

Analysis of the learning evaluation process of nursing staff factions

ANÁLISE DO PROCESSO DE AVALIAÇÃO DA APRENDIZAGEM DE AÇÕES EDUCATIVAS DE PROFISSIONAIS DE ENFERMAGEM

ANÁLISIS DEL PROCESO DE EVALUACIÓN DE APRENDIZAJE DE ACCIONES EDUCATIVAS DE PROFESIONALES DE ENFERMERÍA

Vera Lucia Mira¹, Marina Peduzzi², Marta Maria Melleiro³, Daisy Maria Rizatto Tronchin⁴, Maria de Fátima Fernandes Prado⁵, Patrícia Tavares dos Santos⁶, Enilda Maria de Sousa Lara⁷, Jaqueline Alcântara Marcelino da Silva⁸, Jairo Eduardo Borges-Andrade⁹

ABSTRACT

The objective of this study was to analyze the learning evaluation process of nursing staff training programs, regarding its effectiveness and instrument validity. This is an empirical study whose findings support theoretical formulations for the construction of an evaluation methodology. The analysis consisted of learning evaluations related to six training programs, totaling 993 evaluations. The results of these six training programs showed the need to propose an evaluation methodology, defining criteria, instruments and indicators for this purpose. Furthermore, it is important to improve the diagnosis process on training needs and teaching strategies.

RESUMO

O objetivo deste estudo foi analisar o processo de avaliação da aprendizagem realizado em treinamentos ministrados à equipe de enfermagem, no que se refere à eficácia dos treinamentos e à validade dos instrumentos. Trata-se de um estudo empírico, cujos achados apoiarão formulações teóricas para a construção de uma metodologia de avaliação. A análise consistiu em avaliações da aprendizagem em seis treinamentos, totalizando 993 avaliações. Com base nos resultados dos seis treinamentos avaliados, pôde-se perceber a necessidade de proposição de uma metodologia de avaliação, construindo critérios, instrumentos e indicadores para essa finalidade. Notou-se também que é preciso aprimorar a realização do diagnóstico relativo à necessidade de treinamento e às estratégias de ensino.

RESUMEN

El objetivo de este estudio fue analizar el proceso de evaluación de aprendizaje realizado en entrenamientos brindados al equipo de enfermería, en lo que se refiere a la eficiencia de los mismos y a la validez de los instrumentos. Se trata de un estudio empírico, cuyas conclusiones apoyarán formulaciones teóricas para la construcción de una metodología de evaluación. Se constituyeron en material de análisis las evaluaciones de aprendizaje relativas a seis entrenamientos, sobre un total de 993 evaluaciones. De los resultados de los seis entrenamientos evaluados fue posible percibir la necesidad de la proposición de una metodología de evaluación, construyendo criterios, instrumentos e indicadores para dicha finalidad. Se notó también que es preciso mejorar la realización del diagnóstico de necesidad de entrenamiento y las estrategias de enseñanza.

DESCRIPTORS

Nursing, team
Training
Educational measurement
Personnel administration, hospital

DESCRIPTORES

Equipe de enfermagem
Capacitação
Avaliação da educação
Administração de recursos humanos em hospitais

DESCRIPTORES

Grupo de enfermería
Capacitación
Evaluación educacional
Administración de personal en hospitales

¹Associate Professor, Department of Professional Guidance. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. vlmirag@usp.br ²Associate Professor, Department of Professional Guidance. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. marinape@usp.br ³Associate Professor, Department of Professional Guidance. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. melleiro@usp.br ⁴Ph.D. Professor, Department of Professional Guidance. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. daisyrt@usp.br ⁵Ph.D. Professor, Department of Professional Guidance. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. fatima@usp.br ⁶Nurse, Masters Student of the Graduate Program in Nursing Management. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. patriciatavaress@yahoo.com.br ⁷Nurse, Doctoral Student of the Graduate Program in Nursing Management. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. jaqueline.alc@gmail.com ⁸Nutritionist, Doctoral Student of the Graduate Program in Nursing Management. University of São Paulo, School of Nursing. São Paulo, SP, Brazil. enildalara@usp.br ⁹Full Professor, Institute of Social and Occupational Psychology, University of Brasília. Brasília, DF, Brazil. jairo@unb.br

INTRODUCTION

Developing educational actions for health and nursing professionals is fundamental in order to ensure and improve health care quality. It is also a debate theme under the scope of public health policy in the country. Reflections based on these discussions have resulted in the implementation of the National Policy For Permanent Health Education - PNEPS⁽¹⁻²⁾, which recommends workers training and development according to users/population's needs and profile and the changing procedures in the managerial and assistance model focusing on inter-professional practice integrity and health care effectiveness.

