Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989 - 9572

The role of problem-based Mathematics teaching according to the Kirkpatrick's Model on problem-solving performance of mathematics teachers

El papel de la enseñanza de matemáticas basada en problemas de acuerdo con el modelo de Kirkpatrick sobre el desempeño de resolución de problemas de los profesores de matemáticas

Somayeh Karimi[1],
Ahmad Shahvarani[2],
Majid Haghverdi[3],

Islamic Azad University, Iran

Journal for Educators, Teachers and Trainers, Vol. 10 (1)

http://www.ugr.es/~jett/index.php

Date of reception: 15 July 2018

Date of revision: 12 December 2018

Date of acceptance: 18 March 2019

Karimi, S., Shahvarani, A., & Haghverdi, M. (2019). The role of problem-based Mathematics teaching according to the Kirkpatrick's Model on problem-solving performance of mathematics teachers. *Journal for Educators, Teachers and Trainers*, Vol. 10(1). 12 – 26.

[1] Department of mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2] Department of mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran
[3] Mathematics Department, Islamic Azad University, Arak Branch, Iran

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 10 (1)
ISSN 1989 – 9572

http://www.ugr.es/~jett/index.php

The role of problem-based Mathematics teaching according to the Kirkpatrick's Model on problem-solving performance of mathematics teachers

El papel de la enseñanza de matemáticas basada en problemas de acuerdo con el modelo de Kirkpatrick sobre el desempeño de resolución de problemas de los profesores de matemáticas

Somayeh Karimi, somaiehkarimi60@yahoo.com
Ahmad Shahvarani, shahvarani.1393@gmail.com
Majid Haghverdi, m-haghverdi@iau-arak.ac.ir
Islamic Azad University, Iran

Abstract: The process of evaluation is essentially the process of determining the realization of the educational goals in real terms through curriculum and education and represents the changes that occur in human behavior. Therefore, it is necessary that at the end of each training course (such as training classes, workshops, and training seminars), the teacher or evaluators, evaluate the implemented training program. In the curriculum approaches, learning the problem-solving ability is the ultimate goal of mathematics education. This skill requires empowering teachers with problem solving skills as one of the optimal ways to use capacities and to achieve educational goals. Therefore, the main goal of this study was to examine the problem-based mathematics teaching according to the Krikpatrick's model on problem-solving performance of mathematics teachers. The research design was of a pretest-posttest type with a control group. Using simple random sampling method, 100 male and female mathematics teachers, teaching mathematics at the middle school, were selected from Rabat Karim city, Tehran province. In pre-test and post-test of the traditional teaching and problem-solving based teaching in mathematics, data were collected through mathematical problem-solving performance test and Kirkpatrick's four-level questionnaire. Using SPSS software and R software, the results showed a significant difference between the scores of problem-solving performances between the two groups of control and experiment after the training, and through the equations, we showed that each level of the Kirkpatrick's model has a positive effect on the post-test scores of mathematics teachers

Resumen: El proceso de evaluación es esencialmente el proceso de determinar la realización de los objetivos educativos en términos reales a través del currículo y la educación, y representa los cambios que ocurren en el comportamiento humano. Por lo tanto, es necesario que al final de cada curso de capacitación (como clases de capacitación, talleres y seminarios de capacitación), el maestro o evaluadores evalúen el programa de capacitación implementado. En los enfoques curriculares, aprender la capacidad de resolución de problemas es el objetivo final de la educación matemática. Esta habilidad requiere empoderar a los maestros con habilidades de resolución de problemas como una de las formas óptimas para usar las capacidades y alcanzar metas educativas. Por lo tanto, el objetivo principal de este estudio fue examinar la enseñanza de las matemáticas basada en problemas de acuerdo con el modelo de Krikpatrick sobre el rendimiento en la resolución de problemas de los profesores de matemáticas. El diseño de la investigación fue de un tipo de prueba previa y posterior con un grupo de control. Usando un método de muestreo aleatorio simple, se seleccionaron 100 maestros de matemáticas masculinos y femeninos, que enseñan matemáticas en la escuela secundaria, de la ciudad de Rabat Karim, provincia de Teherán. En las pruebas previas y posteriores de la enseñanza tradicional basada en la enseñanza de la resolución de problemas en matemáticas, los datos se recopilaron mediante la prueba de rendimiento de la resolución de problemas matemáticos y el cuestionario de cuatro niveles de Kirkpatrick. Usando el software SPSS y el software R, los resultados mostraron una diferencia significativa entre los puntajes de desempeño de resolución de problemas entre los dos grupos de control y experimento después del entrenamiento, y a través de las ecuaciones, demostramos que cada nivel del modelo de Kirkpatrick tiene un efecto positivo en las puntuaciones posteriores a la prueba de los profesores de matemáticas

Keywords: Mathematics; Problem solving; Kirkpatrick's model; Performance; Mathematics teachers

Palabras clave: Matemáticas; Resolución de problemas; Modelo de Kirkpatrick; Rendimiento, Profesores de matemáticas

