

Learning and Transfer of Training: a Quasi-Experiment with Longitudinal Design

Maria Cecília dos Santos Queiroz de Araujo¹

Gardênia da Silva Abba²

Luciana Padovez Cualbeta²

¹Agência Nacional de Vigilância Sanitária, Brasília, DF

²Universidade de Brasília, Brasília, DF

Abstract

This paper aimed to evaluate learning in three trainings held at a Brazilian federal public organization. It is a longitudinal quasi-experiment with three waves, pre-test (before training – T1), post-test 1 (right after the end of the training – T2) and post-test 2 (around three months after the training - T3). Learning was assessed with situational tests, so results are based on performance rather than self-assessment. Results show that the experimental group obtained better scores in post-test 1 than it did in the pre-test and better scores than the control group did in post-test 2. There were no difference in the results obtained by the control group, comparing pre and post-test, and that there was no difference in tests according to previous experience and demographic data. Results indicate that learning occurred as a consequence of training and was not explained by other factors of the organizational environment or individuals.

Keywords: training evaluation; learning; training transfer; experimental design; longitudinal design

Aprendizagem e Transferência de Treinamento: um Quase-Experimento com Desenho Longitudinal

Resumo

Esse artigo teve como objetivo mensurar a aprendizagem em três seções de treinamento realizadas em uma organização pública federal brasileira. Trata-se de um quase experimento longitudinal, com três ondas, sendo o pré-teste (antes do treinamento), o pós-teste 1 (imediatamente após o término do treinamento) e o pós-teste 2 (após três meses do término do treinamento). A aprendizagem foi medida usando-se testes situacionais, garantindo que os resultados são baseados em desempenho e não auto relato. Os resultados indicam que o grupo experimental obteve melhores notas no pós-teste 1 do que no pré-teste e melhores notas no pós-teste 2 quando comparado ao pós-teste 1. Não foram encontradas diferenças no grupo controle entre o pré e pós teste, nem diferenças nos resultados dos testes devidas à experiência prévia ou a dados demográficos, indicando que a aprendizagem ocorreu como consequência do treinamento e não de explicações alternativas.

Palavras-chave: avaliação de treinamento; aprendizagem; transferência de treinamento; delineamento experimental; pesquisa longitudinal

Aprendizaje y Transferencia de Entrenamiento: un Casi-Experimento con Diseño Longitudinal

Resumen

Este artículo tuvo como objetivo medir el aprendizaje en tres capacitaciones de entrenamiento realizadas en una organización pública federal brasileña. Se trata de un casi-experimento longitudinal con tres momentos: el pre-test (antes del entrenamiento), el post-test 1 (imediatamente después del término del entrenamiento) y el post-test 2 (después de tres meses del término del entrenamiento). El aprendizaje fue medido usando tests situacionales, garantizando que los resultados se basan en el desempeño y no en el autoinforme. Los resultados indican que el grupo experimental obtuvo mejores notas en el post-test 1 que en el pre-test y mejores notas en el post-test 2 comparado con el post-test 1. No se encontraron diferencias en el grupo control entre el pre y post test, ni diferencias en los resultados de los tests debido a experiencia previa o a datos demográficos, indicando que el aprendizaje ocurrió como consecuencia del entrenamiento y no de explicaciones alternativas.

Palabras clave: evaluación de entrenamiento; aprendizaje; transferencia de entrenamiento; delineamiento experimental; investigación longitudinal

Introduction

In modern society, organizations strive to remain competitive as they face economic, global, technological, and labor market challenges (Noe, Clarke & Klein, 2014). The potential for workplace learning to improve organizational performance is widely recognized (Griffin, 2011) as employees' knowledge is an important source of competitive advantage (Noe *et al.*, 2014).

This means that training and development activities play an important role in developing human capital (Noe *et al.*, 2014). Training refers to learning and development activities that aim to improve individual, team and organizational effectiveness and performance. Development refers to the acquisition of knowledge and skills for purposes of personal growth. These terms are often used interchangeable in the literature (Aguinis & Kraiger, 2009).

Despite the deep knowledge regarding Training and Development over more than one hundred years of research with findings that evidence the training benefits to individuals, teams, organizations and society, (Aguinis & Kraiger, 2009, Bell, Tannenbaum, Ford, Noe & Kraiger, 2017, Ford, Baldwin & Prasad, 2018), there are still gaps regarding how much learning is retained, generalized and transferred to work (Baldwin & Ford, 1988) and there is evidence that much of the investments made in training are wasted (Griffin, 2011). Therefore, more research is necessary regarding training evaluation, so it is possible to understand why training is not achieving its goals, which is not an easy task (Salas & Cannon-Bowers, 2001).

Many companies want to evaluate the impacts of their training and development efforts, but few actually do (Griffin, 2011), because evaluation is costly, labor intensive and difficult to conduct (Salas & Cannon-Bowers, 2001).

Training evaluation is the systematic collection of data to verify if the learning objectives were achieved and whether the achievement of these objectives helped increase performance on the job. Evaluation is done to (1) make decisions about the training (to keep it or to eliminate it), (2) to provide feedback to trainees, trainers and training designers and (3) to market training outcomes to future trainees and other organizations (Salas, Tannenbaum, Kraiger & Smith-Jentsch, 2012).

