A Meta-Analysis of Instructional Management Model for Students' Creative Thinking Development: An Application of Propensity Score Matching

Tanapat Itsarangkul Na Ayutthaya [ID]
Chiang Mai University, THAILAND

Suntonrapot Damrongpanit* [ID]
Chiang Mai University, THAILAND

Abstract: The research emphasized three main objectives: 1) to analyze the propensity score of the research effect size for developing students' creative thinking, 2) to study the attribute variables effect of the research on the effect size of creative thinking before and after the propensity score adjustments, and 3) to compare the effect size between instructional methods to develop creative thinking before and after the propensity score adjustments. The data were obtained from 400 research studies on creative thinking developmpent in Thailand. The research instrument for data collection included the research attribute record form. They were analyzed by calculating effect size, propensity score matching analysis, and fixed effect and random effect meta-regression analysis. The results indicated two research groups with propensity scores that develop students' creative thinking: the low effect size group of 256 research ($\bar{d}$=1.345) and the high effect size group of 144 research ($\bar{d}$=7.284) using 26 attribute variables of creative thinking development research. Moreover, the instructional methods with the creative activities had the highest effect size ($\bar{d}$=3.88). After the analysis of propensity score matching, the effect of 12 research attribute variables was eliminated as follows: manufacturing research institutions, year of publication, educational institutions, curriculum, creative thinking indicators, instructional materials, types of research, research objectives, research groups, research protocols, statistics used in research, quality of research and it was found that integrated instructional model of knowledge using media and technology had the highest effect size ($\bar{d}$=0.41).

Keywords: *Creative thinking, instructional management model, meta-analysis, research synthesis, propensity score matching.*

Introduction

During the current rapid social changes affecting the survival of all human beings thoroughly (Karatas & Zeybek, 2020), each country has started to recognize the importance of developing human beings with sufficient skills to prepare for these changes. Therefore, the new era of education management has to focus on preparing people to face rapid, violent, volatile, and unpredictable changes to create a new generation of people with many skills in learning and adaptation (Kirikkaleli et al., 2021). Instructional management, therefore, emphasizes the development of students' diverse abilities in Cognitive, affective, and psychomotor domains along with advanced thinking skills development such as critical thinking, communication, collaboration, and creativity (Ciğerci, 2020; Herdem, 2019; Trilling & Fadel, 2009). Since the body of knowledge from prior research showed that intelligence and creative thinking were directly related, it was found that students with higher academic background knowledge and abilities were more likely to develop creative thinking highly (Gajda et al., 2017; Kim, 2005). A rapidly changing world thereupon requires the development of the cognitive domain along with creative thinking for the pursuit of new knowledge and adaptation to a globalized society (Ciğerci, 2020). According to the mentioned importance, scholars worldwide are trying to find a way to develop creative thinking for the youths in each country. This makes it possible to see a variety of approaches to developing students' creative thinking in the present day.

In Thailand, researchers have gathered, and synthesized research related to instructional innovations to develop creative thinking; it was found that more than 400 researches were conducted from 1971 to the present. Up to five different teaching-learning groups can be grouped as follows: 1) All-round creative thinking development model (n=80)

---

* **Corresponding author:**

Suntonrapot Damrongpanit, Faculty of Education, Chiang Mai University, Thailand. ✉ suntonrapot.d@cmu.ac.th

emphasizes developing creative thinking through a variety of methods, starting from creating motivation to study, paying attention to students, working in a team, organizing activities creatively, encouraging students to practice thinking, practice, and practice self-development; 2) Creative thinking skills stimulation model (n=25) focuses on the creative arrangement of environment and atmosphere in the classroom, encouraging and motivating learning to motivate the students to learn in a creative environment; 3) Knowledge integration based on creative environment through the technology media usage model (n=68) emphasizes integrating knowledge beyond the subject matter and connecting it to real-life through training for the students to learn by themselves together with the media and technology usage in learning management; 4) Self-knowledge construction with the technology media model (n=124) emphasizes on the students to practice until self-learning by using media and technology to help manage to learn; and 5) Creative learning activities and situations management model (n=103) emphasizes on training the students to learn from activities and situations for them to express themselves creatively. Due to this conclusion, the researcher found three significant observations. *First*, all innovation research results successfully developed students' creative thinking. In contrast, the students' results on the ordinary national educational test (O-NET) also had a low level of creative thinking assessment (National Institute of Educational Testing Service, 2020); it was possible that such research may contain publication bias and should be studied systematically to recognize that bias; *Second,* from the many research results, there had still a lack of organization of information systems that help decision-making. As result, teachers and educational personnel were unable to choose an instructional model to develop creative thinking in their context; and *Third*, the design of innovations was diversified and lacked procedural suggestions for creating a new instructional style to develop the students in the globalization era. Referring to the above observations, the researcher conducted the meta-analysis of preliminary effect size revealed that creative activities and situations of learning activities had the most significant influence on the development of students' creative thinking ($\bar{d}$=3.88). However, as mentioned earlier, the conclusion is based on the conclusion that Effect size characteristics have high variance (I^2=99.227%) and tend to have an effect of publication bias in each study. It reflects that the problems arising from the meta-analysis showed the conclusions about the effectiveness of the instructional model without credibility, and other instructional models may develop more creative thinking. The application of various research methods often causes such problems, the use of research protocols with less control over the effect of variables, the use of one group of samples in the experiment, and the low quality of research instruments (Alinaghi & Reed, 2018; Bom & Rachinger, 2019). This often results in research with high effect sizes being less likely to have lower-quality assessment results. On the other hand, when the research quality is lower, the effect size tends to be higher, considered the significant reason for publication bias; these effects should be controlled or eliminated before testing the differences in the effectiveness of students' creative thinking development from the latest instructional styles.