1. Introduction

In the traditional view of mathematics teaching, the goal of teaching concepts, algorithms and skills is to create problem-solving ability in learners. Of course, in this view, problem solving has a special interpretation, and that is reaching to the correct answer. That is, the final answer is more important than the process. That is, it views the problem-solving method as a skill without considering the content. Advocates of this attitude in mathematics teaching usually refer to the Polya's (1957) "*How to solve the problem*". According to Stanic & Kilpatrick (1988), those who consider the goal of mathematics teaching as learning the problem-solving skill, view the problem solving as a dynamic and constant process in which the final product is not as important as the methods, steps, strategies, and approaches used by learners. Polya's training (1978) activities give credit to this attitude of mathematics teaching and problem-solving status (Gouya, 1999). In recent decades, most of middle school teachers teach their students to solve the problems by following the standard methods. Unfortunately, in these classes, little time was spent on teaching students how to solve the problem (Ho & Hedberg, 2007). The result is that the students have trouble solving unusual problems and the problems that they encounter around them (Golafshani, 2000). Therefore, mathematics educators have criticized the emphasis of teachers on transferring the knowledge and information to students, and instead they have advised them to develop thinking skills in learners. Since the efficient human resources are considered to be the most valuable source of any organization, much of the investment has been devoted to human resources. The most important tool used for this purpose is teaching, which is used with the aim of improving the quality of the individuals' skill, knowledge and attitude level, empowering the individuals in performing their duties and leading to the success of the organization (Khorasani & Hasanzadeh Barani Kord, 2008). Teachers are considered as the starting point and the main driver of any effective change or evolution in the educational system and due to the importance of education in the development of any country, the issue of training the human resources in education has a particular importance compared to the other domains. A look at the status and performance of the Iranian students in TIMSS shows that in our country, the extent of students' access to problem-solving skills, Hypothesis Construction, computational skills estimation, understanding the geometric concepts, and so on, are much less than it is expected (Department of Middle School Education, 2010). Part of these weaknesses appears to be due to the teaching approaches and methods which play a key role in building a meaningful understanding and in generalizing the learned materials to the real situations (Navidi, 2012). The publication of the TIMSS studies in the fields of mathematics and science education in elementary and secondary school levels have greatly shown the harms and weaknesses of this course, and this warning has been given to the teachers and to those involved in the education system, that they should think about a way out of this situation. This report has a special and important message for the teachers, which is that the current state of mathematics and science instruction should be transformed and it should be shifted from the teacher-centric methods and the mere transfer of knowledge and information to the approach of using active teaching methods (Lashkar Bolouki, 2014).

The process of evaluation is essentially the process of determining the realization of the educational goals in real terms through curriculum and education, and in fact represents the changes that occur in human behavior (Tayler, 1902). Therefore, it is necessary that at the end of each training course (such as training classes, workshops, conferences and training seminars), the teacher or evaluators, evaluate the implemented training program. The ultimate goal in evaluation of the program is to judge and make decisions, based on the evidence and documentation, about continuing, modifying, or abandoning the program. Using evaluation models and approaches helps to formulate basic questions and make optimal use of the evaluation program (Kaufman & Keller, 1994). Sometimes educators regard surveys and satisfaction concerning education, separate from the study of change in behavior, and consider each of them as a kind of separate evaluation in the evaluation process, but a model is needed to demonstrate that all evaluation methods and practices are interconnected and complementary and help the principals in evaluating the curriculum (Kirkpatrick & Kirkpatrick, 2006); a model that does not limit the evaluation of the curriculum to the investigation of the extent of learners' learning, can facilitate understanding the evaluation process, and relate the assessment to the educational goals and outcomes. A simple and practical model in evaluating the educational program is the Kirkpatrick's four-level model. This model has been extensively used in evaluating in-service training programs in organizations, and the used models have been mostly inspired by some kind of Kirkpatrick's pattern (Rajeev, Madan, & Jayarajan, 2009). The current goal is to integrate the problem solving into the mathematics curriculum by educating teachers, so that teachers can use the problem situations as interconnected circles in teaching and for introducing new topics and ultimately use problem solving as an educational peak point to achieve meaningful learning. In this regard, empowering teachers in teaching problem-solving skills to the students through the creative combined use of the viewpoints provided by Polya (1968) and Schoenfeld (1985), is emphasized and the evaluation of the extent of effectiveness of the quality of the provided curriculum, through using a functional model is placed on the agenda in order to identify the areas that need improvement and modification. A model that, as a means of determining the value and credibility of education, includes information that can ensure the appropriateness and usefulness of the training program. Using Kirkpatrick's model, the focus of attention is basically on the evaluation of the curriculum in four levels of learner's reaction, learner's level of learning, learner's level of behavior, and learner's level of organizational results. In order to achieve these goals, we need to determine the extent the undergone trainings had influenced the teacher's reaction, learning the problem-solving knowledge and change in their behaviors and thus lead to the acquisition of teachers' scientific and practical skills and improvement in their organizational performance. On the other hand, we need to determine whether the problem-based mathematics teaching, according to Kirkpatrick's model, positively affect the mathematics teachers' problem-solving performance or not? In the following, we will discuss the theoretical foundations of the problem-solving method and the Kirkpatrick's model.

1.1. Kirkpatrick's four-level model

Kirkpatrick published a four-level model for evaluating educational effectiveness. He recently updated his publications in 2006 with detailed case studies and current samples, but his model has remained the same. The four levels of Kirkpatrick's model include: 1) reaction, 2) learning, 3) behavior, and 4) results (Kirkpatrick & Kirkpatrick, 2006):

The first level, the reaction, addresses the way trainees react to the given exercise. Do they like the instruction? And do they observe its immediate application to their job.

The second level, namely, learning, estimates the apprenticeship learning rate in comparison with the specific learning objectives of the training session. This is assessed by the use of the examinations or through the training sessions at the highest level of educational activities.

The third level, namely behavior, measures the actual changes in behavior while working on tasks that are specifically related to educational goals. For example, does the trainee use the techniques and skills trained during this training program or use some other methods? The third level addresses the measurement of the extent of the trainings being transferred to the workplace.

The fourth level, the results, measures the impact of training on the benefits of the organization. Examples of measurable results, in case that efficiency is the goal, include the time of work completion or the reduction of the number of errors. Each of the factors used to measure the results should be selected in such a way as to match the main educational objectives. Otherwise, the evaluation of the effectiveness of training will be wrong (Jung, 2013).