Bell *et al* (2017) argue that despite the increase in the understanding of what learning is, more empirical studies should be conducted to examine the dimensions of learning, training and performance, improving our knowledge about training effectiveness. To meet the aforementioned, this paper aims to evaluate learning in three trainings held at a Brazilian federal public organization. Learning was measured in three different moments through the application of three situational tests due to the nature and complexity of instructional objectives that reached the creation level of the Taxonomy by Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintirch, Rath and Wittrock, (2001).

Salas and Cannon-Bowers, (2001) point the need to conduct longitudinal studies in the field of training evaluation. More, Noe *et al.*, (2014) claim for longitudinal designs as learning and human capital development involve change over time. Authors like as Cromwell and Kolb (2004), Laing and Andrews (2011) and Zumrah (2014) point the need to conduct longitudinal studies

for a more in-depth knowledge of factors that influence learning and transfer of training to work. Therefore, in this investigation, participants were tested at three different moments: T1 (pre-test), T2 (post-test at the end of the training) and T3 (post-test applied three months after the training).

There is also a predominance of correlational studies in the field (Noe *et al.*, 2014) and several authors, such as Steensma and Groeneveld (2010), and Homklin, Takahashi and Techakanont (2013), point out the need to conduct experimental or quasi-experimental designs. This research proposes itself to address such recommendations, using pre and two post-tests, in addition to a control group (untrained) and an experimental group (trained), seeking to move closer to causal explanations of learning.

Learning Theories, Taxonomies of Educational Objectives and Instructional Theories

Learning refers to the processes of retention, generalization and application, at work, of knowledge, skills and attitudes (KSAs) acquired during training (Baldwin & Ford, 1988; Ford & Weissbein, 1997). Many theories have been elaborated to describe learning processes and results. The cognitivist approach adopted as reference in this study defines learning as a relatively long-lasting change in behavior, not attributable only to growth and development processes, that occurs as a result of an individual's interaction with his context. Learning occurrence has been inferred comparing a person's behavior before and after a training situation (Gagné & Medsker, 1996). This definition, adopted in this study, was chosen because it is compatible with the instructional theory (Gagné, 1985) and taxonomies of educational objectives, which enable the formulation and measurement of learning results in terms of KSAs and procedures for acquisition and transfer of KSAs to work.

Learning in work environment happens in two ways: (1) in a formal, planned and induced manner, through TD&E actions; and (2) in an informal or spontaneous manner independent of a deliberate initiative from the organization, through contact with coworkers, attempts and errors, among others (Clarke, 2004; Manuti, Pastore, Scardigno, Giancaspro & Morciano, 2015).

Learning cognitive processes includes acquisition, retention, generalization and transfer. Acquisition refers to the phase of apprehension by the individual of new pieces of knowledge, skills and/or attitudes.

Retention is storing of information in the long-term memory. Generalization refers to the application of competences apprehended in situations and conditions different from those of acquisition. Finally, transfer refers to application, at work, of knowledge, skills or attitudes learned in training situations. Transfer comprehends retention and generalization, with the latter being a necessary condition for an effective use, at work, of new types of learning (Baldwin & Ford, 1988; Ford & Weissbein, 1997).

The Taxonomy by Anderson *et al.* (2001) consists of a reformulation of the Taxonomy by Bloom *et al.* (1972) and was chosen to guide this research for being a comprehensive taxonomy involving not only the complexity of the cognitive process, but also types of knowledge. Additionally, for providing practical orientations as to the elaboration of instructional objectives, which help the construction of learning assessment items. The categories of the Taxonomy by Anderson *et al.* (2011) are (1) remembering, (2) understanding, (3) applying, (4) analyzing, (5) assessing and (6) creating, and the types of knowledge defined in such taxonomy are (a) factual, (b) conceptual, (c) procedural and (d) meta-cognitive. The authors created a double-entry matrix that allows identifying what type of knowledge needs to be taught and how complex is the cognitive process required by learning, and, based on that, establishing specific instructional objectives, defining a teaching sequence, strategies and instructional resources, as well as learning assessment criteria.

Complementarily, the Instructional Theory describes how external conditions can facilitate internal learning process and prescribes general instructional events applicable to any type of training. The instructional approach by Gagné and Medsker (1985), adopted in this study, explains how external events of instruction facilitate or enhance internal learning process towards achieving desired learning results.

Learning assessment

Most of the research on learning assessment measure leaning only in terms of declarative knowledge (Bell *et al.*, 2017) and use multiple choice tests. Tracey, Tannenbaum and Mathieu (2001) examined the influence of pretraining motivation on different levels of training reactions and knowledge acquisition, and the hierarchical relationships between levels of training outcomes in a private organization that owns about forty hotels in the United States. Learning was assessed using a posttraining test with contained eleven multiple choice items

that were developed to assess the trainees' declarative knowledge of the training program and eleven questions that aimed to assess trainees' abilities to apply the course information on the job situations. The authors have identified a positive relationship between reaction, learning and transfer, and argued that positive reactions may influence the individual's willingness to learn and that learning is fundamental for training transfer.

Tan, Hall and Boyce (2003) proposed that employees' reactions are learning predictors and distinguished between trainees' affective and cognitive reactions. Learning was defined by the authors as the principles, facts, and techniques absorbed by the trainees and that change in behavior can only be expected if learning objectives are accomplished.

Their research was conducted in a large company with 283 automotive technicians. Learning was measured with thirty-nine multiple choice questions test related to the training content. Results indicated that trainees who did not like the program also showed the higher levels of learning.