Therefore, it should be highlighted that a permanent training proposal refers to acknowledging educational actions limitations as they are being developed, and points at the implementation of a set of changes with a view to transform health practice, specially guiding them according to the specific user needs of each health establishment⁽³⁾.

Among the limitations, there are scarce educational evaluation activities of active professionals in health services, which is a particular evaluation object, acknowledged as necessary for educational actions and knowledge improvement and for its effects in health care quality⁽⁴⁻⁶⁾.

In Brazil, evaluating health professionals' educational actions and programs has not yet been consolidated as a research tradition. As a result, it have been considered a relevant theme that should be investigated.

To do this, a long course in following practices and research must be pursued, including a variety of aspects that will allow for understanding the survey and diagnosis processes of the needs and the expected outcomes, the variables affecting those outcomes, the compliance to educational actions and programs objectives and their effectiveness on health care, as well as defining evaluation criteria and parameters for specific quality indicators of workers' actions and programs, and the evaluation methodology.

The first, and internationally awarded, educational action evaluation model for professionals is that by Donald Kirkpatrick's, who in the 1950's published articles that established four evaluation levels: reaction evaluation – a measure of how participants felt about their education and their personal reactions towards the learning experience; learning evaluation – a measure of the increased knowledge or intellectual capacity, before and after the educational action; behavioral evaluation – a measure of whether participants applied what was learnt with a change in behavior, which could be verified immediately or a few months after the educational activity; and results

evaluation – a measure of the effect on the service or work environment resulting from the improvement in the participants' performance⁽⁵⁾.

This model has received contributions for its improvement by Anthony Cradell Hamblin, who subdivided the assessment level four into two sublevels: level 4 – organization evaluation, regarding the changes in the organizations' functioning, and level 5 – final value evaluation, regarding the changes in achieving final objectives for the organizations⁽⁵⁾.

The Kirkpatrick model was also used in interprofessional education (IPE) studies⁽⁷⁻⁸⁾ as the foundation for a six-level-proposal to evaluate IPE effectiveness: reaction – participants' perception regarding their learning experience and the nature of IPE; perception/attitude changes – reciprocal attitude/perception changes between participants, and the use of teamwork and values in health care for specific groups; abilities and knowledge acquisition – knowledge and abilities associated to IPE learning; behavioral changes – the individual's transfer of interprofessional learning to his/her practice;

organizational practice changes – identifying changes in the health care provided by the organization; benefits for users/patients, family members and community – identifying improvements in patients/clients health conditions. IPE effectiveness is conceived as the health care that produces positive outcomes with acceptable costs with no unacceptable secondary effects⁽⁸⁾.

In Brazil, two models of educational action evaluation were developed by authors from the social psychology field, founded on Kirkpatrick's studies. The first is the Modelo de Avaliação Integrado e Somativo (MAIS -

Summative Integrated Evaluation Model) by Borges-Andrade which adds environmental and processes variables proposing information analysis and integrated interpretation with a view to provide organizational policies and strategies⁽⁹⁾. The second is the Modelo Integrado de Avaliação do Impacto do Treinamento no Trabalho (IMPACT - Integrated Model for Evaluating the Impact of Work Training), developed from MAIS⁽¹⁰⁾, and analyzing the relation between the reaction, learning and impact evaluation levels. Impact is evaluated in this model by the transfer of learning and the influence that training has on the participants' overall performance.

Specifically, regarding results and the impact of educational processes in services quality, it is important to stress that workers are not able to automatically operate a mechanical and direct transposition of capacity to a working situation, since it involves a compound of aspects that regard work processes features, work conditions and the organizational structure, supporting MAIS and IMPACT model proposals.

In Brazil, evaluating health professionals' educational actions and programs has not yet been consolidated as a research tradition. As a result, it have been considered a relevant theme that should be investigated.

In addition, not only changes in professional performance on a short term basis, originated from the employment of new knowledge, abilities and attitudes that would represent results must be considered, but also that these changes must be employed in the mid and long term, and followed by positive effects on the professionals' overall performance. This way, the concept of work training impact is understood as the effects that training exercises have on overall performance, motivation and work abandonment, in the mid and long term⁽¹¹⁻¹²⁾.