Based on the above problems, the researcher studied the concept and applied the analysis of the Propensity Score Matching (PSM) (Rosenbaum & Rubin, 1983) to control the excess variables from the research in order to eliminate the effect of the extraneous variables affecting the research results (Fitzmaurice, 2006; Sturmer et al., 2006). This is achieved by matching the study group and the control group through the Propensity Score analysis, which is the effect of excess variables with equal or similar values. As a result, the causal arguments and study results are completer and more precise. This method is often applied to reduce publication bias with many quasi-experimental and observational bias studies (Gray et al., 2017; Justus, et al., 2014; Yu et al., 2017). It is in line with research by Arikan et al. (2018), which studied "The use of propensity score matching helps to understand sources of DIF and mathematics performance differences of Indonesian, Turkish, Australian, and Dutch students in PISA." The findings revealed that the propensity score provided a new tool to control better sources of DIF and country differences in PISA mathematics performance. The propensity score matching is a new and beneficial innovative method for controlling the extraneous variables that influence the dependent variable well. Moreover, the propensity scores effectively eliminate all extraneous variables to analyze according to the research objectives to obtain the complete research results.

According to the reasons and importance mentioned above, this research aims to apply propensity score matching to eliminate the effect of research attribute variables on the effect size of instructional styles from different research methods in individual research. The research synthesis revealed that there were three groups of 26 variables of the most favored attribute variables used in innovative research to develop creative thinking, namely: 1) The research base variables consisted of 6 variables, namely the researcher's gender, manufacturing research institutions, the field of research, types of research, year of publication, and educational institutions; 2) The content variables of the research consisted of 13 variables as course provision, curriculum, creative thinking indicators, creative thinking theories, educational level, number of lesson plans, numbers of instructional weeks, amount of time per lesson plan, total instructional time, learning process, evaluation, instructional materials, and learning resources; and 3) The methodological variables consisted included seven variables, namely research objectives, research groups, sample selection, research protocols, quality of tools, research statistics as well as the quality of research. Therefore, the researcher speculated that the use of propensity score matching would help to eliminate the effect of research attribute variables resulting in the individual studies and reduce publication bias. Nevertheless, it will reveal the facts about the propensity score matching with effective instructional styles in developing the students' creative thinking in the academic field more accurately and reliably than ever.

Literature Review

Creative Thinking

Educational management for the students in the 21st century focuses on developing the students to acquire knowledge, skills, and expertise and be able to apply knowledge for self-development and innovate to appropriately solve social problems (Karatas & Zeybek, 2020). In the past, each country's educational management aimed at human resource development to prepare for the rapid and more complex changes of the new world (Kirikkaleli et al., 2021). These thinking skills are the primary key to developing the lifelong learning process. Among these factors, scholars agree that creative thinking is an essential quality and essential to living in the 21st century (In'am & Sutrisno, 2021; Özyurt & Özyurt, 2020; Shah & Gustafsson, 2021; Wang, 2020).

Creative thinking refers to the ability of the brain to see the relationship of surroundings, resulting in learning and understanding. Imaginative thinking leads to inventing new things to meet the needs or solve problems arising in daily life (Dani et al., 2018; Royston & Reiter-Palmon, 2019; Taylor & Kaufman, 2021). Bloom's Taxonomy cognitive skills theory describes creative thinking as the ultimate cognitive skill for explaining a person's abilities (Brown, 2004). Children with high creative thinking have broad, fast, and multi-directional thinking abilities, leading to the discovery of answers that may be new ideas or innovations. Conversely, for children with a lack of creative thinking, their brain processes are so shut down that they cannot integrate the basic ideas arising from learning and experiences into their imaginations (Algahtani, 2017). The factors affecting the development of creative thinking can be divided into two areas: 1) Demographic factors such as age, gender, educational level, etc. (Royston & Reiter-Palmon, 2019; Taylor & Kaufman, 2021); 2) Environmental factors such as the current environment, parenting, instructional management model, and assessment, etc. (Dani et al., 2018; Ulger, 2019). Therefore, creative thinking is essential to surviving in the 21st century to respond to the rapid changes in the modern world.

Meta-Analysis of Instructional Management Model

A meta-analysis is a quantitative analysis technique to conclude a synthesis of several studies aimed at the same issue with statistical methods used to estimate effect sizes (Cohen, 1988). This technique is applied to draw conclusions, solve the conflict problems in each study, and research the same variable in different contexts. As a result, it becomes an essential moderating variable and influences the change in the relationship between variables (Dowdy et al., 2020). Therefore, to conclude the effectiveness of instructional methods in developing the creative thinking in Thailand from 1971 to the present, up to 400 studies related to the development of students' creative thinking can be obtained from a search in ThaiLIS Digital Collection database includes Chiang Mai University E-Library, Chulalongkorn University E-Library and Mahidol University E-Library. Among these, the researcher has organized such research groups according to five groups of the instructional nature, as shown in Table 1.