Rowold (2007) tested the links between individual variables and knowledge acquisition in call centers. At the end of the training program, trainees' declarative knowledge was assessed through a test, in the form of a questionnaire. Results showed that education, motivation to learn and expectation fulfillment were positively related to knowledge acquisition.

Iqbal, Maharvi, Malik, Khan and Road (2011) tested the relationship between training characteristics, reaction and learning. Learning was defined as an increase in knowledge and change in trainees' skills and attitudes because of the training program. Their participants self-reported on learning by responding to six items. Galanou and Priporas (2009) and Dalston and Turner (2011) used multiple item tests to measure learning and have identified that training contributes to knowledge gains. Mollahoseini and Farjad (2012) and Homklin *et al.* (2013) used self-reported Likert scales to measure learning.

Schuchter, Rutt, Satariano and Seto (2015) relied on interviews and asked about knowledge before and after training to measure trainees' learning. Ruud *et al.* (2012) used a preexperimental pretest-posttest design with paper-and-pencil tests to assess learning improvements. The results indicated that the participants gained knowledge as a result from the training program.

Steensma and Groeneveld (2010) adopted a quasi-experimental design to assess learning measures.

Method

Knowledge was measured by a multiple choice test with 33 questions. Posttest scores evidence growth in knowledge in both experimental and control groups, but post training knowledge scores were significantly higher in the experimental group than in the control group, thus suggesting that training was responsible for learning.

Therefore, the study's hypotheses are proposed:

H1. There will be no difference in the results obtained by the experimental group (trained) and the control group (untrained) in the knowledge pre-test.

H2. The experimental group (trained) will obtain better results in learning post-test 1 compared to the control group.

H3. The experimental group will have better scores in post-test 2 compared to the control group.

H4. The experimental group will have better scores in post-test 1 than in the pre-test.

H5. The experimental group will have better scores in post-test 2 than in post-test 1.

H6. There will be no difference in the results obtained by the control group in the pre-test and post-test 1.

H7. There will be no difference in the results obtained by the control group in post-test 1 and post-test 2.

H8. There will be no difference in the experimental group's tests according to previous experience and demographic data.

Galanou and Priporas (2009), Ruud, Leland, Liesinger, Johnson, Majka and Naessens (2012) and Homklin *et al.* (2013) argue that self-reported measures are limited. In this study, learning was assessed through the application of three situational tests, the first one before the training, the second one at the end of the training, and the third one three months after the training. The application at three different moments aims to assess the participants' initial repertoire, learning and long-term retention and generalization. The situational test was chosen due to the nature and complexity of instructional objectives that reached the creation level of the Taxonomy by Anderson *et al.* (2001). Bell *et al.* (2017) state that the objectives of training guide what is delivered and have implications for what can be measured as training outcomes. The authors argue that learning assessment shouldn't focus only on declarative knowledge and that other forms of evaluation should be used. Therefore, using situational tests is an innovation in the field and a contribution of the present research.

This study was conducted at a health-related Federal Regulatory Agency in Brazil from August 2014 to June 2015, and the three trainings evaluate were called Training on Health Indicators, Training on Writing Norms, and Workshop on Goals and Indicators. They were all short courses and were part of the Annual Qualification Plan for the organization's employees.

Training on Health Indicators

The training, offered to the employees of a national health surveillance agency, aimed to enable them to formulate, monitor and evaluate the health situation and health surveillance using indicators. The instructional objectives were: prepare the federal employee to recognize the inherent properties of information necessary to build a good health indicator, to understand typical technical terms, to identify, in the various information systems available, necessary variables to elaborate indicators, focusing on their expertise.

The training contained six topics: evaluation and monitoring in health, focusing on health surveillance, introduction on indicators, establishment and use of indicators, information systems, analysis and interpretation of indicators, health surveillance indicators. The target audience was the federal health surveillance agency's employees. It was a classroom course and lasted 20 hours.

Training on Writing Norms

The purpose of this training was to train the national health surveillance agency employees to write normative texts, using writing norms, ensuring that the text is clear, coherent and well-founded. The instructional objectives were: to differentiate normative and argumentative texts, to distinguish what kind of normative act should be edited, considering the content of the norm and its objectives, to formulate the first paragraph of a norm. The target audience was the federal health surveillance agency's employees. It was a classroom course and lasted 20 hours.

Workshop on Goals and Indicators

The purpose of the workshop was to enable the federal employees to discuss the concepts of performance management in the public sector, performance evaluation, indicators, and goals and to elaborate an individual proposal of goals and performance indicators. It was called a workshop due to its practical design.

The target audience was employees and managers who had recently mapped competencies in the agency, and employees that work in Planning and are directly responsible for elaborating and monitoring the agency's goals and indicators. Table 1 displays an overview of the study.

The study's objective was to assess learning during three training waves (T1, T2 e T3) held at a Federal Health Agency through a quasi-experiment, with experimental group (trained) and control group (untrained) chosen by convenience, and application of pre-test (before training – T1), post-test 1 (right after the end of the training – T2) and post-test 2 (around three months after the training - T3). Eight hypotheses were defined for the study, which have been presented before.

Participants

A total of 150 people joined the experimental group, and 80 people were in the control group, totaling 230 participants. In the experimental group, 77,2% were female, age average was 35,02 (SD = 7,34) and the average time of work in the organization was 64,79 months (SD = 53,67). In the control group, 63,75% were female, age average was 34,8 (SD = 6,63) and the average time of work in the organization was 60,10 months (SD = 51,67).