Of the available models, the most employed in the health area are reaction and learning evaluations. In this study, we will approach the learning evaluation that presents efficiency and outcomes for educational programs and actions or, as previously mentioned, whether the educational action was capable of promoting or improving the participants' knowledge.

Therefore, the objective of the present study is to support the construction of an evaluation methodology through the analysis, particularly, of learning evaluation focused on the decision-making that favors management and the definition of managerial instruments for the process of planning health educational programs.

OBJECTIVE

To analyze the learning evaluation process performed in training programs provided to a nursing team regarding the efficiency of the programs and the validation of the instruments.

METHOD

This is an empirical study, as its findings will support theoretic formulations for the construction of an evaluation methodology providing a stronger foundation for facts and empirical data, dependant on the theoretical framework, and that will add impact in facilitating practical approximation⁽¹³⁾.

This quantitative, correlation study tested, at the same time, variables to verify how much one variable changed due to the change of another⁽¹⁴⁾. Therefore, variables in the learning evaluation were measured by scores, before and after educational actions, hereon generally referred to as training (T) due to its skills and cognitive learning feature.

The analysis is focused on the learning evaluation process. Hence, this study features as an evaluation research, since it aims at determining the value of a program, resorting to scientific procedures⁽¹⁴⁾.

The analysis is based on two items, specifically, in training efficiency, i.e., whether it contributed for improving the trainee's knowledge, and the validation of the instrument or whether the knowledge test was able to mea-

sure what it was supposed to measure. Although these aspects are interdependent, the primary goal of this division in the analysis is to discuss on the findings, proposing the interventions needed for each of them, and, next, for the training process as a whole.

This study was developed at two hospitals in the city of São Paulo. One of the hospitals is a public teaching hospital - University of São Paulo University Hospital (HU-USP)- and the other is a private philanthropy institution (HPSP). Both are large hospitals providing general services. Both institutions provide a structured educational service, exclusively for the nursing team, and periodically develop programs according to the identified needs.

The analyzed material consisted of learning evaluations regarding six training programs; three from each hospital, totaling 486 evaluations from HU-USP and 507 from HPSP, as follows:

1. Pressure Ulcer Prevention and Treatment Training Program (PUT) – performed at HU-USP from May to July 2005, counting with 96 participants.
2. Adult Cardio-pulmonary Resuscitation Training Program (ACPRT) – performed at HU-USP in July and August 2005, counting with 218 participants.
3. Nursing Entries Training Program (NE) – performed at HU-USP in 2005, counting with 172 participants.
4. Indwelling *catheter care* Training Program (ICC) – performed at HPSP in December 2006, counting with 76 participants.
5. Medication Administration Means Training Program (MAM)– performed at HPSP in December 2007 and January 2008, counting with 410 participants.
6. Dressing Training Program (DT) – performed at HPSP in April 2007, counting with 21 participants.

Data collection was performed using tests developed by the nurse instructors, based on the theoretical contents of the training, consisting of questions to verify the specific knowledge in the trainings. These tests were applied immediately after the program with a view to verify whether the knowledge difference or aimed improvement could be attributed to the T. Pre and post-training tests were identical so there would be no difference in the difficulty level and to compare performance by scores.

The evaluations were applied by the T instructors, appropriately guided in a way to ensure homogeneity in the collection, and also to make participants aware of the objective of the evaluation and the use of the results.

The findings were analyzed by inferential and descriptive statistics, and the Shapiro Wilk test was used to verify the distribution of all variables in order to guide the inferential test of choice.

In order to identify the existence of a statistically significant difference between the pre- and post-training

scores, a comparison of these two moments was performed by applying the Wilcoxon non-parametric test, which is used in paired data, to test the hypothesis of no statistically significant difference; generally, it involves a variable measured in the same individual at two different times. Between measures, there is an intervention over the subjects with a view to verify if the intervention affected the responses. Therefore, for each individual, a difference between final and initial measures was calculated⁽¹⁵⁾.

In order to verify the discriminative power of the learning evaluation instrument, the Chi-Square test was employed using the error interval of $p < 0.05$ in the NE of the HU. For the other trainings, the Fisher exact test was employed to compare the rates of incorrect questions in both times in order to verify whether the questions were able to measure the knowledge it was supposed to measure.

