Abbad, Gardênia; Borges Andrade, Jairo Eduardo; Henriques Sallorenzo, Lúcia
Self-Assessment of Training Impact at Work: Validation of a Measurement Scale
Sociedad Interamericana de Psicología
Austin, Organismo Internacional

Available in: http://www.redalyc.org/articulo.oa?id=28438213
This paper reports an experience of developing and validating a measurement scale of Training Impact at Work. This research has made possible the creation of a psychometrically valid tool applicable in studies related to the effects of training on individual performance. According to Borges-Andrade and Abbad (1996) and Abbad (1999), this sort of investigation is rare both in national and foreign literature. Few attempts have been made in connection with organizing measurements to determine the level of training effectiveness involving trained individuals. The lack of reliable, valid measurements of learning, reaction (satisfaction with training) and training impact at work may be one of the factors that have been hindering research progress in the training evaluation area.

Training impact at work is one of the main criterion variables of training evaluation models and corresponds to the third level of evaluation in more traditional approaches such as those of Kirkpatrick (1976, 1977) and Hamblin (1978). Job impact is an indirect result of training, and it is defined as the effect of training on the trainees’ performance, motivation and/or attitudes. Kirkpatrick’s (1976, 1977) and Hamblin’s (1978) approaches suggest that criterion variables such as reactions, learning, job performance (impact) and results (or organization change and final value) keep a highly positive relation among them. However, according to Alliger and Janak (1989), specialized literature on training evaluation has shown that those relations are not always significant or not always in the direction foreseen by such approaches. Research results have revealed situations in which a participant—despite showing contentment (a favorable reaction) with training and having obtained good scores in learning evaluations—would not apply his or her newly acquired skills at work. In this case, such a trainee has failed to produce a positive training transfer, not necessarily because of poor memory, poor retention capability, and poor generalization, or due to shortcomings on the training program, but because opportunities to put the things he or she has learned into practice in the work environment have been missing.

Self-Assessment of Training Impact at Work: Validation of a Measurement Scale

Gardênia Abbad
Jairo Eduardo Borges-Andrade
Lúcia Henriques Sallorenzo
Universidade de Brasília, Brasil

Abstract
This paper describes the development and validation of a measurement scale for Training Impact. Impact is defined as the training long-term effect on work performance, motivation and/or attitudes. Two weeks after training, a questionnaire with 12 impact evaluation items was applied to participants from 226 courses that had been offered by a Brazilian public organization. The 1,270 valid answers were submitted to factor analyses (PAF, direct-oblimin) and to reliability analyses (Cronbach’s Alpha). A two sub-scale structure was found (a=0.86; r=.56) accounting for 60 percent of the impact variability. A single factor structure was also found and it is similarly reliable (a=.90), accounting for 45 percent of the variability. Both structures are useful, reliable and valid.

Keywords: Labour; program evaluation; measurement; test construction.

Auto-Avaliação de Impacto do Treinamento no Trabalho: Validação de uma Escala

Resumo
Este artigo descreve o desenvolvimento e a validação de uma escala de avaliação do Impacto do Treinamento no Trabalho. Impacto é definido como o efeito do treinamento a longo prazo no desempenho, motivação e/ou atitudes. Duas semanas após o treinamento, foi aplicado um questionário com 12 itens de avaliação de impacto, em participantes de 226 cursos oferecidos por uma organização pública brasileira. As 1,270 respostas válidas foram submetidas à análise fatorial (PAF, direct-oblimin) e de confiabilidade (Alpha de Cronbach). Dois tipos de estruturas fatoriais foram encontrados: uma com duasescalas (a=0.86; r=0.56) e outra unifatorial (a=0.90) explicando, respectivamente, 60% e 45% da variabilidade de impacto. Ambas as estruturas são úteis, confiáveis e válidas.

Palavras-chave: Trabalho, avaliação de programa; medidas; construção do teste.
Researchers have given little attention to the evaluation of assumptions related to Kirkpatrick’s model (1976, 1977) and have neglected the importance of doing further studies on the interrelation of the criterion variable. Goldstein (1991) has defended the idea that reaction measurements are of little avail as learning predictors and make sense only when related to training needs. Furthermore, according to Goldstein as well as Tannenbaum and Yukl (1992), learning should be considered only as a needed - although not sufficient - condition for transfer or impact at work.