In both groups the predominant schooling was specialization (33,58% of participants in the experimental group and 43,75% in the control group). The predominant degree was in Pharmacy (29,33% in the experimental group and 43,75% in control group). Previous experience with the training theme was present in 71% of the participants from the experimental group, and in 28% from the control group.

In accordance with recommendations by Goodwin and Goodwin (2013), the power of the test was calculated by means of GPower 3.1.9.2 software. The power of the test for pre-test was 0.59, 0.45 for post-test 1, and 0.16 for post-test 2 (effect size 0.3 and $p = 0.05$).

Instruments

Instruments for data collection consisted of three equivalent situational judgment tests. Situational Judgment Tests (SJT) are a low-fidelity measurement tool commonly used as a selection tool in human resources (Fritzsche, Stagi, Salas & Kurke, 2006), that capture job-related competences and skills (Lievens, Peeters & Schollaert, 2008) as they present to respondents typical work-related situations and ask them to respond what they should or would do in each of them (Whetzel & McDaniel, 2009). Therefore, they can be better performance predictors than other methods, like self-report measures (Whetzel & McDaniel, 2006; (Motowidlo, Hooper & Jackson, 2006).

SJTs were adopted because they measure trainees' performance in tasks that are related to behavior expected by the organization because of training. It is explained that the tests used are not only knowledge tests but also include measurement of skills, since participants had to put into practice technical skills acquired during courses to answer the test questions, which refer to the last complexity level of the Taxonomy by Anderson *et al.* (2001), which is creation and comprehends learning of procedures towards the solution of problems at work.

SJTs were built from the instructional objectives of each training, with one being for application before

Table 1.

Method

Objective	Participants	Instruments	Moment of conduction	Collection procedures	Analysis procedures
Learning assessment	Experimental Group and Control Group	Situational test 1-T1	Before the start of the training	Pre-test application	Means, Standard Deviations, Minimums and Maximums,
	Experimental Group and Control Group	Situational test 2 -T2	At the end of the training	Post-test 1 application	Mann-Whitney, Kruskal Wallis, Friedman's
	Experimental Group and Control Group	Situational test 3 – T3	3 months after the end of the training	Post-test 2 application	Anova, Ancova with bootstrap

the training (pre-test) to assess the participants' initial repertoire as to course contents; another for application at the end of the training (post-test 2) to assess long-term retention and generalization. The third test corresponds to the assessment of a measure of transfer of training. Tests were considered equivalent in terms of content and complexity, because they've evaluated the same instructional goals and had the same number of questions (up to 4 questions for each wave).

The tests were prepared by the instructors of each training and the researcher, based on the instructional objectives of the training and on the Taxonomy by Anderson *et al.* (2001). The instructors used the Taxonomy Table (Anderson et al., 2001) to elaborate the course's instructional objectives. Subsequently, they inserted each objective into the cell corresponding to the intersection between the type of knowledge and degree of complexity of the cognitive process. The researcher analyzed the submitted material and suggested improvements, which were accepted by the instructors. Items were elaborated to cover a representative sample of the contents referred to in the instructional objectives of the three courses. Figure 1 exemplifies the work carried out.

It is worth noting that the first objective set for the training on Health Indicators is not at all basic, as it requires a cognitive process of analysis, the fourth level of the complexity gradation of learning cognitive processes. It is also possible to observe that, though there are two objectives at the same complexity level, each one of them refers to one type of knowledge: conceptual and procedural. Finally, the third objective relates to procedure as well but requires more than analysis, it requires creation. The individual, then, based on the acquired knowledge, should be capable of creating something new.

Each test was made up by an average of four questions. Each question was a situation based on the work reality, that should be solved by the participant and

included the content of at least one instructional objective. The tests were applied by means of an electronic form. The solution to each situation were answered by the participants using opened answers that were evaluated by the instructors, based on criteria previously established by the instructors based on the standards and technical guidelines of each course.

In the training on Health Indicators, for example, for the instructional objective "to elaborate health indicators, focusing in their occupation area", one of the situational tests required that the participants elaborated a health indicator, its concept, interpretation, possible uses, limitations and calculation method, aiming to monitor and protect populations' health.

Procedures

The sample of participants of the experimental groups was defined according to the availability of access to the participants of the two trainings evaluated. Control group was chosen randomly from a list of employees that were in the same job position as the experimental group but did not participate of the evaluated courses.

Participants of the experimental group received an invitation to participate of the research right after confirming their participation in the training. Control group participants received an invitation in the days that preceded the training.

One of the authors participated of the first class of each training, to clarify participants about the research goals, ethical procedures and explain that their participation was voluntary.

SJTs were inserted in an electronic form and the links to access the form were sent by email. Pre-tests were applied in the week before training, post-test 1 was applied immediately after training and post-test 2 was applied three months after the course ended. The response time of each test 30 minutes on average.

Types of knowledge	Degrees of complexity of the cognitive process	
	4. Analyze	5. Create
Conceptual	Recognize properties necessary to the construction of a good health indicator.	Elaborate indicators with a focus on their area of activity.
Procedural	Identify in the several information systems available variables necessary to the elaboration of indicators.	

Figure 1. Instructional objectives of the training on Health Indicators

Data analysis

Data analysis procedures consisted of means, standard deviations, minimums and maximums, Mann-Whitney and Kruskal Wallis tests, Friedman's ANOVA, and ANCOVA with bootstrap, after verification of compliance with statistical assumptions (Field, 2013).