The research project was approved by the HU-USP Research Committee and by the Research and Ethics Boards of both institutions (document CEP 555/05). Participants were instructed regarding the objectives of the evaluation, so that they would participate, spontaneously, with no constraint or fear for sanctions depending on the tests results. For this reason, the authorization from all participants was required to use the study evaluations information, according to a Commitment Agreement, complying with the principles of Resolution 196/96.

RESULTS

Because the objective of this study is focused on the global analysis of learning data, the technical aspects of the training programs will not be analyzed. For the same reason, the taught contents and questions will not be presented and discussed herein. Therefore, the results will be presented into synthesis charts, according to the training program, regarding the efficiency measured by the scores on the learning tests and the capacity to differentiate each question.

According to Table 1, regarding the evaluations of the training programs delivered at HU-USP, we observe there is a significant alteration in the score and in the capacity to differentiate the following questions: 2, 5, 8a, 8d, 10b and 10d from ACPRT and 1a, 1b, 2c, 3a, 3c and 3d from NE.

All questions in PUT had a significant alteration in post scores; however, questions were not able to measure a difference between the pre and post training. Apart from these, questions 1, 3, 4, 6 and 10c from ACPRT presented the same condition.

Questions 8b, 8c and 10a from ACPRT, and 1c, 1d, 2a and 3b from NE did not show a statistical significance between the pre and post-training, since these questions were able to demonstrate power of differentiation.

Questions 7 and 9 from ACPRT and 2b from NE did not show significant alterations in scores and no differentiation power between pre and post-training.

Learning evaluations results from all three trainings taught in HPSP are described in Tables 2 and 3. Questions 1 from ICC and 1 from MAM were not able to discriminate differences between the two moments; neither there was a significant difference between moments.

Table 1 – Comparison of pre- and post training scores and the capacity to differentiate questions between times (pre and post), in trainings performed at HU-USP, São Paulo – 2007

Question	Time	Mean	p* Significance between pre and post-moments	p** Question differentiation capacity
PUT				
1a	Pre	0.21	<0.001	0.289
	Post	0.30		
1b	Pre	0.16	<0.001	0.478
	Post	0.30		
1c	Pre	0.24	0.007	1.000
	Post	0.28		
2	Pre	1.84	<0.001	0.448
	Post	2.03		
3	Pre	1.51	<0.001	1.000
	Post	2.20		
4	Pre	2.52	<0.001	0.453
	Post	3.19		
Total Score	Pre	6.48	<0.001	----
	Post	8.30		
ACPRT				
1	Pre	0.25	< 0.001	0.516
	Post	0.69		
2	Pre	0.67	< 0.001	0.001
	Post	0.91		
3	Pre	0.37	< 0.001	0.253
	Post	0.81		
4	Pre	0.25	< 0.001	0.289
	Post	0.67		
5	Pre	0.36	< 0.001	0.003
	Post	0.60		
6	Pre	0.52	< 0.001	1.000
	Post	0.98		
7	Pre	0.84	0.088	0.335
	Post	0.90		
8a	Pre	0.19	0.016	0.042
	Post	0.20		
8b	Pre	0.24	0.564	0.031
	Post	0.24		
8c	Pre	0.19	0.257	0.021
	Post	0.20		
8d	Pre	0.18	< 0.001	<0.001
	Post	0.23		
9	Pre	0.94	0.527	0.277
	Post	0.96		
10a	Pre	0.15	0.631	<0.00
	Post	0.15		
10b	Pre	0.21	0.005	0.015
	Post	0.24		
10c	Pre	0.16	< 0.001	0.711
	Post	0.24		
10d	Pre	0.07	< 0.001	0.004
	Post	0.12		
Total	Pre	5.16	< 0.001	---
	Post	8.07		

Continued...

...Continuation

Question	Time	Mean	p* Significance between pre and post-moments	p** Question differentiation capacity
NE				
1a	Pre	1.00	0.002	0.001
	Post	1.08		
1b	Pre	0.99	< 0.001	0.0017
	Post	0.80		
1c	Pre	0.95	0.050	0.001
	Post	1.02		
1d	Pre	0.99	0.275	0.0015
	Post	0.96		
2a	Pre	0.35	0.862	0.0024
	Post	0.34		
2b	Pre	0.39	0.655	0.426
	Post	0.40		
2c	Pre	0.30	0.020	0.001
	Post	0.26		
3a	Pre	0.96	0.007	0.001
	Post	1.04		
3b	Pre	0.93	0.050	0.001
	Post	0.99		
3c	Pre	1.04	0.003	0.02
	Post	1.09		
3d	Pre	0.58	0.008	0.001
	Post	0.70		
Total	Pre	8.31	0.007	
	Post	8.64		

Questions 3 from DT and 3 from MAM showed a significant difference between moments, although these questions have not presented a power to differentiate such a difference.