Table 1. Instructional Management Model Developing the Creative Thinking in each Group.

Instructional Management Models	Instructional Methods	Frequency
1) All-Round Creative Thinking Development Model (n=80) means supervising and enhancing learning motivation, practicing thinking and practicing independently, working as a team, and organizing creative activities. It will be effective the primary school ($\bar{d}$=2.94) as a creative skill enhancement subject ($\bar{d}$=3.35).	1) Creative Development Theory	28
	2) 4MAT Learning Cycle Model	8
	3) Multiple Intelligences	9
	4) Project-Based Learning	24
	5) Brain-Based Learning	6
	6) Six Thinking Hats	5
2) Creative Thinking Skills Stimulation Model (n=25) means arranging classrooms' environment and atmosphere, encouraging and motivating to do teamwork. It will be effective in the primary school ($\bar{d}$=3.40) as a science subject ($\bar{d}$=3.77).	1) Team-Based Learning	8
	2) Brainstorming	7
	3) Cooperative Learning	5
	4) Creativity based Learning	2
	5) Forecasting Method and Concept mappin	3
3) Knowledge Integration based on Creative Environment through the Technology Media Usage Model (n=68) means integrating knowledge and real-life through the media and technology usage. It will be effective with the primary school ($\bar{d}$=4.54) as an art subject ($\bar{d}$=5.06).	1) Science Technology Engineering and Mathematics Education (STEM Education	17
	2) Science Technology Engineering Art and Mathematics Education (STEAM Education)	5
	3) Synectics	26
	4) Learning Integration	12
	5) Blended Learning	8

Table 1. Continued

Instructional Management Models	Instructional Methods	Frequency
4) Self-Knowledge Construction with the Technology and Media Model (n=124) promotes self-learning by using media and technology to help manage to learn. It will be effective with the primary school ($\bar{d}$=5.02) as a mathematics subject ($\bar{d}$=4.87).	1) Inquiry-Based Learning (5Es, 7Es, 9Es)	22
	2) Inquiry Method	9
	3) Constructionism	8
	4) Laboratory Method	8
	5) Learning Activity Set and Skill Set	35
	6) Programmed Learning and Online Learning	42
5) Creative Learning Activities and Situations Management Model (n=103) means organizing activities and situations/a learning model that allows students to act according to their imaginations. It will be effective in the primary school ($\bar{d}$=5.12) as a home economics subject ($\bar{d}$=12.05).	1) Learning Centers, Camps, and Gatherings	10
	2) Story Line Method	21
	3) Games Based Learning	13
	4) Learning Experience	14
	5) Art Activities	45

Propensity Score Matching

The propensity score matching was gained widespread attention and was applied in pervasively. (Rosenbaum & Rubin, 1983; Thoemmes & Kim, 2011), particularly in education and research, and evaluation (Hong & Raudenbush, 2005; Thoemmes & West, 2011) as a method for summarizing data for multiple covariates to reach a unified conclusion. Moreover, it is for determining the likelihood of an intervention for the variables using a probabilistic score calculated from the covariates (as in this research, they refer to attributes of each research such as gender of researcher, educational institution, educational level, instructional methods, research instruments, sample selection, including research methodology, etc.). The two significant analytical conditions of the propensity score matching include (Marshall & Paul, 1999): 1) Intervention depends on fundamental characteristics of the research; and 2) Every unit of the variable must have an opportunity to be used in an experimental study. The "propensity score" estimation relies on data analysis with the logistic regression model, which defines the dependent variable as intervention or treatment. Independent variables refer to related factors (Fitzmaurice, 2006). The analysis results will show the propensity score between 0 to 1, known as the "balancing score," used to see if the distribution of covariate characteristics is similar or not (Rosenbaum & Rubin, 1983). A propensity score is thereupon an alternative to assessing the likelihood of obtaining treatment according to the fundamental nature of the research. It is determined by the balance of the probabilistic score calculated from several common variables (Gray et al., 2017; Yu et al., 2017).

Nowadays, the application of propensity score matching to eliminate the effect of the extraneous variable for causation studies can be achieved using matching, also known as PSM (Fitzmaurice, 2006; Sturmer et al., 2006). It is a statistical method used to match experimental and control subjects using propensity scores of equal or similar value. Mostly, the researchers will apply a 1:1 matching method, which is the traditional method with the limited cases where the sample size is small or where the sample is difficult to find; therefore, 1:2, 1:3, 1:4, 1:many, or many: many matching methods can be substituted (Gray et al., 2017; Justus et al., 2014; Yu et al. , 2017) This pairing can help simplify cause-and-effect reasoning for the quasi experiment and observational bias in reducing publication bias and comparing propensity scores with independent t-tests or paired t-test. It can be concluded that propensity score matching is advantageous for controlling extraneous variables influencing dependent variables well (Thoemmes & West, 2011).

Methodology

Research Goals

This quantitative research emphasized three main objectives: 1) to analyze the propensity score of the research effect size for developing students' creative thinking; 2) to study the attribute variables effect of the research on the effect size of creative thinking before and after propensity score adjustment; and 3) to compare the effect size between instructional methods to develop creative thinking before and after the propensity score adjustments.

