IJHP

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

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,

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

Table 6: Standardized Estimates for the Confirmatory Factor Analysis

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.