However, question 3b from DT, which had the ability to differentiate the difference between pre and post training times did not present a significant difference.

Table 2 - Percentage distribution of correct answers in the pre and post-training and of the question differentiation capacity in trainings performed in HPSP, São Paulo - 2008

Question	Time	Percentage Frequency	p* Significance between pre and post-moments	p** Question differentiation capacity
ICC				
1	Pre	88.31	1.000	1.000
	Post	98.70		
2	Pre	42.47	0.001	0.001
	Post	80.00		
DT				
4	Pre	94.12		---
	Post	100.00		
5	Pre	17.65		0.669
	Post	11.76		
MAM				
1	Pre	97.80	0.085	1.000
	Post	99.02		
2	Pre	61.92	<0.001	0.00
	Post	96.09		
3	Pre	91.71	0.036	0.057
	Post	99.02		

Questions 2, 3 and 4 from ICC and 4 from MAM presented a significant difference between times, and they also presented the capacity to demonstrate such difference. In questions 3c, 3d and 3e there was no difference between power or differentiation.

Table 3 - Scores results comparison in the pre and post training moments and the question discriminative capacity between moments, in trainings performed in HPSP, São Paulo - 2007

Question	Time	Mean	p* Significance between pre and post-moments	p** Question differentiation capacity
ICC				
3	Pre	2.89	0.01	0.025
	Post	3.38		
4	Pre	1.25	<0.001	0.017
	Post	1.44		
Total	Pre	7.24	<0.001	---
	Post	8.75		
DT				
1	Pre	1.20	0.08	0.537
	Post	1.51		
2	Pre	1.85	0.52	0.118
	Post	1.94		
3	Pre	1.13	0.05	---
	Post	1.38		
Total	Pre	7.89	0.10	
	Post	8.62		
MAM				
4	Pre	1.45	1	0.001
	Post	1.59		
Total	Pre	8.6	<0.001	---
	Post	9.7		

(*) Wilcoxon non-parametric test probability descriptive level
 (**) Fischer descriptive level

DISCUSSION

We observed that the situation is not the ideal, since it was expected that questions would be able to differentiate differences, above all, in the presence of significant alterations in the scores. These results demand a quality analysis on the tests and on the level of difficulty of the questions or, even, participants' prior knowledge level.

The questions were expected to demonstrate a power to measure the difference in scores between pre and post training times. However, some questions demonstrated that they were not able to measure differences.

Recalling the main results regarding learning, we observed the following:

- In HU-USP, all three training programs presented a statistically significant improvement in the final score. In PUT, all seven questions demonstrated a significant increase; however, they did not show differentiation power. In ACPRT, from all 16 questions, 11 presented a significant increase; among them, six had power of differentiation and five did not; five questions presented no significant difference. In 11 questions in NE, six presented a signifi-

cant alteration, among them, four demonstrated increase and two a decrease in scores; five questions presented no significant alterations; except for one question that presented no difference, all other questions could differentiate times.

- In HPSP, only DT presented no significant final score increase. In ICC, from four questions, three presented significant increase by questions that could differentiate; the other question presented no alteration and no ability to differentiate. From nine questions in DT, only one question presented a significant increase, however, it could not differentiate. From the remaining questions, only one could differentiate. The small number of participants (17) may have harmed the statistics analysis. In MAM, from two questions that presented a significant increase, one could differentiate times and the other could not; the same occurred with two questions that were not altered.

Considering, at first, the significant alterations in learning evaluation scores, generally, trainings can be confirmed to influence scores in post-training compared to pre-training.

Also, test questions that regarded new or unusual contents in professional education courses were essential to evaluate knowledge acquisition, since they demonstrated lower levels of correct answers in the pre-training and higher improvement levels in the post-training. ACPRT is an example among nursing technicians and assistants; a fact that demonstrates that new knowledge was acquired and doubts were clarified.