Sample and Data Collection

Examples of research include graduate theses and research articles on creative thinking development in Thailand. The researcher defined the keywords for searching as "creative thinking," "creative development," "teaching method," "teaching techniques," "teaching style," and "instructional management model." At first, 826 related research reports were discovered. Then the researches were selected according to four criteria: 1) it is a graduate thesis work, and published research articles can be searched from the ThaiLIS Digital Collection and Electronic libraries of each university; 2) It is experimental research with an experimental model designed to study the primary variable as instructional management style, and the dependent variable as creative thinking; 3) It is research reporting the necessary statistical values sufficient for effect size calculations (d) such as mean ($\bar{x}$), standard deviation (SD), number of the sample group,

p-value, t-value from the t-test, F-value from F-test; and 4) It is a research conducted in Thailand between 1971 and 2020. This research has four major research search and screening steps: 1) Identification, a total of 826 related subjects were found, and six duplicates were removed; 2) Screening, 24 inaccessible entries and 154 removed article entries; 3) Eligibility, 206 non-experimental studies and 36 subjects with incomplete statistics to calculate effects size; and 4) Included, 426 subjects that did not qualify for the meta-analysis study, and were excluded, and remained 400 suitable studies in this study. After that, 400 subjects on the effect size results of the instructional management model: were grouped based on mean, with 144 subjects as effect size high groups and 256 subjects as effect size low groups before conducting the propensity score matching analysis. As mentioned above, the research and data collection review are illustrated in Figure 1.
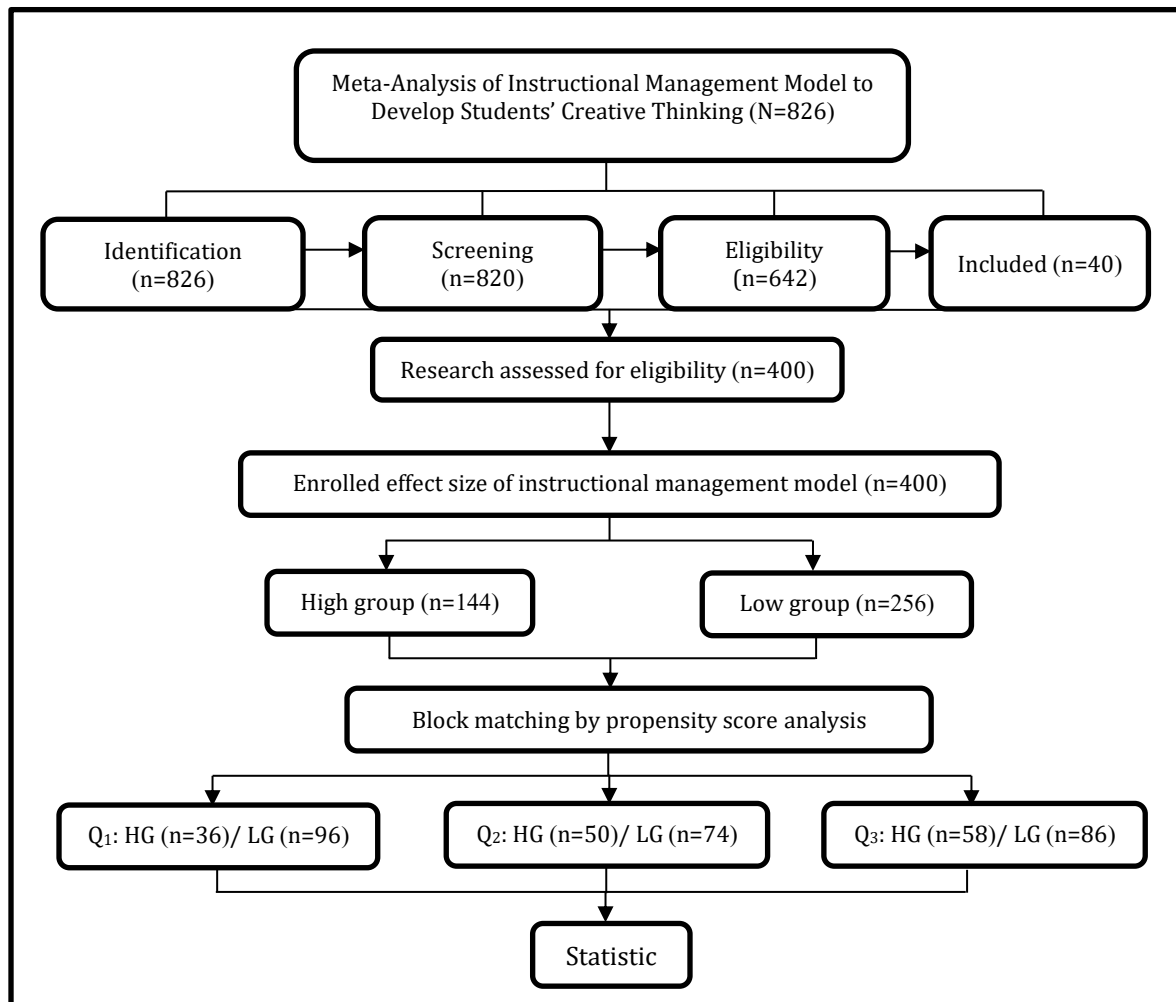

Figure 1. Procedures for Research Review and Data Collection

Research Reliability: This meta-analysis examined the reliability of the record form by a method of determining the concordance of record results between two independent evaluators in the research record. The codebook and coding manual was prepared, consisting of research characteristics in four areas: 1) research background data; 2) research content information; 3) research methodological data; and 4) research results. The coherence among the assessors was analyzed with Cohen's Kappa statistic (Lipsey & Wilson, 2001), with a reliability value of 0.98, indicating that the recording results were of good consistency (Card, 2012).