Results of recent research presented by Tannenbaum and Yukl (1992) have not confirmed the significant relation between learning and transfer (impact) measurements as well as between reaction measurements and other criteria (learning, impact, and results). Some results have shown, however, that the reaction level played a moderating role on the motivation/learning relation. In view of this, instructional program efficiency should not be determined by evaluating only one variable level, as it has commonly been done in the area.

Such findings related to the weak relationships between reaction, learning, and impact are not conclusive. Few attempts have been made to identify predicting variables of each of these evaluation levels and to build up reliable measurements of different evaluation levels have been made. In this context, the purpose of this paper is to present a reliable measurement of training impact at work which can make possible a careful and more precise evaluation of relationships among criterion variables of training effectiveness.

Method

This article describes the development and validation of a Training Impact at Work Scale. The tool has been developed in three stages – development, semantic validation, and statistical validation. The Training Impact at Work Scale described in this paper is part of the data collecting tools used by Abbad (1999) in a research effort analyzing 226 courses from March 1997 to June 1998 at a training agency in a Brazilian public organization.

Training Impact at Work has been measured by applying a scale of self-assessment at two different moments and two months after training. On both occasions, questionnaires were sent to participants to rate the impact of the training they had participated in. A Blue Format questionnaire was sent to participants two weeks after the end of the course, while a Blue Format was sent three months after the end of the course.

The aforementioned questionnaires comprised part - items related to Training Impact at Work, about Situational Supporting Factors, Methods of Training, and Consequences Associated to the Use of Newly Acquired Skills at Work. The questions included also instructions for filling in, deadlines on how they should be returned to the training agency, evaluation items were associated to a 5-point Likert scale where 1 corresponded to I totally disagree with this statement, and 5 to I totally agree with this statement. The questionnaires were submitted to semantic validation in a sample that was similar to the one of training participants.

Statistical Validation Procedures

The 12-item questionnaire on Training Impact at Work (IMTT) was applied on two different occasions of workers from the organization that received training at agency headquarters - two weeks and two months after training. Before completing the questionnaires, participants had already completed two other questionnaires, which contained differences between the Green and the Blue Format questionnaires, sample participants had already completed two other questionnaires, and 12 impact items, however, were the same.

Two thousand nine hundred and ninety IMTT questionnaires - and the same number of questionnaires were sent via internal mail to participants who had participated in training activities. Instructions and sample participants to return their fully completed questionnaires to the training agency within no later than 15 days.
More than half of the sample (55.7%) were males, 74.5% of them were middle-aged (30 to 50 years of age) and most of them (74.1%) had an university degree. Regarding functional characteristics, the sample had a balanced structure as to job position (55.9% had a job position due to an university degree; the remaining ones were high school graduates). For most of the sample (65.5%), seniority ranged from one to ten years.

Firstly, as Tabachnick and Fidell (1996) recommend, data obtained through questionnaires were submitted to descriptive and exploratory statistical analyses. Secondly, they were submitted to an analysis of the principal components, a factor analysis (PAF, oblique rotation - direct oblimin), as well as an analysis of their internal consistency (Cronbach’s Alpha). Such analyses have allowed examining the data empirical structure, validating the questionnaires’ content and evaluating their reliability.

**Results**

This section presents the results gathered from the statistical validation process of the Training Impact at Work questionnaires, for both formats.

**Training Impact at Work Scale Validation - Green Format**

The return rate of the Training Impact at Work - Green Format questionnaire was approximately of 43.66%. This corresponds to a final sample of approximately 1,270 valid answers per item. Answers were submitted to an analysis of their principal components in line with Kaiser’s method - eigenvalue > 1. As verified by KMO (.93), the data matrix could be factorized. Figure 1 shows the scree plot containing the scale principal components.

Results issued after analyzing the data reveal an empirical structure consisting of two highly correlated factors ($r=.56$), which explain 60.06% of total variance in answers. Such components include training impact on performance and training impact on attitudes. Based on the scree plot (Figure 1) and the high magnitude of bivariate correlation among components, they suggest that the scale has a predominantly unifactorial structure. For this reason, data were analyzed one more time by using the principal axis method (PAF, oblique rotation - direct oblimin), thus forcing the solution into one factor only. The scale has also been submitted to a reliability analysis (Cronbach’s Alpha). A summary of such results can be found in Table 1.

Table 1 shows the results of such analyses in terms of items, descriptive statistics, factorial loadings and the Cronbach’s Alpha value. A high level of reliability has been obtained and the items express the percentage of effects produced by training on performance and motivation at work. This factor explains 45.15% of total variance in answers.