Results

In a comparison between groups, H1 assumed that there would not be difference between groups before the training, and H2 and H3 predicted that the experimental group would have higher means compared to the control group in post-test 1 and post-test 2, respectively. No significant difference was found between the control groups and the experimental group's pre-test scores, $U = 3.43$, $z = -0.641$, $p = 0.522$ ($p > 0.05$), $r = -0.05$. On the other hand, comparing means between both groups in post-test 1, the difference found in the statistics proved significant, $U = 2.156$, $z = 3.611$, $p = 0.000$ ($p < 0.05$), $r = 0.37$, showing that those who participated in the training achieved better scores than those who did not participate in the training. Comparing means between the two groups in post-test 2, difference was not significant, $U = 77.5$, $z = 1.362$, $p = 0.173$ ($p > 0.05$), $r = 0.29$; however, the experimental group's mean in post-test 2 is superior to the control group's mean in the same test. Therefore, H1 and H2

were corroborated, whereas H3 was not. Table 2 summarizes the results.

As for intra-group comparison, H4 predicted that the experimental group would have better scores in post-test 1 than in the pre-test, and H5 established that the experimental group would have better scores in post-test 2 than in post-test 1. Results showed that this difference between the experimental group's scores in the pre-test and post-test 1 was significant, $\chi^2(1) = 22.26$, $p = 0.000$ ($p < 0.05$), $r = 0.55$, corroborating H4. On the other hand, the difference between post-test 1 mean score and post-test 2 score was not significant, $\chi^2(1) = 1.80$, $p = 0.180$, ($p > 0.05$), $r = 0.48$, refuting H5.

Still about intra-group comparison, H6 set that there would not be difference between the pre-test and post-test 1 for the control group, and H7 predicted that there would not be difference between the control group's post-test 1 and post-test 2. Comparing differences between the control group's pre and post-test 1, it was possible to observe that there was no significant difference $\chi^2(1) = 1.29$, $p = 0.257$ ($p > 0.05$), $r = 0.12$, in the same way that the difference between post-test 1 and post-test 2 was not significant $\chi^2(1) = 1.60$, $p = 0.206$ ($p > 0.05$), $r = 0.62$, corroborating H6 and H7, respectively. Table 3 synthesizes results of intra-group comparisons.

In an explanatory model, there are normally other variables that influence the dependent variable. Because

Table 2.

Comparison between control group and experimental group

Test	Control Group	Experimental Group	Mann-Whitney U	Sig*
Pre-test (0-100)	63.32	63.22	3.43	0.522
Post-test 1 (0-100)	62.17	80.21	2.156	0.001
Post-test 2 (0-100)	73.72	88.43	77.5	0.173

*Exact significance, 1-tailed.

Table 3.

Intra-group comparison

Test	Friedman's ANOVA Pre x Post 1	Sig*	Friedman's ANOVA Post 1 x Post 2	Sig*
Experimental Group	22.26	0.000	1.80	0.180
Control Group	1.29	0.257	1.60	0.206

* Asymptotic significance

in the present study there was no random choice of participants, it was necessary to investigate the influence of sociodemographic variables. Thus, H8 established that the scores of the experimental group's test would not vary according to previous experience and demographic data. No significant difference was found as to the scores of participants that had previous experience compared to those that did not, in none of the tests (pre-test, $U = 943.50$, $z = -1.52$, $p = 0.128$ ($p > 0.05$), $r = -0.16$; post-test 1, $U = 488.50$, $z = -1.73$, $p = 0.08$ ($p > 0.05$), $r = -0.20$; post-test 2, $U = 16.00$, $z = 0.081$, $p = 0.190$ ($p > 0.05$), $r = 0.027$).

When it comes to the influence of gender on test results, there was no significant difference between scores in men's and women's tests (pre-test: $U = 937.500$, $z = 1.708$, $p = 0.088$ ($p > 0.05$), $r = 0.17$; post-test 1: $U = 420.500$, $z = 1.086$, $p = 0.278$ ($p > 0.05$), $r = 0.13$; post-test 2: $U = 6.00$, $z = 0.920$, $p = 0.357$ ($p > 0.05$), $r = 0.31$).

Analyzing the influence of educational attainment, results indicated that differences were not significant (pre-test: $H(6) = 6.58$, $p = 0.362$ ($p > 0.05$), $r = 0.04$; post-test 1: $H(6) = 13.26$, $p > 0.05$), $r = 0.2$; post-test 2: $H(3) = 7.62$, $p = 0.054$ ($p > 0.05$), $r = 0.67$). In the analysis by job position, differences in the pre-test and

post-test 1 were significant, $H(4) = 14.86$, $p = 0.005$ ($p < 0.05$), $r = 0.30$ and $H(3) = 14.34$, $p = 0.002$ ($p < 0.05$), $r = 0.47$, respectively. Job position had no influence on post-test 2 scores, $H(2) = 1.18$, $p = 0.553$ ($p > 0.05$), $r = 0.66$.

Analyzing the influence of the age co-variable on the experimental group's post-test 1 scores, the age co-variable was not significantly related to the post-test 1 score, $F(1.69) = 2.96$, $p = 0.090$ ($p > 0.05$), $\eta^2 = 0.04$. Time working in the sector was not significantly related to post-test 1 results, $F(1.69) = 0.34$, $p = 0.564$ ($p > 0.05$), $r = 0.07$. Face the results, H8 was partially corroborated because job position influenced pre-test and post-test 1 scores. Table 4 displays a synthesis of the study's hypothesis tests.