Research Accuracy: The researcher considered using the research quality variables to indicate the importance of the research process as part of the research characteristics variables by conducting research quality assessments using the five-rating scale form. It consists of 20 rubric criteria for assessment, covering seven parts of research quality content: 1) Background and significance of research problems, 2) Documents and related researches, 3) Research methods, 4) Results of data analysis, 5) Conclusion, discussion, and recommendations, 6) Presentation of the research report, and 7) Benefits of the research. The researchers examined the rubric's quality with the individual results of two independent recorders. They then analyzed them by Cohen's Kappa statistic for the consistency of the assessment results or the confidence among the assessors (Lipsey & Wilson, 2001). As a result, it had a reliability of 0.97, indicating that the rubric was consistent and concise and suitable for research assessment application (Card, 2012).

Analyzing of Data

Meta-Analysis of Instructional Management Model Developing Students' Creative Thinking

The researcher analyzed the effect size (d) by Cohen's computational method (1988) for the overall effect resulting from meta-analysis. It has been stated that Cohen's computational method can yield more accurate results than other methods (Lipsey & Wilson, 2001). The interpretation of effect size is d=0.20-0.50, meaning low level of effect, d=0.50-0.80 means a moderate level of effect, and d above 0.80 means a high level of effect.

Propensity Score Matching Analysis

The propensity score matching analysis is a statistical method for assessing the variable impact (research attributes) that affects the dependent variables (effect size). This research obtained the data from 400 experimental studies and worked on seven steps of the propensity score matching analysis: 1) To divide the effect size into two groups based on the mean of all the effect sizes ($\bar{d}$=3.48) as the separation criteria and define groups with the effect size more significant than average (high group=1); 2) To test for comparison of the instructional styles with the high effect size group and the low effect size group by independent t-test analysis to determine the difference in effect size between high and low-effect size groups; 3) To select attribute variables of research before analysis by comparing the research attribute variables with the high and low-effect size groups. It should be analyzed by using an independent t-test analysis to obtain the research attribute variables affecting the classification of high and low research opportunities; 4) To calculate the propensity scores with the logistic regression with the following equation.

$$\log\left(\frac{\Pr(y = 1|x)}{\Pr(y = 0|x)}\right) = a + \beta x$$

When x is the set of research attribute variables, Y is group member, where y = 1 as a high-effect size group and y = 0 as a low- effect size group.

This analysis results led each sample group to have the propensity scores predicted from a set of research attributes variables. They will have values in the range of 0 to 1 after analysis. Every sample group will have the propensity scores showing the probability of members of high and low effect size groups. The samples that were influenced by the same research attribute variables will have the same probability; 5) To check the distribution of each propensity score of the group for comparing the similar probability attributes in the manner of matching, which reduces the effect of research attribute variables. This is because it compares similar probability intervals between high and low groups; 6) To divide the propensity score into three ranges: Q_1 (between 0.00-0.33), Q_2 (between 0.33001-0.66), and Q_3 (between 0.66001-1.00); and 7) To verify the results of the adjustment of variables, followed by two-way ANOVA analysis, where the dependent variables were the research attribute variables while the primary variable was the members of high and low groups, and members in the three probability score groups (Q_1-Q_3). If the statistical differences of the extraneous factors were still found, step four should be performed repeatedly by adding the extraneous variables to the equation or adding the interactions of the variables to the equation. The probability score should be adjusted until it was found that none of the independent variables was statistically significant.

The Study of Effect of Instructional Management Model on Students' Creative Thinking

The meta-analysis of research effect size differences was tested by applying the random effect model in cases where the variance of the research effect size was different. The fixed effect model in cases where it was found that the variance of the research effect size was not different can be studied from 1) Omnibus test of model coefficients (Q_a); If the test results show that the Q_a value is statistically significant (p-value<.05), It means that the total effect size is different from zero. However, If Q_a is not statistically significant (p-value $\geq$ .05), the total effect size mean is not significantly different from zero; 2) According to the estimation of the effect size as zero, it is an interpretation according to Q_b with the statistical significance (p-value<.05), it indicates that there is a heterogeneity or has a non-zero residual. The random effect model estimation method should be chosen. In contrast, if Q_b is not statistically significant (p-value $\geq$ .05), there is no heterogeneity, or the residual is zero. The fixed effect model method should be chosen so that the estimation result is unbiased (Hedges & Vevea, 1998); 3) The Z value indicates that the effect size (d) differs from zero after adjustment. If the Z value is statistically significant (p-value<.05), the effect size differs from zero and has a positive or negative tendency towards the total mean of effect size. However, if the Z value does not have the statistical significance (p-value $\geq$ .05), the effect size is not different from zero and has no influence on the total mean effect size; 4) The τ^2 value explains the variance of the effect size (d) of each study. The interpretation can be considered in conjunction with the statistic of I^2, which represents the variance percentage level of the effect size for each research subject. It can be interpreted into three levels (Cooper et al., 2009) 25% referring to the low difference or no difference, 50% referring to the moderate level of difference, and 75% referring to the high level of difference; 5) Forest Plot is used to describe the effect of five different instructional models. The effect size means shown as a square box and 95% confidence interval. The size of the large box indicates a larger study sample size. If the confidence interval crosses the Line of on effect (value 0), the result is non-significant; and 6) Funnel plot has the test principle to create a scatter plot between the values representing the effect