![Figure 1. Principal components of the Training Impact at Work - Green Format Scale](image-url)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description of Items</th>
<th>Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imp6</td>
<td>The quality of the work I do has improved.</td>
<td>0.79</td>
</tr>
<tr>
<td>Imp3</td>
<td>I make fewer mistakes at work.</td>
<td>0.76</td>
</tr>
<tr>
<td>Imp5</td>
<td>I do my work faster.</td>
<td>0.75</td>
</tr>
<tr>
<td>Imp9</td>
<td>My self-confidence has increased.</td>
<td>0.75</td>
</tr>
<tr>
<td>Imp8</td>
<td>My motivation for working has improved.</td>
<td>0.68</td>
</tr>
<tr>
<td>Imp7</td>
<td>The quality of the work I do has improved in tasks not related to the course.</td>
<td>0.67</td>
</tr>
<tr>
<td>Imp10</td>
<td>I suggest more frequent changes in work routine.</td>
<td>0.66</td>
</tr>
<tr>
<td>Imp1</td>
<td>I often make use of skills learned during training.</td>
<td>0.64</td>
</tr>
<tr>
<td>Imp11</td>
<td>I feel more receptive to changes.</td>
<td>0.63</td>
</tr>
<tr>
<td>Imp2</td>
<td>I take advantage of opportunities to practice my newly acquired skills.</td>
<td>0.62</td>
</tr>
<tr>
<td>Imp12</td>
<td>My workmates can learn from me.</td>
<td>0.61</td>
</tr>
<tr>
<td>Imp4</td>
<td>I can remember well the course content.</td>
<td>0.42</td>
</tr>
</tbody>
</table>
In a second phase, the data matrix was analyzed, thus forcing the analysis solution (PAF) in two factors, as it is suggested by the scree plot, format (see Figure 1). These two components explained 60.06% of total variance in answers and were highly correlated ($r=0.67$). The first scale (where $\alpha=0.86$) refers to training impact on performance and consists of 7 items (Imp3, Imp1, Imp6, Imp2, Imp5, Imp4, and Imp12), while the second scale refers to training impact on attitudes (where $\alpha=0.86$) and consists of 5 items (Imp11, Imp9, Imp8, Imp10, and Imp7) (Table 1). The two solutions reveal that both the unidimensional structure and the bidimensional one are useful, reliable, and valid. Therefore, depending on the evaluation purposes, they can be used altogether or separately.

**Training Impact at Work Scale Validation - Blue Format**

The answers provided by approximately 710 trainees to evaluation items on the Blue Format were submitted to the same statistical validation process used with the Green Format. As verified by KMO (0.94), the data matrix could be factorized. The results of the analyses of the principal component indicate an empirical structure with two highly correlated factors ($r=0.60$), which explain 65.58% of total variance in answers. However, taking the scree plot (similar to Figure 1) as well as the high magnitude of bivariate correlation among components, a predominantly unifactorial structure can be observed. For this reason, data have been analyzed in two stages, by using the same techniques described in connection with the Green Format.

The results obtained from the analyses of the empirical structure of the Blue Format, when taken alone, almost identical to the one obtained with Green Format. Table 2 presents the Blue Format structure and descriptive statistics.

It can be noted from Tables 1 and 2 that the consistency rate of items increased and the position of some of them changed in scale when the second questionnaire was applied. This is true for items “Imp7” and “Imp10”, whose relative importance increased, and items “Imp1” and “Imp8” decreased when the second questionnaire was applied. In a second stage, the data matrix was forced in two factors, as suggested by the scree plot format. These two components explained 59.02% of total variance in answers and kept a high magnitude correlation ($r=0.71$) among items. The first scale (where $\alpha=0.90$) refers to training impact in skills, and consists of 7 items (Imp1b, Imp3b, Imp4b, and Imp12b), while the second scale refers to training impact in attitudes (where $\alpha=0.89$) and consists of 5 items (Imp11b, Imp9b, Imp8b, Imp10b, and Imp7b) (Table 2). The two solutions reveal that both the unidimensional structure and the bidimensional one are useful, reliable, and valid.

A matched-samples t test, confirming the stability of results confirm those obtained with the first questionnaire, showing that the unidimensional empirical structures composed of the *Training Impact at Work* scale have seemingly been observed.