Discussion

The results of this study showed that there was no difference between the experimental group and the control group before training; that the experimental group obtained better scores in post-test 1 than it did in the pre-test and that the experimental group obtained better scores than the control group did in post-test 2. On the other hand, the difference between post-test 1

Table 4.
Synthesis of Study 1's hypothesis test

	Hypothesis	Result
H1	There will be no difference in the results obtained by the experimental group (trained) and control group (untrained) in the pre-test.	Corroborated
H2	The experimental group will obtain better results in post-test 1 compared to the control group.	Corroborated
H3	The experimental group will obtain better scores in post-test 2 compared to the control group.	Not corroborated
H4	The experimental group will have better scores in post-test 1 than in the pre-test.	Corroborated
H5	The experimental group will have better scores in post-test 2 than in post-test 1.	Not corroborated
H6	There will be no difference in the results obtained by the control group in the pre-test and post-test 1.	Corroborated
H7	There will be no difference in the results obtained by the control group in post-test 1 and post-test 2.	Corroborated
H8	There will be no difference in the experimental group's tests according to previous experience and demographic data.	Partially corroborated

mean score and post-test 2 score for the experimental group was not statistically significant.

The study indicated, additionally, that there was no difference as to results obtained by the control group, comparing pre and post-test, and that there was no difference in tests according to previous experience and demographic data, except for job position, which influenced pre and post-test 1 scores. Facing these results, it is possible to analyze that learning occurred as a consequence of training; therefore, training was effective, and learning was not explained by other factors of the organizational environment or individuals. Informal learning at work, possible alternative explanation to results, seems to not have had enough influence to reduce the effect of training on the performance of former trainees in skill tests.

According to Shadish, Cook and Campbell (2002), a quasi-experiment is a research design with no random distribution of participants, but in which other conditions for the conduction of an experiment are met (use of pre and post-test and control group). According to the authors, a quasi-experiment creates a reasonable approximation to the causes of a phenomenon. The present research, by means of a quasi-experimental design, was capable of identifying the occurrence of learning, as preconized by the authors. In Galanou and Priporas (2009) and Zumrah (2014), training also influenced learning in a positive way.

Observing that the difference between the experimental group's mean score in post-test 1 and the score in post-test 2 was not statistically significant, it can be suspected that former trainees are not transferring to work competences learned during training and, for this reason, are not applying what has been learned. On the other hand, although they were not statically different, scores in the post-test 2 for the experimental group were a lot higher than the scores in post-test 2 for the control group. A possible explanation is that the sample was not big enough not to reject the hypothesis.

The influence of previous experience and demographic characteristics in the tests was analyzed. Previous experience was not significant. Therefore, there was no difference between those who had and those who did not have experience with the training theme in the test results, reinforcing the idea that training was responsible for learning.

About demographic data, only job position influenced test scores, but no results were found in the literature for comparison. It is suspected that the job position that obtained the highest scores is one more

related to the themes of the training, which would explain the higher scores.

Analyzing the influence of educational level, the latter did not prove significant. This result opposes previous studies. Ruud *et al.* (2012) found positive correlation between education level and scores obtained in post-test for each test of each session and in the overall result. In Rowold (2007), education predicted learning significantly.

Other demographic data did not influence test results, reinforcing that learning happened because of training. Homklin *et al.* (2013) confirm the influence of an individual's characteristics on training improve formation effectiveness, which was not seen in the present research.

Bearing in mind that instructional objectives were built from observable behaviors and that the Taxonomy by Anderson *et al.* (2001) was used to analyze the type of knowledge to be taught and the complexity of the cognitive process, it is judged that the learning assessments were adequate and managed to obtain results that meet reality.

Moreover, the performance of learning assessment through a quasi-experiment with longitudinal design and control group has contributed to fulfilling the research agenda in the TD&E field (Cromwell and Kolb, 2004; Steensma and Groeneveld, 2010; Laing and Andrews, 2011; Homklin *et al.* 2013 and Zumrah, 2014).

Final considerations

The paper aimed to investigate if learning is transferred to work and if transfer was caused by training or alternative explanations. Therefore, it contributes to knowledge in Training and Development, as it addressed research gaps pointed in literature reviews and meta-analysis (Aguinis e Kraiger, 2009, Bell *et al* 2017; Ford, Baldwin & Prasad, 2018).

A quasi-experiment was conducted, which are always recommended in research agenda of scientific studies (Steensma and Groeneveld, 2010; Homklin *et al.*, 2013) but rarely put into practice. Quasi-experiments provide greater security in relation to results found, due to a greater control, which involves comparison group and application of tests at different moments.

Another contribution refers to the conduction of a longitudinal study, pointed out as necessary by several authors (Cromwell and Kolb, 2004; Laing and Andrews, 2011; Zumrah, 2014). Measurement of what has been learned at different moments allows assessing not only

learning but retention and generalization as well based on tests applied at three different moments.

Using situational test, based on performance rather than self-assessment, was relevant, considering that information then does not depend on perceptual measurement through self-report but on right or wrong answers that can be corrected by an expert.