size on the x-axis and the value representing the sample size on the y-axis considering the symmetry of the funnel plot. Suppose the data does not identify publication bias. In that case, it is found that the effect size in the mini research has the extensive dispersion data, and it is equally distributed around the mean effect size. However, if the dispersion data is asymmetric from the mean effect size, it can be concluded that publication bias is a problem.

As mentioned above, each part of the analysis was performed with JASP version 0.14.1.

Findings / Results

Finding the Propensity Scores of the Research Effect Size for Developing Students' Creative Thinking Analysis

Due to the propensity score matching technique requires a small amount of data in the analysis, and the data must have a similar number of matching groups, the researcher divided the research into two groups with different effect sizes with the mean effect size ($\bar{d}$=3.483) of the total research as the criterion for grouping. They included a high effect size group ($\bar{d}$>3.483) and a low effect size group ($\bar{d}$≤3.483). After that, the propensity score matching analyzed the data to answer the first objective.

The analysis of logistic regression yielded the propensity scores of the instructional model effect size. When taking the mean and standard deviation of the propensity scores for the low and high effect size groups, it was found that the propensity scores for the low effect size group consisted of 256 subjects ($\bar{d}$=1.345, SD=0.960), and the propensity scores for the high effect group had 144 subjects ($\bar{d}$=7.284, SD=0.277). The results of the analysis are detailed in Table 2.

Table 2. Mean and Standard Deviation of the Propensity Scores by Low and High Effect Size Groups

Group of Propensity Score	n	$\bar{d}$	SD
Low group	256	1.345	0.960
High group	144	7.284	0.277

Mean and standard deviation of the propensity score interval (Q_1- Q_3) were found that Q_1 of propensity score interval covered 132 subjects classified into 96 subjects of the low group ($\bar{d}$=0.105, SD=0.025) and 36 subjects of the high group ($\bar{d}$=0.031, SD=0.021). The propensity score interval of Q_2 included 124 subjects classified into 74 subjects of the low group ($\bar{d}$=0.432, SD=0.068) and the high group with 50 subjects ($\bar{d}$=0.210, SD=0.041). The propensity score interval of Q_3 had a totally of 144 subjects classified as the low group with 86 subjects ($\bar{d}$=0.684, SD=0.033) as the high group with 58 subjects ($\bar{d}$=0.582, SD=0.007). It can be seen in Table 3.

Table 3. Mean and Standard Deviation of the Propensity Scores (Q1 - Q3)

Quartiles Propensity Score	Group of Propensity Score	n	$\bar{d}$	SD
Q_1	Low group	96	0.105	0.025
	High group	36	0.031	0.021
Q_2	Low group	74	0.432	0.068
	High group	50	0.210	0.041
Q_3	Low group	86	0.684	0.033
	High group	58	0.582	0.007

When the research from Table 1 and Table 2 are distributed, the frequency (Y-axis) of the ascending probability (X-axis) is classified into high and low groups (between 0.000-1.000); it appears as a bar chart above-showing research frequency distributions with the low effect size. Therefore, the overall picture has a right-skewed distribution. Figure 2. The bar chart section below shows the research frequency distributions with the high effect size. Overall, there is a distribution of left-skew. This makes it possible to estimate the matching considerations for both research groups to achieve equality, as shown in Figure 2.
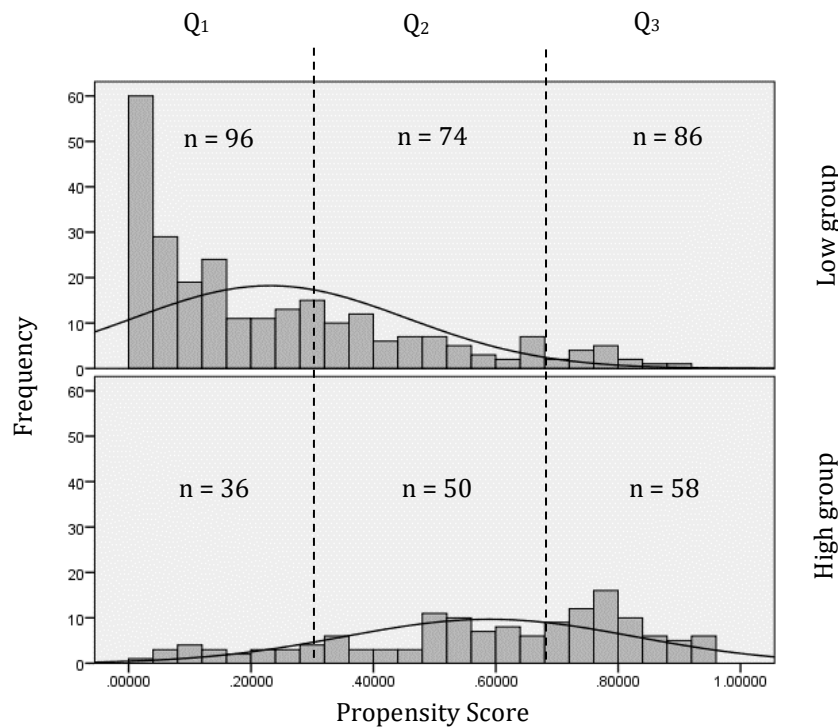
Figure 2. Comparison of the Propensity Scores Distributions for Low and High Effect Size Groups