The Kirkpatrick model is the most commonly used method for evaluating the educational effectiveness (Sugrue & Rivera, 2005; Aguinis & Kraiger, 2009; Twitchell, 2000). It is the model used in all state institutions, corporations, and academic institutions (Tourville, 2005; American

Management Association, 2007; Kirkpatrick Partners, 2013; Praslova, 2010). It has proved to be a successful model to assess the effectiveness of training after the completion of an educational activity and it helps designing and preparing instructional materials and methods (Patel, 2010). Kirkpatrick encourages the use of return on expectations (ROE) instead of return on investments (ROI) as a public effectiveness criterion of the instruction (Kirkpatrick Partners, 2013). The following are the examples of similar research on the current subject of the study in recent years:

In a study on assessing a 12-week accelerated ultrasound intensive course through the Kirkpatrick's four level model of evaluation, Sim (2017) showed that all of these four levels were achieved while being evaluated. Since this intensive course fulfills the requirements of the workplace with regard to the reduced level of supervision and student work-readiness, it is considered worthy. Moreover, the fact that each criterion has been valued to be excellent, adds to the worthiness of this intensive course.

In a study on a model for the evaluation of the training results or outcomes, Moldovan (2016) found that the Kirkpatrick's evaluation model by the use of 4 levels, has a lot of adaptations to several fields of training. The new eQvet-us model of training results and outcome evaluation created in his study, include an improvement of the Kirkpatrick's model by associating the corresponding objectives to the evaluation level. Thus, we had the deduction of two levels. The objectives' level includes motivation, performance, outcome, and knowledge, which is related to the classical evaluation level including the evaluations of reaction, learning, behavior and result. An experiment for a VET course for quality experts in services industry was carried out to evaluate the efficiency of the proposed approach. It was revealed that not only the methodology is helpful in training professionals to plan the intervention, but also it is helpful for the trainee's employers in producing business outcomes, and in understanding the factors that facilitate training transfer.

In a case study of PIA on the evaluation of training in an organization through the model of Kirkpatrick, Rafiq (2015) evaluated the effectiveness of the training on PIA by using the four levels of Kirkpatrick's model comprising of reaction, learning, behavior and Results, respectively. By the use of the theme/construct created based on literature review, Rafiq (2015) evaluated the effectiveness of training at the different levels. Those employees whose training had lately been completed were interviewed for the case of evaluation for level one; Those who had passed the same training about three months were selected for level 2, those who had passed the same training about 6 months were selected for level 3 and those who had passed the same training about year earlier were selected for level 4. The results showed the positive reaction of the participants towards the training except for its short duration, moreover they showed that the respondents have used knowledge and skills which they had learnt from their training course.

In a study on the role of workplace factors according to the Kirkpatrick's Model of training evaluation in distance learning transfer by Aluko & Shonubi (2014), which was originally adopted from a longitudinal study on the impact of a teacher training distance education program on the job performance of graduates, the authors added to the findings of the prior pilot study. One of the researchers with regard to the previous study which applied the Kirkpatrick's model of training evaluation, found a strong relationship between the graduates' class performance and their completion of the program. It was also found that the schools' organizational environment strongly affects the learning transfer in the workplace. They also presented some Suggestions on how the school principals and the educators can effectively work together.

2. Research methodology

This study was of fundamental or applied type and the research method was semi-experimental; in which the participants were selected and then divided into two groups of control and experimental, in a way that the pre-test conditions were the same for them. To collect data in this research project, the pre-test, post-test was used along with control group, with reliance on problem-based teaching before and after training for experimental group and traditional training for the control group; and in order to evaluate the quality of problem-based educational program, Kirkpatrick's questionnaire was implemented for the teachers in experimental group, before and after the training. In this training course, the problem-solving assessment based on the expected goals and the use of the trained techniques in the form of the implementation of the four levels was considered. In order to determine the level of teacher's mastery and ability in problem-solving instruction, the pre-test from subjects related to eight topics of the middle school mathematics book was first taken. Then, to assess the teachers' ability in the provided instructions, a post-test was prepared in the form of questions from the discussed topics, and they were asked to proceed in accordance with the problem-solving process. This was a good opportunity to learn how to use these levels in problem solving instruction. Questionnaire is another instrument that is presented and implemented to evaluate the quality of the curriculum based on the Kirkpatrick's model and represent the learner's attitudes, his amount of learning, learner's behavior and organizational results of his work.

2.1. Research instruments

For pre-test, eight questions were considered from the topics of similarity, mathematical reasoning, factorization, in equation, linear relations, and system of linear equations from the ninth grade and the topics of greatest common divisor of two numbers and volume from the seventh grade, and 2 points were considered for each of the questions, where the total score of the pre-test is calculated out of 16 points and the method of scoring allocated 1 point to the problem solving-based responses and 1 point to the correct responses. Moreover, given that the covered topics in the form of 7 lessons were from the topics of similarity, factorization, inequation, linear relations, system of linear equations, greatest common divisor of two numbers and volume, for post-test, eight questions were from the same topics and 2 points was accounted for each question, and the total score of post-tests was calculated from 16 points. The method of scoring the responses in the post- test for the experimental group was using the four levels of problem solving and 0.5 points was considered for each level, and for the control group 1point was considered for the solution method and 1 point was considered for the correct response, and considering the extent the test questions are representative of the content and objectives of the research, we consulted with the experienced colleagues and it was approved by mathematics experts. Another data collection instrument used in this study was the evaluation questionnaire on the effectiveness of the training course, which is built based on the Kirkpatrick's evaluation model, and includes 33 questions, and the respondents identify their thoughts on the topic by choosing their response from five given options.