A limitation in this research is the difficulty in composing groups and obtaining answers from participants to the questionnaires. In addition, the size of the sample was not big enough to evidence the validity of scales used. The arguments used by the employees were excessive work and busy schedules. In addition, it was not possible to make a random choice of subjects for participation in control and experimental groups, which caused initial differences between groups, which had to be statistically extracted as co-variants to isolate the effect of trainings on post-test results. These limitations did not allow for a representative number for pairing between results obtained by the participants in the tests, which also did not allow for bigger samples, more sensitive to training effects. However, it was noticed that people might have not participated in the research due to the request for identification in situational tests. Further studies should adopt the method presented in Ruud *et al.* (2012), according to which each participant creates a number of identifications for tests so only he knows who is answering, but in a way that is possible to register the different questionnaires answered by each participant.

Another limitation is the fact that the experimental groups have not been separated by training. The objective was only to test whether the trainings explained learning results. For this reason, it is not possible to know if a course was better or worse than another as to production of results.

A suggestion for research agenda is expanding the sample would be an adequate action, with application of situational tests at other organizations, in order to enable more solid conclusions on found results. Further quasi-experimental studies or even experimental studies need to be conducted as to identify what variables actually lead to learning during trainings. By doing so, it will be possible to promote enhancements in T&D actions.

References

Aguinis, H. & Kraiger, K. (2009). "Benefits of training and development for individuals and teams, organizations, and society", *Annual review of psychology*,

60, 451–474. doi: <https://doi.org/10.1146/annurev.psych.60.110707.163505>

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintirch, P. R., Rath, J. & Wittrock, M. C. (2001). *A Taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Baldwin, T.T. & Ford, J.K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41(1), 63-105. doi: <https://doi.org/10.1111/j.1744-6570.1988.tb00632.x>

Bell, B. S., Tannenbaum, S. I., Ford, J. K., Noe, R. A. & Kraiger, K. (2017) "100 years of training and development research: What we know and where we should go.", *Journal of Applied Psychology*, 101(3), 305–323. doi: <https://doi.org/10.1037/apl0000142>

Clarke, N. (2004) "HRD and the challenges of assessing learning in the workplace", *International Journal of Training and Development*, 8(2), 140–156. doi: <https://doi.org/10.1111/j.1468-2419.2004.00203.x>

Cromwell, S. E. & Kolb, J. A. (2004). "An examination of work-environment support factors affecting transfer of supervisory skills training to the workplace". *Human Resource Development Quarterly*, 15(4), 449-471. doi: <https://doi.org/10.1002/hrdq.1115>

Dalston, T. R. & Turner, P. M. (2011) "An Evaluation of Four Levels of Online Training in Public Libraries", *Public Library Quarterly*, 30(1), 12–33. doi: <https://doi.org/10.1080/01616846.2011.551041>

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. London: Sage.

Ford, J. K., Baldwin, T. P., & Prasad, J. (2018). Transfer of Training: The Known and the Unknown. *Annu. Rev. Organ. Psychol. Organ. Behav.*, 5, 5.1-5.25. doi: <https://doi.org/10.1146/annurev-orgpsych032117-104443>.

Ford, J.K. & Weissbein, D.A. (1997). "Transfer of training: An updated review and analysis", *Performance Improvement Quarterly*, 10(2), 22-41. doi: <https://doi.org/10.1111/j.1937-8327.1997.tb00047.x>

Fritzsche, B. A., Stagi, K. C., Salas, E., & Kurke, C. S. (2006). Enhancing the Design, Delivery and Evaluation of Scenario-Based Training: Can Situational Judgment Tests Contribute? In J. A. Weekley & R.