A matched-samples t test, confirming the stability of measures for items constituting such scales revealed that after three months *Training Impact at Work* scales have seemingly been observed.

### Table 2

**Empirical Structure of the Training Impact at Work - Blue Format Factor ($\alpha=0.93$)**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description of Items</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imp6b</td>
<td>The quality of the work I do has improved</td>
<td>0.83</td>
</tr>
<tr>
<td>Imp3b</td>
<td>I make fewer mistakes at work</td>
<td>0.80</td>
</tr>
</tbody>
</table>
self-evaluations were quite similar to the ones collected two weeks after course completion. However, two significant differences between such variable averages have been found - the first being related to the item “I can remember well the course content” (Imp4, \( t=5.41, df=548, p=.005 \)); the second being related to “I take advantage of opportunities to practice” (Imp2, \( t=2.41, df=540, p=.016 \)). In both cases, averages associated with such variables were significantly lower in the second application of the same items.

**Discussion**

The evaluation questionnaire for *Training Impact at Work* presented in this paper can be used in research involving the global effects of training on trainees’ behavior. The 12-item scale is more comprehensive than the bidimensional one with two sub-scales. Both are reliable structures, and both can be applied, depending on research purposes - to investigate the global impact of training on participants’ behavior or on more specific impacts (on participants’ productive performance or attitudes).

The unidimensional scale was used in a number of nation wide projects (Borges-Andrade, Azevedo, Pereira, Rocha, & Puente, 1999; Borges-Andrade, Gama, & Oliveira-Simões, 1999; see also Britto, 1999, e.g.). It had its empirical structure preserved and its validity confirmed in different organizational environments.

The present study has not validated a questionnaire for heteroevaluations of training impact at work. Greater objectivity and reliability of results would have been ensured if similar questionnaires had been used with supervisors or peers. Although the data collecting questionnaire has obtained fine validity and reliability rates, it needs improvement.

When he evaluated the courses offered by three Brazilian organizations with headquarters in the Federal District – a financial institution, an airport administration enterprise and a private telephone company – Meneses & cols. (1999) in Banco do Brasil, 1999; see also Britto, 1999, e.g.). It had its empirical structure preserved and its validity confirmed in different organizational environments.

The aforementioned described results of factorial analysis have indicated that item “I can remember well the course content” (Imp4, \( t=5.41, df=548, p=.005 \)) had the highest factorial loading in the selfevaluation scale of training impact. Abbad (1999) has made a stepwise multiple regression analysis having trainees’ ability to remember content as a criterion variable and has found that most variables explaining retention (ability to remember content) were not correlated with impact, should not be considered as a criterion variable.

Such differences in predicting the ability to remember content and impact - should be considered in the light of conceptual and methodological issues involved in the development of the questionnaire used in this research. This issue refers to the use of the expression “I can remember content” in the questionnaire. In the statement has been made, the content is neither a necessary nor a sufficient criterion for transfer of training. Participants, for example, may not remember a content, but they may be able to transfer acquired skills into practice at work.

The 12 items included in Meneses’s (2002) heteroevaluation scale of training impact explained 50% of the variance of answers and reached a reliability index of 0.94. Abbad (1999) has made a stepwise multiple regression analysis having trainees’ ability to remember content as a criterion variable and has found that most variables explaining retention (ability to remember content) were not correlated with impact, should not be considered as a criterion variable.
Conversely, participants may remember content, but they may not know how to explain it. In such a case, not taking into account an eventual lack of adequate instructional planning, participants may be said to have learned what, but not how. Such trainees do not know how to behave the prescribed way. However, they can say what must be done.

In the context of transfer and impact, being able to retain is knowing how to do something after training is completed. Item “Imp4”, which is used to measure retention, does not suggest the use of such concept. This variable would be more intensely correlated with impact, if the corresponding item were asked about the participants’ ability to act in harmony with effectiveness standards and criteria taught during the course.

Referring again about the hypothetical example involving a computer technology course, trainees who have learned how to work out different tables, lines and fonts by using a word processor must remember steps and results of actions in order to reproduce them correctly in their work environment. Here, the expression “ability to remember content” does not seem adequate for two reasons. The first is that the questionnaire statement seems to refer only to trainees’ ability to say things about the content, not to their ability to make evident performance learned during training. The second problem involves the idea beneath the object to be remembered. Item “Imp4” seems to have made participants to focus on the idea beneath the object to be remembered. Item “Imp4” is used to measure retention, does not suggest the use of such concept. This variable would be more intensely correlated with impact, if the corresponding item were asked about the participants’ ability to act in harmony with effectiveness standards and criteria taught during the course.