Since the lower limit of the points for each question equals one and given the fact that the number of questions in the questionnaire is 33, the lower limit of the questionnaire scores is obtained 33, and as the same way add the points from the above 33 items in the questionnaire. The minimum possible score will be 33 and the maximum possible score will be 165. Considering content validity, the pre- and post- test questions as well as Kirkpatrick's questionnaire were reviewed and approved in accordance with the opinions of experts in the field of mathematics education. On the other hand, the validity or reliability of the mentioned instruments was confirmed through the use of Cronbach Alpha with the values above 0.75. Finally, the analyzes of this study according to two independent sample t–test, paired-Samples t-test, and Analysis of covariance (ANCOVA) were performed using SPSS software; and for analyzing the conceptual model of the research, R software, ggplot2, lavaan and semPlot packages were used.

2.2. Participants

In this study, the population includes mathematical teachers in Rabat Karim city, who were teaching mathematics in the middle school during the academic year 2017-2018. Using simple random sampling method, 100 participants (62 in the experimental group and 38 in the control group) were selected and trained. These teachers were employed at girls' middle schools and boys' middle schools in the city of Rabat Karim during the academic year of 2017-2018. These participants were selected with the experience of teaching mathematics in the ninth grade and seventh grade of Middle school, for one semester.

2.3. Procedure

After choosing the educational topics, such as similarity, volume, factorization, inequation, linear relations, and system of linear equations and greatest common divisor, in educational sessions, four levels and sublevels were completed for each selected question respectively. At the beginning of the work, we explained to the mathematics teachers how they need to act according to the researchers' wish, if they want to teach a subject. The teachers were asked to complete the levels 1 to 4 and it was discussed with them, and the teachers were required to do the discussions in the manner they were asked to do, and they should have been trained. The used framework for classes, under the titles of levels and sublevels is according to Pinter (2012), and it is based on the Hungarian main curriculum, which creates and reinforces the problem-solving skills in student teachers. In this study, during an in-service course and two training workshop courses, the teachers of the experimental group received the necessary trainings for teaching based on problem solving theories of Polya (1965) and Schoenfeld (1985) and in the form of the four mentioned levels, in seven topics of similarity, volume, factorization, inequation, linear relations, system of linear equations and greatest common divisor, for 12 weeks and 4 hours in two days (Thursday and Friday) of per week. The selected problems in this study were from the topics of the middle school (seventh and ninth grades) mathematics books that have been the main concern of the teachers and have been selected in consultation with them, and they are in accordance with the opinion of the teachers who were teaching in teacher training centers and believed that teaching these topics is important given the focus of the subjects in mathematical education and are prioritized to the problems in the eighth grade.

3. Findings

In this section, the statistical findings of the data analysis through the use of SPSS software and the packages of ggplot2, lavaan and semPlot are shown below. First, the descriptive statistics containing demographic information are described:

Table 1.
Distribution of absolute and relative frequencies of control and experimental groups

Relative frequency percentage	Frequency	Group
38	38	Control
62	62	Experimental
100	100	Total

According to the results of table 1, 38% of the mathematics teachers were assigned to the experimental group and 62% were assigned to the control group.

Table 2.
Distribution of absolute and relative frequencies of mathematics teachers' gender

Relative frequency percentage	Frequency	Gender
48	48	Female
52	52	Male
100	100	Total

According to the results of table 2, 48% of the mathematics teachers were female and 52% were male.

Table 3.
Distribution of the absolute and relative frequencies of mathematics teachers' age groups

Relative frequency percentage	Frequency	Age groups
81	81	Less than 25 years old
15	15	Between 25 to 30 years
4	4	Over 30 years old
100	100	Total

According to the results of table 3, 81% of the mathematic teachers under study were in the age group of 25 years, 15% of them were in the age group of 25-30, and 4% were in the age group of over 30 years old.

Table 4.
Distribution of absolute and relative frequencies of mathematics teachers' teaching experience

Relative frequency percentage	Frequency	Teaching experience
75	75	Less than 5 years
21	21	5 to 10 years
4	4	10 to 20 years
100	100	Total

According to the results of table 4, the teaching experience of 75% of mathematics teachers were less than 5 years, 21% were between 5 to 10 years, and 4% was between 10 to 20 years.

Table 5.
Distribution of absolute and relative frequencies of mathematics teachers' educational degree

Relative frequency percentage	Frequency	Educational degree
80	80	B.A
20	20	M.A.
100	100	Total

According to the results of table 5, the educational degrees of 80% of the mathematics teachers under study were B.A, and 20% of them were M.A. Here, we examine and review the extent of problem-solving performance of the mathematics teachers. At this stage, we review the difference between the problem-solving performance scores between the two groups of control and experimental, before and after the training course. Table 6 shows the results of the inferential analysis.

Table 6.
Review of the difference between the problem-solving performance scores between the two groups of control and experimental, before and after the training course

P-value	df	t	Standard deviation	Mean	Intervention	Group
P<0.05	37	4.279	1.9	9.7	Before	control
			1.7	8.6	After	
P<0.05	61	-4.125	2.09	9.1	Before	experimental
			1.5	14.4	After	

Table 6 shows that the relationship between the problem-solving performance score of the training course in the experimental group, before and after the training, is significant with regard to the statistical test (t-test) ($p < 0.05$). In table 7, we review the difference between the problem-solving performance score between the two groups of control and experimental, before and after the problem-based teaching. The results of table 6 show that the mean score of problem solving performance for the control group was 9.7 ± 1.9, before training and 8.6 ± 1.7, after training. The above table shows that the difference between the problem solving performance score of the training course in the control group before and after the training is significant with regard to the statistical test (t-test) ($p < 0.05$). Therefore, it seems that in later analysis, this variable should be controlled by statistical methods such as covariance analysis. Moreover, the mean score of problem solving performance for the experimental group was 9.1 ± 2.09, before training and 14.4 ± 1.5, after training. Table 6 shows that the relationship between the problem solving performance scores of the training course in the experimental group before and after the training is significant with regard to the statistical test (t-test) ($p < 0.05$). In table 7, we examine and review the difference between the problem-solving performance score between the two groups of control and experimental, before and after the problem-based training.