For an objective confirmation of this aspect, verifying the relationship between the score variable and the content knowledge variable is still needed in each training, which can be obtained by the reaction or satisfaction evaluation.

Questions presenting high levels of correct answers, both in the pre and post-training, demonstrate that participants had information regarding the theme before the training, as presented in NE.

On the one hand, results demonstrate an important aspect regarding knowledge acquisition by a constant overcoming and the possibility of knowledge improvement when there is prior knowledge about the theme⁽¹⁶⁾. It also supports the understanding that having previous information helps to acquire new contents and allows for relating new information to the participants' prior knowledge.

On the other hand, results suggest that the participants' and teachers' needs were above the contents taught in these training programs and the people were inadequately selected for the program, or that teachers could not elaborate contents that could meet the participants' needs. We understand this fact as a mistake in surveying the participants and their needs.

Regarding training themes, we found that except for the cardio-respiratory arrest training, which is more complex, and the pressure ulcer training for its new products offers, all other themes - entries, dressings, medication administration and indwelling catheter - regard routine nursing activities. Hence, participants were expected to present a considerable prior knowledge for the training programs.

For instance, from all three training programs performed at the HU, the higher increase in the average score from pre to post training was presented in the ACPRT and the lowest in the NE. It is concluded that, by relating theme and score, ACPRT presented a more complex and new contents to participants, while NE involved the most common and routine content in professional education.

In HPSP, the DT did not present a higher final score in post-training, despite the low *n*. This confirms that, among the themes approached in trainings, the taught contents were already known by participants.

Therefore, each training program demonstrated an improvement in the taught content despite previous knowledge, since PUT demonstrated a significant increase in all questions in the test; in ACPRT, 11 from 16 questions and in NE four out of 11; two questions demonstrated a decrease.

In HPSP, however, improvement was subtle. There was a significant increase in one of four questions in ICC; DT presented one of nine, and two in MAM.

In the instrument perspective - knowledge test-, the ideal scenario is for all questions to be able to measure knowledge under a comparative mode between times. However, only a few more than half the questions demonstrated this feature, considering all six training programs.

Under the learning evaluation point of view, the fact that the instrument presented no capacity to differentiate makes it difficult to confirm the training efficiency, since, although there was a statistical significance between the two times, there were also problems in elaborating the questions.

Such result requires a detailed analysis of the tests, their construction and use, relating contents and objectives for the educational activity and analyzing, in a qualitative way, each of the questions.

Contents measured in the test must be investigated regarding their compatibility with the training contents, since one of the learning evaluation limitations is the non-correspondence between contents required in the tests by the instructors while trainings⁽¹⁷⁾.

Likewise, in order to answer the tests, participants mentally exercise acknowledging the applicability of questions, understanding them as if they were actually using them. The instructor must follow the same thinking to formulate the tests, so that questions will present a practical sense.

In addition, learning occurs when the individual fully dominates the learning object, and this domain is translated by the acquisition of new abilities⁽¹⁸⁾.

Knowing that learning is one of the needed conditions to transfer knowledge to work, however not sufficient for that, deepening evaluation techniques is advised, as well as verifying the relations between learning variables with those of impact and satisfaction.

These results, although not conclusive, support the analysis that detected that there is a poor relationship between the reaction and satisfaction level of evaluation and learning; these variables are strongly correlated with impact⁽¹¹⁾. Also, they corroborate the study regarding educational actions in the health area that demonstrated a predominance of traditional teaching strategies and incipient evaluation experiences⁽¹⁹⁾.

Impact evaluation indicates behavioral changes in the position and in the effectiveness of the training actions at the individual level⁽¹²⁾, from which the need for improving evaluation techniques is presumed, since the final objective is to cause changes in the working environment, hence, impact must be evaluated.