Findings of the Study on the Attribute Variables Effect of the Research on the Effect Size of Creative Thinking Before and After the Propensity Score Analysis

26 attribute variables were studied for the effect on the effect size of creative thinking. The difference between the pre-adjusted and post-adjusted research attribute variables with the propensity scores revealed that the mean of research attribute variables was statistically significantly different at the .05 level with 13 variables. These variables included manufacturing research institutions, year of publication, educational institutions, curriculum, creative thinking indicators, instructional materials, types of research, research objectives, research groups, sample selection, research protocols, statistics used in research, and quality of research.

After the propensity score adjustments, it was found that the 12 variables of the research attribute variables are probable and different with no statistical significance at the .05 level. They were manufacturing research institutions, year of publication, educational institutions, curriculum, creative thinking indicators, instructional materials, types of research, research objectives, research groups, research protocols, statistics used in research, and quality of research. It was shown that the propensity scores eliminated the differences in the extraneous variables of the samples with low and high effect sizes. Details are as in Table 3

Table 3. Fundamental Statistics and Results of Comparison of Pre- and Post-Adjusted Effect Size with the Propensity Scores Matching Between Low and High Effect Sizes with Research Attribute Variables

Categorical Moderators	Effect size				Before Propensity Score Matching			After Propensity Score Matching		
	d	n	x̄	SD	t	p	F	p	F*	p
Gender of researcher	low	256	0.730	0.445	-1.914	0.050	1.147	0.285	0.017	0.895
	high	144	0.810	0.392						
Manufacturing research institutions	low	256	1.160	0.367	-4.888***	0.000	1.118	0.291	0.685	0.408
	high	144	1.389	0.489						
Field of research	low	256	3.144	1.391	1.352	0.177	0.230	0.632	0.923	0.337
	high	144	2.951	1.334						
Types of research	low	256	1.080	0.275	-2.378*	0.018	0.023	0.880	0.107	0.744
	high	144	1.170	0.374						
Year of publication	low	256	2.812	1.179	-7.687***	0.000	3.797	0.052	5.093	0.052
	high	144	3.604	0.862						
Educational institutions	low	256	1.492	0.501	2.431*	0.016	0.158	0.691	0.121	0.728
	high	144	1.368	0.484						

Table 3. Continued

Categorical Moderators		Effect size			Before Propensity Score Matching			After Propensity Score Matching		
	d	n	$\bar{x}$	SD	t	p	F	p	F*	p
Course provision	low	256	4.664	2.367	-1.287	0.199	0.149	0.699	0.643	0.423
	high	144	4.986	2.461						
Curriculum	low	256	2.812	1.180	-7.687***	0.000	3.797	0.052	5.093	0.052
	high	144	3.604	0.862						
Creative thinking indicators	low	256	2.156	1.207	2.708**	0.007	0.576	0.448	0.921	0.338
	high	144	1.826	1.099						
Creative thinking theories	low	256	1.855	1.273	-1.012	0.312	0.510	0.475	4.852	0.082
	high	144	1.986	1.176						
Educational level	low	256	2.270	1.190	1.870	0.062	0.115	0.734	0.002	0.969
	high	144	2.030	1.287						
Numbers of the lesson plan	low	256	2.101	1.073	1.148	0.251	0.002	0.968	0.989	0.321
	high	144	1.972	1.096						
Numbers of instructional weeks	low	256	2.191	1.009	-0.805	0.421	0.007	0.935	0.045	0.833
	high	144	2.278	1.067						
Amount of time per lesson plan	low	256	2.172	0.752	-0.581	0.562	0.013	0.908	0.167	0.683
	high	144	2.222	0.872						
Total instructional time	low	256	2.172	0.877	-1.863	0.063	0.223	0.637	1.310	0.253
	high	144	2.347	0.948						
Learning process	low	256	3.023	2.800	-1.220	0.223	0.014	0.908	0.501	0.479
	high	144	3.403	3.082						
Evaluation	low	256	1.359	0.804	1.450	0.148	1.182	0.278	2.625	0.106
	high	144	1.243	0.750						
Instructional materials	low	256	3.160	1.247	-1.978*	0.049	0.000	0.992	0.049	0.824
	high	144	3.390	1.025						
Learning resources	low	256	1.031	0.174	1.185	0.237	0.303	0.582	0.065	0.799
	high	144	1.013	0.117						
Research objectives	low	256	1.851	0.788	-4.552***	0.000	0.801	0.371	4.850	0.028
	high	144	2.222	0.770						
Research groups	low	256	1.690	0.590	10.447***	0.000	4.505	0.034*	2.706	0.011*
	high	144	1.170	0.410						
Sample selection	low	256	2.312	1.467	-2.594**	0.010	0.494	0.483	9.979	0.002
	high	144	2.680	1.299						
Research protocols	low	256	1.437	0.563	-9.085***	0.000	0.001	0.970	0.411	0.522
	high	144	1.861	0.366						
Quality of instruments	low	256	1.676	0.831	-0.765	0.445	0.006	0.940	3.351	0.068
	high	144	1.743	0.867						
Statistics used in research	low	256	1.789	0.658	11.836***	0.000	2.292	0.131	2.292	0.131
	high	144	1.159	0.404						
Quality of research	low	256	4.140	0.390	4.212***	0.000	1.727	0.190	1.578	0.210
	high	144	3.993	0.301						