Table 7.
Review of the difference between the problem-solving performance score between the two groups of control and experimental, before and after the training

P-value	df	t	Standard deviation	Mean	Group	Intervention
0.1	98	1.64	1.9	9.7	Control	Before
			2.09	9.1	Experimental	
P<0.05	98	-17.72	1.7	8.6	Control	After
			1.5	14.4	Experimental	

The results of table 7 show that the mean of problem-solving performance score of the training course before training was 9.7 ± 1.9 in the control group and 9.1 ± 2.09 in the experimental group. Table 7 shows that the difference in problem solving performance score before training between the two groups of control and experimental is not significant with regard to the statistical test (t-test) ($p > 0.05$). Moreover, the mean of problem-solving performance score of the training course after training was 8.6 ± 1.7 in the control group and 14.4 ± 1.5 in the experimental group. Table 7 shows that the difference in problem solving performance score after training between the two groups of control and experimental is significant with regard to the statistical test (t-test) ($p < 0.05$). In the following, in table 8, we examine and review the difference between the mean scores of the levels of Kirkpatrick's model in training course between the two groups of control and experiment.

Table 8.
Review of the difference between the mean scores in the levels of Kirkpatrick's model in training course between the two groups of control and experiment

Level of significance (P-value)	df	t	Standard deviation	Mean	Group	Domain
P<0.05	98	-17.9	6.6	22.8	Control	Reaction
			8.2	51.1	Experimental	
P<0.05	98	-15.59	3.2	10.9	Control	Learning
			4.1	24.1	Experimental	
P<0.05	98	-12.70	3.2	8.6	Control	Behavior
			4.1	18.6	Experimental	
P<0.05	98	-12.41	5.7	15.8	Control	Organizational results
			7.5	33.5	Experimental	

The results of table 8 show that the mean of the reaction domain was 22.8 ± 6.6 in the control group and 51.1 ± 2.8 in the experimental group. Table 8 shows that the difference between the control and experimental groups in the domain of reaction is significant with regard to the statistical test (t-test) (p <0.05). The mean of learning domain in the control group was 10.9 ± 3.2 and in the experimental group was 24.1 ± 4.1. Table 8 shows that the difference in learning domain between the two groups of control and experimental is significant with regard to the statistical test (t-test) (p <0.05). The mean of behavior domain in the control group was 8.6 ± 2.3 and in the experimental group was 18.6 ± 4.1. Table 8 shows that the difference in behavior domain between the two groups of control and experimental is significant with regard to the statistical test (t-test) (p <0.05). The mean of organizational results domain in the control group was 15.8 ± 7.5 and in the experimental group was 33.5 ± 7.5. Table 8 shows that the difference in organizational results domain between the two groups of control and experimental is significant with regard to the statistical test (t-test) (p <0.05). In this section, we will check the difference of the problem-solving performance scores between the two groups of control and experimental after the training, taking into account the teacher's score before the training. Since a significant difference had been observed between the teachers' scores before the training, it is necessary to check the significance of the difference in teachers' score after the training by eliminating the effect of teachers' scores before the training. To achieve this goal, the covariance analysis (ANCOVA) was used. One of the important assumptions in taking this test is the equality of the regression line slope of the post-test score regression on the pre-test score regression. To do so, the covariance analysis model was conducted despite the interaction between the pre-test score and the training group. If this interaction is not meaningful, it implies that the assumption of the slope equality is observed. Information of doing this test is presented in table 9.

Table 9.
Review of the equality of regression line slope in the covariance analysis model

P-value	F	Mean-square	Degree of freedom	Sum of squares	Source of error
P<0.05	36.27	68.77	1	68.77	Effect of the educational group
P<0.05	34.80	65.97	1	65.97	Pre-teaching score
0.09	2.89	5.49	1	5.49	The intraction between the previous score and the educational group
		1.90	96	181.99	Model error
			100	16074.00	Total

According to the results of this test, as shown in table 9, the interaction between the pre-teaching score and the educational group is not significant (p <0.05). Then, we review the covariance analysis for the difference between the problem-solving performance scores between the two groups of control and experimental after teaching, with regard to the teacher's scores before teaching:

Table 10.
Analysis of covariance (ANCOVA)

P-value	F	Mean-square	Degree of freedom	Sum of squares	Source of error
P<0.05	31.31	60.51	1.00	60.51	Pre-teaching score
P<0.05	437.94	846.44	1.00	846.44	Effect of the educational group
		1.93	97.00	187.48	Model error
			100.00	16074.00	Total

According to the results of this test, as shown in table 10, the difference between the problem-solving performance score between the two groups of control and experimental, after training is significant, with regard to the teacher's scores before training (p <0.05). Now, in order to review the research concept model that was designed by the researchers, we would like to examine the problem solving performance of the teachers based on the levels of Kirkpatrick's model after the problem-based training in the experimental group more precisely and with more fitting details. The conceptual model of the research was designed as follows:

