILLION members

wayForward

- Screen & Identify
- Self-Use Programs
- Access to Experts
- Deeper Analytics

Contact **Tom Shjerven @ tom.s@wayforward.io** or **253-737-5580** to find out how you can digitally-enable your EAP to cost-effectively deliver expertise and provide on-demand support. ***It's the wayForward!***

Your EAP is only as good as its results

Approach

Personally arranged services

Meaningful engagement

Internal + external integration

Measurable results

96% of cases resolved within EAP benefit

90% participant follow-through rate

30% improvement in quality of life

98% account retention rate

Better, not cheaper

Don't stumble over a few pennies only to leave thousands of dollars in savings on the table. Purchase EAP services based on value, not just price.

800-356-0845 | www.paseap.com