- E. Ployhart (Eds.), *Situational judgment tests* Mahwah, NJ: Lawrence Erlbaum Associates.
- Gagné, R. M. (1985). *The conditions of learning*. (4 ed.). New York: Holt, Rinehart & Winston.
- Griffin, R. (2011) “Seeing the wood for the trees: workplace learning evaluation”, *Journal of European Industrial Training*, 35(8), 841–850. doi: <https://doi.org/10.1108/03090591111168357>
- Gagné, R. M. & Medsker, K. L. (1996). *The conditions of learning – training applications*. Belmont: Wadsworth Group / Thomson Learning.
- Galanou, E. & Priporas, V. (2009). “A model for evaluating the effectiveness of middle managers’ training courses: evidence from a major banking organization in Greece”, *International Journal of Training and Development*, 13 (4), 221-246. doi: <https://doi.org/10.1111/j.1468-2419.2009.00329.x>
- Homklin, T., Takahashi, Y. & Techakanont, K. (2013). “Effects of individual and work environment characteristics on training effectiveness: Evidence from skill certification system for automotive industry in Thailand”, *International Business Research*, 6(12), 1-17. doi: <https://doi.org/10.5539/ibr.v6n12p1>
- Iqbal, M. Z., Maharvi, M. W., Malik, S. A., Khan, M. M. & Road, P. (2011). “An empirical analysis of the relationship between characteristics and formative evaluation of training”, *International Business Research*, 4(1), 273-286. <https://pdfs.semanticscholar.org/9111/622dd7cbdf70aad759e36e6ccb545f4377c4.pdf>
- Laing, G. K. & Andrews, P. (2011). “An empirical test of training outcomes in the Australian hotel industry.”, *International Journal of Business and Social Science*, 2 (21), 81-87. http://ijbssnet.com/journals/Vol_2_No_21_Special_Issue_November_2011/10.pdf
- Lievens, S., & de Soete, B. (2015). Situational Judgment Test. *International Encyclopedia of the Social & Behavioral Sciences*, 22, 13-18.
- Manuti, A., Pastore, S., Scardigno, A. F., Giancaspro, M. L. & Morciano, D. (2015) “Formal and informal learning in the workplace: A research review”, *International Journal of Training and Development*, 19(1), 1–17. doi: <https://doi.org/10.1111/ijtd.12044>
- Mollahoseini, A. & Farjad, S. (2012) “Assessment Effectiveness on the Job Training in Higher Education (Case Study: Takestan University)”, *Procedia - Social and Behavioral Sciences*, 47, 1310–1314. doi: <https://doi.org/10.1016/j.sbspro.2012.06.817>
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). A Theoretical Basis for Situational Judgment Tests. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* Mahwah, NJ: Lawrence Erlbaum Associates.
- Noe, R. A., Clarke, A. D. M. & Klein, H. J. (2014) “Learning in the Twenty-First-Century Workplace”, *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 245–275. doi: <https://doi.org/10.1146/annurev-orgpsych-031413-091321>
- Rowold, J. (2007). “Individual influences on knowledge acquisition in a call center training context in Germany”, *International Journal of Training and Development*, 11(1), 21-34. doi: <https://doi.org/10.1111/j.1468-2419.2007.00267.x>
- Ruud, K. L., Leland, J. R., Liesinger, J. T., Johnson, M.G., Majka, A. J. & Naessens, J. M.. (2012). “Effectiveness of a quality improvement training course: Mayo Clinic quality academy”, *American Journal of Medical Quality*, 27, 130-138. doi: <https://doi.org/10.1177/1062860611415391>
- Salas, E. & Cannon-bowers, J. A. (2001) “The science of training: a decade of progress”, *Annual review of psychology*, 52, 471-499. doi: <https://doi.org/10.1146/annurev.psych.52.1.471>
- Salas, E. Tannebaum, S. I. Kraiger, K. & Smith-Jentsch, K. A. (2012) “The science of training and development in organizations: what matters in practice”, *Psychological Science in the Public Interest*, 13(2), 74-101. doi: <https://doi.org/10.2307/23484697>
- Schuchter, J., Rutt, C., Satariano, W. A. & Seto, E. (2015) “Building capacity for Health Impact Assessment: Training outcomes from the United States”, *Environmental Impact Assessment Review*, 50, 190–195. doi: <https://doi.org/10.1016/j.eiar.2014.10.002>
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalizes causal inference*. Boston: Houghton Mifflin Company.
- Steenasma, H. & Groeneveld, K. (2010). “Evaluating a training using the four levels model”, *Journal of Workplace Learning*, 22(5), 319-331. doi: <https://doi.org/10.1108/13665621011053226>

- Tan, J. a., Hall, R. J. and Boyce, C. (2003) “The role of employee reactions in predicting training effectiveness”, *Human Resource Development Quarterly*, 14(14), 397–411. doi: <https://doi.org/10.1002/hrdq.1076>
- Tracey, J. B., Tannenbaum, S. I. and Mathieu, J. E. (2001) “The influence of individual characteristics and the work environment on varying levels of training outcomes”, *Human Resource Development Quarterly*, 12(1), 5–23. doi: [https://doi.org/10.1002/1532-1096\(200101/02\)12:1<5::AID-HRDQ2>3.0.CO;2-J](https://doi.org/10.1002/1532-1096(200101/02)12:1<5::AID-HRDQ2>3.0.CO;2-J)
- Whetzel, D. L., & McDaniel, M. A. (2009). “Situational judgment tests: An overview of current research”. *Human Resource Management Review*, 19, 188-202. doi: <https://doi.org/10.1016/j.hrmr.2009.03.007>
- Zumrah, A. R. (2014). “Service quality in Malaysian public sector: The role of transfer of training”, *Procedia - Social and Behavioral Sciences*, 144, 111 – 117. doi: <https://doi.org/10.1016/j.sbspro.2014.07.279>

Recebido em: 22/01/2018

Reformulado em: 19/09/2018

Aprovado em: 28/09/2018

Sobre as autoras:

Maria Cecília dos Santos Queiroz de Araujo, Masters in Social, Work, and Organizational Psychology from the Universidade de Brasília (UnB). Lato Sensu Post graduation course in Sanitary Surveillance and Work Psychodynamics. Psychologist with a Bachelor’s degree in Psychology. Holds the position of Administrative Analyst at the National Health Surveillance Agency (Anvisa), since 2007, working in the areas of human resources management and planning. ORCID: orcid.org/0000-0002-3845-8354
E-mail: ceciliasqs@gmail.com

Gardênia da Silva Abbad holds a PhD in Psychology from the Universidade de Brasília (UnB). She is an associate professor at the Graduate School of Psychology, Institute of Psychology, at Universidade de Brasília (UnB). Her research interests are focused on work and organizational psychology, learning in organizations, training and development and training, and educational program evaluation. ORCID: [Orcid.org/0000-0003-0807-3549](https://orcid.org/0000-0003-0807-3549)
E-mail: gardenia.abbad@gmail.com

Luciana Padovez Cualheta, Doctoral student in Business Administration at the Universidade de Brasília (UnB). Masters in Business Administration from the Universidade Federal de Goiás (UFG). Lato Sensu Graduate course in Marketing from Fundação Getúlio Vargas (FGV). Consultant and facilitator of Entrepreneurship programs. ORCID: orcid.org/0000-0003-3893-6181
E-mail: lucianapadovez@outlook.com

Contato com as autoras:

Maria Cecília dos Santos Queiroz de Araujo
SQN 215 Bloco K Apartamento 106 - Asa Norte
Brasília-DF, Brasil
CEP: 70874-110