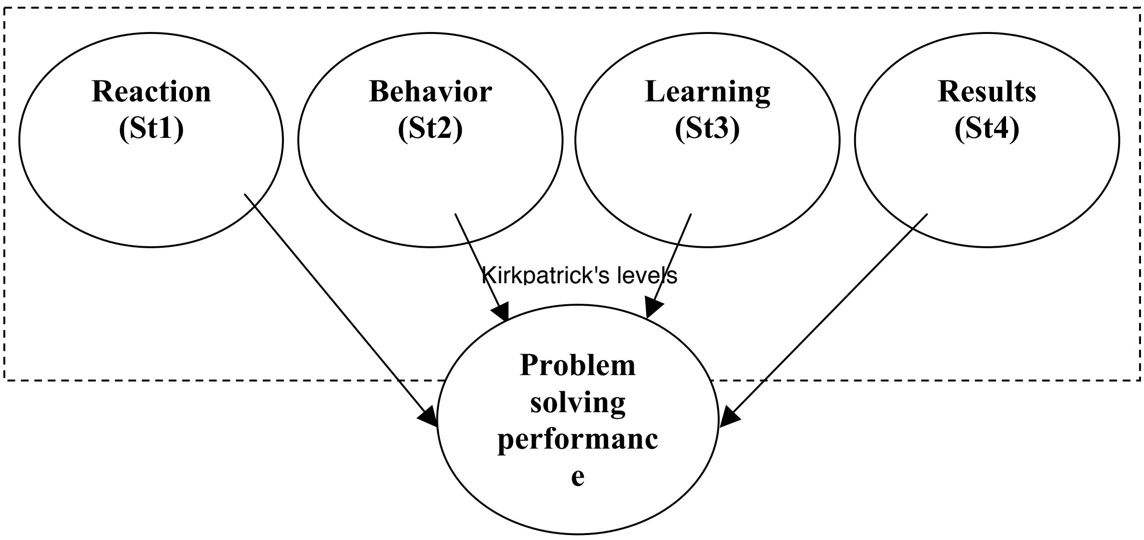

Figure 1. Conceptual model of research

To study the conceptual model of research in figure 1, we use the structural equation modeling, with regard to the analysis of the R software, and the packages of ggplot2, lavaan and semPlot. In figure 2, the variable PS, refers to the post-test score of the mathematical teacher, the rest of the symbols are also specified, and figure 2 is the output of the software:

In figure 2, we see the conceptual model of research along with the standard factor loads, and in table 11, we see the goodness of fit indicators. Also, in table 12 the non-standard factor loads are represented with their degree of significance.

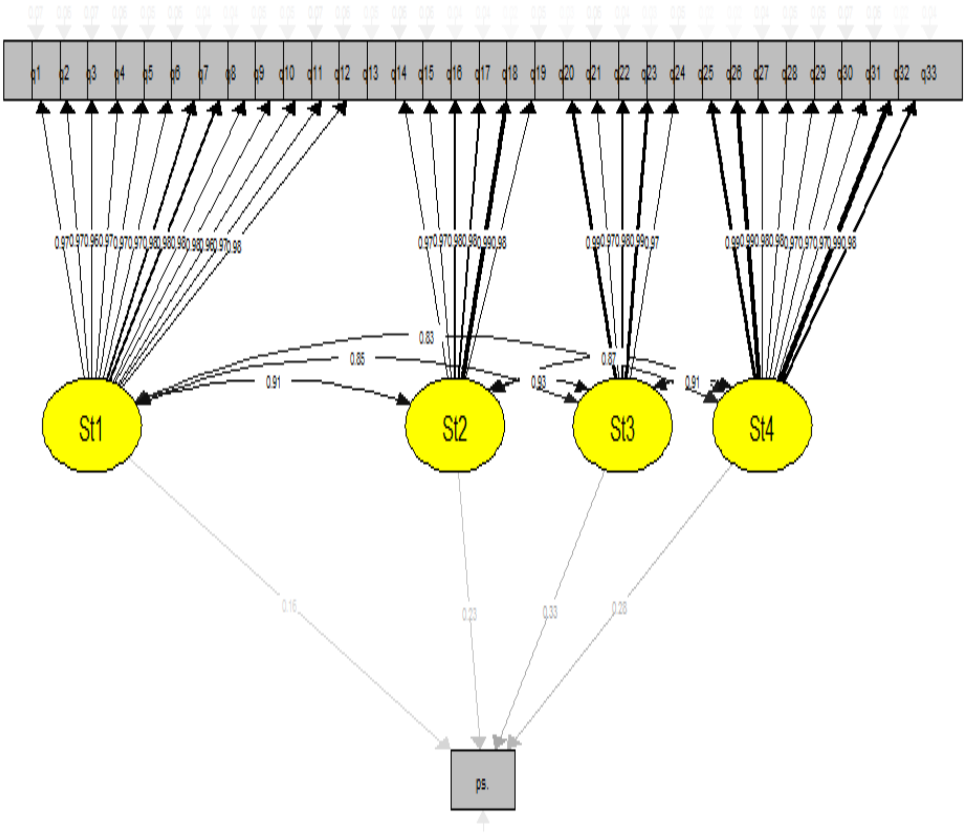

Figure 2. The implemented conceptual model with standard factor loads

Table 11.
Goodness of fit indicators

Result	The observed goodness of fit indicators	Index limits for the goodness of fit	Index limits for the acceptable fit	Types of the goodness of fit indicators
Good fitting	(518)726.37	Ratio of X^2 statistic to the degree of freedom 3	Ratio of X^2 statistic to the degree of freedom less than 5	(df)X^2 statistic value
Inappropriate fitting	<0.05			X2 test P-value
Good fitting	1.40			Ratio of X^2 statistic to the degree of freedom
Acceptable fitting	0.06	Less than 0.05	Less than 0.08	RMSEA
Good fitting	0.09	More than 0.1	more than 0.05	P(RMSEA<0.05)
Good fitting	0.97	More than 0.95	more than 0.90	CFI
Acceptable	0.89	More than 0.95	more than 0.90	NNFI
Acceptable	0.86	More than 0.90	more than 0.85	GFI
Acceptable	0.89	More than 0.90	more than 0.85	AGFI