REFERENCES

1. Brasil. Ministério da Saúde. Portaria n. 198/GM, de 13 de fevereiro de 2004. Institui a Política Nacional de Educação Permanente em Saúde como estratégia do Sistema Único de Saúde para a formação e o desenvolvimento de trabalhadores para o setor e dá outras providências. Brasília; 2004.
2. Brasil. Ministério da Saúde. Portaria n. 1.996/GM, de 20 de agosto de 2007. Dispõe sobre as diretrizes para a implementação da Política Nacional de Educação Permanente em Saúde e dá outras providências [Internet]. Brasília; 2007 [citado 2011 ago. 30]. Disponível em: http://portal.saude.gov.br/portal/arquivos/pdf/Portaria_N_1996_GMMS.pdf
3. Peduzzi M, Guerra DAD, Braga CP, Lucena FS, Silva JAM. Atividades educativas de trabalhadores na atenção primária: concepções de educação permanente e de educação continuada em saúde presentes no cotidiano de Unidades Básicas de Saúde em São Paulo. *Interface Comunic Saúde Educ.* 2009;13(30):121-34.
4. Organización Panamericana de la Salud (OPAS). Capacitación del personal de los servicios de salud. Quito; 2002. (Proyectos relacionados com los Procesos de Reforma Sectorial, 137)
5. Mira VL. Avaliação de programas de treinamento e desenvolvimento da equipe de enfermagem de dois hospitais do Município de São Paulo [tese livre-docência]. São Paulo: Escola de Enfermagem, Universidade de São Paulo; 2010.
6. Otrenti E. Avaliação de processos educativos formais para profissionais da área da saúde: revisão integrativa de literatura [dissertação]. São Paulo: Escola de Enfermagem; Universidade de São Paulo; 2011.
7. Barr H, Koppel I, Reeves S, Hammick M, Freeth D. Effective interprofessional education: arguments, assumption & evidence. Oxford: Blackwell/CAIPE; 2005.
8. Freeth D, Hammick M, Reeves S, Koppel I, Barr H. Effective interprofessional education: development, delivery & evaluation. Oxford: Blackwell; 2005.
9. Borges-Andrade JE. Avaliação integrada e somativa em TD&E. In: Borges-Andrade JE, Abbad GS, Mourão L. Treinamento, desenvolvimento e educação em organizações e trabalho: fundamentos para a gestão de pessoas. Porto Alegre: Artmed; 2006. Parte III – Avaliação dos sistemas de TD&E; p. 343-58.
10. Abbad G. Um modelo integrado de avaliação de impacto de treinamento no trabalho [tese doutorado]. Brasília: Universidade de Brasília; 1999.
11. Abbad G, Gama AL, Borges-Andrade JE. Treinamento: análise do relacionamento da avaliação nos níveis de reação, aprendizagem e impacto no trabalho. *Rev Adm Contemp.* 2000;4(3):25-45.
12. Pilati R, Abbad G. Análise fatorial confirmatória da Escala de Impacto do Treinamento no Trabalho. *Psicol Teoria Pesq.* 2005;21(1):43-51.

CONCLUSION

Regarding learning, results demonstrate training efficacy, although acquisition significance or intended knowledge improvement were weakened by measurement instruments that presented differentiation problems of the score variable between the pre and post-training times.

Instructors and participants should, together, analyze the construction of each question, comparing score behavior regarding the proposed objectives to the participants' previous knowledge and trainings contents, which, in addition to explaining the errors of the instrument, will allow for evaluating education, thus leading to a full appreciation of the teaching-learning process and not to learning alone.

From all six training programs evaluated herein, it was observed there is a need for proposing an evaluation methodology, building criteria, instruments and indicators for the purpose. Furthermore, it was also found there is a need to improve the diagnosis of training needs and teaching strategies.

13. Demo P. Metodologia do conhecimento científico. São Paulo: Atlas; 2008.
14. LoBiondo WG, Haber J. Pesquisa em enfermagem: métodos, avaliação crítica e utilização. 4ª ed. Rio de Janeiro: Guanabara Koogan; 2001.
15. Menezes RX, Azevedo RS. Bioestatística não paramétrica. In: Massad E, Menese RX, Silveira PSP, Ortega NRS. Métodos quantitativos em medicina. Barueri: Manole; 2004. p. 307-18.
16. Freire P. Educação e mudança. São Paulo: Paz e Terra; 2001.
17. Schaan MH. Avaliação sistemática de treinamento: guia prático. São Paulo: LTr; 2001.
18. Carvalho AVC, Nascimento LP. Administração de recursos humanos. São Paulo: Pioneira; 2000.
19. Tronchin DMR, Mira VL, Peduzzi M, Ciampone MHT, Mel-leiro MM, Silva JAM, et al. Permanent education of health professionals in public hospital organizations. Rev Esc Enferm USP [Internet]. 2009 [cited 2011 Ago 30];43(n. esp 2):1210-5. Available from: http://www.scielo.br/pdf/reeusp/v43nspe2/en_a11v43s2.pdf