Analyses have shown that retention is the effect of training measured in terms of ability to perform actions learned during the same level of competence that was acquired, without tests. Transfer of training is the application of skills in the work environment, which involves measuring the effects of training on participants’ performance and attitudes. However, they can say what must be done.

In the context of transfer and impact, being able to retain is knowing how to do something after training is completed. Item “Imp4”, which is used to measure retention, does not suggest the use of such concept. This variable would be more intensely correlated with impact, if the corresponding item were asked about the participants’ ability to act in harmony with effectiveness standards and criteria taught during the course.

Referring again about the hypothetical example involving a computer technology course, trainees who have learned how to work out different tables, lines and fonts by using a word processor must remember steps and results of actions in order to reproduce them correctly in their work environment. Here, the expression “ability to remember content” does not seem adequate for two reasons. The first is that the questionnaire statement seems to refer only to trainees’ ability to say things about the content, not to their ability to make evident performance learned during training. The second problem involves the idea beneath the object to be remembered. Item “Imp4” seems to have made participants to focus on the content of theoretical explanations.

In line with previously described conceptual analyses, the aforementioned item should check if, after training, participants still knew how to do the things taught during the course. But it did not. Apparently, the item seems to have limited the concept of retention to being able to remember things said by the instructor about course themes.

For such reasons, the concept of retention in research involving transfer of training and training impact needs to be extended. Analyses have shown that retention is the effect of training measured in terms of ability to perform actions learned during the same level of competence that was acquired, without tests. Transfer of training is the application of skills in the work environment, which involves measuring the effects of training on participants’ performance and attitudes. However, they can say what must be done.

In the context of transfer and impact, being able to retain is knowing how to do something after training is completed. Item “Imp4”, which is used to measure retention, does not suggest the use of such concept. This variable would be more intensely correlated with impact, if the corresponding item were asked about the participants’ ability to act in harmony with effectiveness standards and criteria taught during the course.

Referring again about the hypothetical example involving a computer technology course, trainees who have learned how to work out different tables, lines and fonts by using a word processor must remember steps and results of actions in order to reproduce them correctly in their work environment. Here, the expression “ability to remember content” does not seem adequate for two reasons. The first is that the questionnaire statement seems to refer only to trainees’ ability to say things about the content, not to their ability to make evident performance learned during training. The second problem involves the idea beneath the object to be remembered. Item “Imp4” seems to have made participants to focus on the content of theoretical explanations.

In line with previously described conceptual analyses, the aforementioned item should check if, after training, participants still knew how to do the things taught during the course. But it did not. Apparently, the item seems to have limited the concept of retention to being able to remember things said by the instructor about course themes.

For such reasons, the concept of retention in research involving transfer of training and training impact needs to be extended.
knowledge acquired during training articulates with the individual’s previous experience should be made in order to ensure better performance. It seems also necessary to investigate possible unfavorable impacts of new knowledges on further performances of training participants.

Bransford and Schwartz (1999) claim that traditional studies in the area of transfer of training cannot explain how a set of pre-existing abilities (the use of word processors, for example) affect people’s ability to learn a second set of similar abilities (the use of a different word processor, for example), because, traditionally, studies focus the effects of training solely on knowing how and knowing what.

In this context, longitudinal studies would also facilitate understanding the effects of training on knowing what, knowing how and knowing with. According to Bransford and Schwartz (1999), knowing with refers to the set of previous knowledge and experience accumulated by the individual from which he or she perceives, interprets and judges present and future situations. In this case, the different kinds of knowledge would not be independent. Knowing how would depend not only on the ability to interpret but also on the way the individual uses his or her previous experience (from knowing with). Also, for applying new skills in a smart, efficient way, the individual should develop metacognitive abilities to learn from new experiences and to adapt to the environment in order to be successful.

Additional studies need to be done to improve the way training impact at work is measured. A research agenda is suggested here and should include:

1. Developing and validating evaluation questionnaires of training impact at work that may contain specific items for long-term retention evaluation (based on instructional objectives, for instance).
GARDÊNIA ABBAD, JAIRO EDUARDO BORGES-ANDRADE & LÚCIA HENRIQUES SALLORENZO