Regarding to the values of goodness of fit indicators in table 11, this model is at a good and at least acceptable level in terms of all goodness of fit indicators and according to the data of this study. The only indicator that indicates that this model does not have goodness of fit, is chi-square test p-value, which reason can be the high sample size. In table 12, the extent of their direct effects and their significant are examined and reviewed for each of the direct paths of the second conceptual framework of the study.

Table 12.
Review of the significance of path coefficients related to direct effects according to the conceptual model of research

P-value	t-statistic value	standard error	Estimation of standard path coefficients	Estimation of nonstandard path coefficients	Effect		
<0.05	2.05	0.19	0.16	0.41	Post.test	--->	St1
<0.05	2.15	0.29	0.23	0.63	Post.test	--->	St2
<0.01	3.19	0.27	0.33	0.86	Post.test	--->	St3
<0.01	0.20	0.20	0.27	0.71	Post.test	--->	St4

According to the p-value obtained from the t-test for the significance of each of these paths in table 12, the effects of these pathways are significant (p <0.01) and according to these paths' signs of coefficients, this indicates the positive effect of each level in Kirkpatrick's model on the post-test scores of the mathematics teachers. Moreover, with regard to the standard load factor values, since the third level of the Kirkpatrick's model has a larger standardized coefficient, it can be said that the third level of the Kirkpatrick's model had the most effect on the post-test scores of the mathematical teachers. Also, the second most important level was the fourth level, the third largest level is the second level, and finally, it can be said that the first level of Kirkpatrick’s model had the least effect on the post-test scores of the mathematics teachers.

4. Conclusion

In the curriculum approaches, learning the problem-solving ability is the ultimate goal of mathematics education. This skill requires all the skills and abilities that exist in mathematics and it somehow applies all the mathematical skills. Teachers' empowerment in problem solving skills is one of the optimal ways to use capacities and to achieve educational goals. Designing and solving practical problems that can address real-life situations in mathematical language requires powerful teachers who have various representations of the problems and activities, relevant to the stage of students' progress. What is at the heart of the efforts to recognize the common professional development is teachers' learning, learning how to learn, and transferring the knowledge to action to help students' progress (Avalos, 2011). Any kind of teaching, including teaching mathematics, requires certain terms and conditions. First, the student and his abilities must be recognized and then he must be guided in a correct way. However, this guidance must be done by a teacher familiar with the correct way of teaching. If our goal is to provide a context for teaching problem solving skill to the students, the teachers, themselves, must first learn the problem solving clearly (Chapman, 2005). The mathematics teacher training curriculum in our country has remained unchanged for more than 20 years. Therefore, one of our priorities should be the inclusion of a problem-solving approach in the mathematics teacher training curriculum. But this alone is not enough, and for the current mathematics teachers, appropriate in-service training courses should be held. We expect mathematics teachers to do something they have not been trained for; we ask them to teach in such a way that they have not learned during their training course. Typically, mathematics teachers teach in the same way as they are taught to teach. Human resources are considered as the most important and strategic source of any organization. Since every year the various educational organizations spend a lot of money to train specific skills, and given that the implementation of such programs and training courses take a lot of time, doing an accurate and scientific evaluation that show the strengths and weaknesses, as well as the ways to improve and correct them, and also the extent to which goals are achieved, and in general, illustrate the status of educational effectiveness of these courses, is necessary. Kirkpatrick's four-level model provides a practical framework for training practitioners to evaluate the effectiveness of educational programs.

In the study, first a problem-based education program for teaching middle school mathematics in a training course was designed and implemented for middle school mathematics teachers based on the theoretical frameworks of Polya (1978) and Schoenfeld (1985). After the problem-based trainings in four levels, as it was outlined in the procedure section of the study, a test and a questionnaire were used, where the standard questionnaire based on the four levels of Kirkpatrick were distributed among the teachers. After examining the pre-tests and post-tests of the problem-solving performance, a significant difference was found between the problem-solving performance in two groups of control and experimental after the training, with regard to the teachers' scores before the training, then we designed a new mode to compare the relationship between Kirkpatrick's levels with the post-test of mathematics teacher's problem-solving performance. Therefore, by structural equations in R software, we showed that each level of Kirkpatrick model has a positive effect on the mathematics teachers' post-test scores. Moreover, with regard to the standard load factor values, since the third level of the Kirkpatrick model has a larger standardized coefficient, it can be said that the third level of the Kirkpatrick's model has the greatest impact on the post-test score of the mathematics teachers. Also, the second most important level is the fourth level, the third most important level is the second level, and finally, it can be said that the first level of Kirkpatrick model has the least effect on the post-test scores of the mathematics teachers. When the assessment is done in a coherent way and in a pre-determined manner, such as in Kirkpatrick's model, educational approaches and training courses for teachers are examined in a more coherent way. The problem-solving approach was implemented for mathematics teachers in accordance with Polya (1978) and Schoenfeld (1985). frameworks, so that mathematics teachers can develop and expand teaching mathematic problems in middle school with a more recent approach to discover problem solving knowledge and skills among students. Therefore, with regard to the better performance of mathematics teachers after the problem-based training course in several topics of the middle school mathematics book, suggestions for further research are presented. It is suggested that the Kilpatrick's levels be held, checked and reviewed at different time periods during a school year among mathematics teachers, so that the in-service training courses will be estimated and evaluated with coherence and with a more accurate estimation of the mathematical performance of the teachers. It is also suggested that the problem-solving approach and in-service training courses be implemented based on the frameworks of Polya (1978) and Schoenfeld (1985).in the same way as geometry, and then the outcomes of the training course be evaluated according to the Kirkpatrick's four levels.

5. 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**.**

Aluko, F.R. & Shonubi, O.K. (2014). Going beyond Kirkpatrick's Training Evaluation Model: The role of workplace factors in distance learning transfer, *Africa Education Review,* 11(4), 638-657.

American Management Association. (2007). Effective Training: Implementing the Four Levels.

Avalos, B. (2011). Teacher professional development in teaching and teacher education oven ten years, *Teaching and Teacher Education*, 27(1), 10-20.

Chapman, O. (2005). Constructing pedagogical knowledge of problems solving pre-service mathematics teachers. In H.L. Chick & J.L. Vincent (Eds.), Proceedings of the 29th Meeting of the International Group for the Psychology of Mathematics Education (vol. 2, pp. 225-232). Melbourne, AU.

Department of middle school education. (2010). The program of enriching learning experiences in mathematics and science.

Golafshani, N. (2000). Teachers conception of mathematical and their instructional practices, university of Toronto.

Gouya, Z. (1999). *The role of met cognition in learning mathematical problem solving*. Journal of Mathematics Education, 53,13-18.

Jung, C. (2013). Evaluation in Instructional Design - Kirkpatrick's 4 Level Model from: Hewlett-Packard, 2004 Summary Process for measuring ROI of hp, HewlettPackard Development Company, L.P.

Ho, K.F., & Hedberg, J.G. (2007). Teacher's pedagogies and their impact on student's mathematical Problem solving, *Journal of Mathematical Behavior*, 24, 238-252

Kaufman, R., & Keller, JM. (1994). Levels of evaluation: beyond Kirkpatrick, *Human Resource Development Quarterly*, 5(4), 371-8.

Kirkpatrick, DL., & Kirkpatrick, JD. (2006). Evaluating Training Programs. The Four Levels. 3rd ed. San Francisco, California: Berrett-Koehler Store.

Kirkpatrick Partners, LLC. (2013). Training Evaluation That Delivers Measurable Business Results. 2013 [cited 2013 17 April]; Available from: http://www.kirkpatrickpartners.com/Home/tabid/38/Default.aspx.

Khorasani, A. & Hasanzadeh Barani Kord, B.(2008). Educational Needs Assessment Strategies. Tehran: Iranian Industrial Research and Training Center.

Lashkar Bolouki, Gh,R.(2014). Iranian students in the mirror of TIMSS 2011, Journal of developments in middle school education, 18(8).

Moldovan, L. (2016). Training Outcome Evaluation Model, *Procedia Technology*, 22, 1184-1190.

Navidi, A. (2012). An evaluation of the trial Execution of "the program of enriching learning experiences in mathematics and science" in the first grade of middle school during the academic year 2011-2012.

Patel, L. (2010). ASTD State of the Industry Report 2010, American Society for Training & Development: Alexandria, VA.

Pinter, K. (2012). On Teaching Mathematical Problem-Solving and Problem Posing.

Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education, *Educational Assessment, Evaluation and Accountability*, 22(3), 215-225.

Polya, G. (1957). How to solve it. Second edition. Princeton: Princeton university press.

Polya, G. & MAA. (1965). Let Us Teach Guessing [Film]. Produced by MAA (Mathematical Association of America). Washington, DC: MAA. Retrieved March 7, 2015 from https://vimeo.com/48768091

Polya, G. (1968). Mathematics and plausible reasoning (Vol. 2): Patterns of plausible inference. Princeton: Princeton university press. (Original work published 1954).

Polya, G. (1978). Guessing and proving. The Two-Year College Mathematics Journal, 9(1), 21-27.

Rafiq, M.(2015). Training Evaluation in an Organization using Kirkpatrick Model: A Case Study of PIA, *Journal of Entrepreneurship & Organization Management*, 4(151).

Rajeev, P., Madan, M., Jayarajan, K. (2009). Revisiting Kirkpatrick's model-an evaluation of an academic training course, *Current science*, 96(2), 272-6

Sugrue, B., & Rivera, R.J. (2005). State of the Industry: ASTD's Annual Review of Trends in Workplace Learning and Performance, American Society of Training and Development: Alexandria, VA.

Schoenfeld, A. (1985). Mathematical problem solving. New York: Academic Press.

Stanic, G., & Kilpatrick, J. (1988). Historical perspectives on problem solving in the mathematics curriculum. In R. Charles & E. A. Silver (Eds.), The teaching and assessing of mathematical problem solving: Research agenda for mathematics education series (Vol. 3, pp. 1-22). Reston, Virginia: National Council of Teachers of Mathematics.

Sim, J. (2017). Using Kirkpatrick Four Level Evaluation model to assess a 12-week accelerated ultrasound intensive course, *Sonography*, 4(3), 110-119.

Tayler, R, W. (1902). Basic Principles of Curriculum and Instruction; [trans.] Ali Taqipour Zahir(2010). Tehran: Agah Publications.

Twitchell, S. E.F. (2000). Holton III, and J.W. Trott Jr, Technical Training Evaluation Practices in the United States, *Perform Improvement Quarterly*, 13(1), 84-109.

Tourville, S. (2005). Lockheed Martin Training Solutions, Lockheed Martin Global Training and Logistics: Orlando, FL.