Note: F*=The statistical value of the test for the effect of interaction between the research attribute variables and the propensity score group

Findings of the Instructional Models in Developing Students' Creative Thinking Effect for Pre- and Post-Adjusted the Propensity Score Comparison

The results of the effect size of the instructional management model on the students' creative thinking before propensity score matching by random effects model analysis revealed that the effect size differed from zero (Q_a=8.68). The residual was not zero (Q_b=34134.92***), the variance of the effect size was high (τ^2=10.83), and the percentage variance of the effect size of the influence size was high (I^2=99.21%). It showed that each research subject had a statistically significant difference in influence on the creative thinking and effect of those researches. Moreover, it can be revealed that each group of instructional management methods had a different influence on the effect size and differed according to the research attributes.

When considering only the effect of instructional management models on creative thinking development, it was shown that the three instructional management models had statistically significantly different effect sizes from zero. The model

with the average effect size in descending order was the creative learning activities management group ($\bar{d}$=3.88, SE=0.37, [3.16, 4.62]), self-knowledge construction with the technology media group ($\bar{d}$=3.69, SE=0.29, [3.12, 4.27]), and all-rounded creative thinking development group ($\bar{d}$=2.64, SE=0.35, [1.95, 3.34]).

The findings analysis of the effect size of the instructional management model developing the students' creative thinking after the propensity score matching by fixed effects model were shown that the effect size did not differ from zero (Q_a=7.51) and had the residual as zero (Q_b=294.48) without the variance of the effect size. Therefore, those research had no statistically significant difference, and then it can be concluded that the difference in the effect size resulted only from instructional management models.

The study results on the effect of instructional management models on creative thinking development found that the two instructional models had statistically different effect sizes from zero. The instructional models were varied with the average effect size in descending order as) the knowledge integration-based media and technology model group ($\bar{d}$=0.41, SE=0.03, [0.35, 0.48]), and all-rounded creative thinking development group ($\bar{d}$=0.29, SE=0.03, [0.24, 0.36]). Details are shown in Table 4 and Figure 3.

Table 4. The Results of the Effect Size of Instructional Methods on the Development of Creativity Before and After the Propensity Score Matching Analysis

Instructional Management Models	Before Propensity Score Matching							After Propensity Score Matching						
	k	$\bar{d}$	SE	95%CI	z	Q_a	Q_b	k	$\bar{d}$	SE	95%CI	z	Q_a	Q_b
1. All-Rounded Creative Thinking Development Group	80	2.64	0.35	[1.95, 3.34]	7.13***	8.68	34134.92***	80	0.29	0.03	[0.24,0.36]	9.34***	7.51	294.48
2. Creative Thinking Skills Stimulation Group	25	2.58	0.55	[1.50, 3.67]	-0.07			25	0.32	0.06	[0.21,0.45]	0.44		
3. Knowledge Integration based on Media and Technology Group	68	3.52	0.38	[2.77, 4.29]	1.62			68	0.41	0.03	[0.35,0.48]	2.44**		
4. Self-Knowledge Construction with the Technology and Media Group	124	3.69	0.29	[3.12, 4.27]	2.20*			124	0.31	0.02	[0.26,0.36]	0.29		
5. Creative Learning Activities Management Group	103	3.88	0.37	[3.16, 4.62]	2.46*			103	0.31	0.03	[0.26,0.38]	0.47		
Residual Heterogeneity Estimates: τ^2= 10.83, I^2 = 99.21%														

Note: Q_a=Omnibus test of model coefficients, Q_b=Test of residual heterogeneity. *p<.05, **p<.01, ***p<.001

Uncombined Findings of Effect Size Analysis in Accordance with Creative Variable

The results of considering the forest plot for describing the mean effect size of each instructional models before and after the propensity score analysis they described as square boxes. 95% of the reliability interval revealed that before adjustment based on the propensity scores, the instructional model with the effect size affected the total mean effect size statistically significant (p<.05). Considering the effect size of instructional management models in descending order, they were creative learning activities management group, self-knowledge construction with the technology and media group, and all-rounded creative thinking development group. After the propensity scores adjustment, the instructional model with the effect size affected the total mean effect size statistically significant (p<.05). Considering the effect size of instructional management models in descending order, they were integrated knowledge-based on technology and media group and all-rounded creative thinking development group. Therefore, it can be summarized that the propensity score matching analysis could indicate the two instructional models that are suitable for developing the students' creative thinking. Details are described in Figure 3.
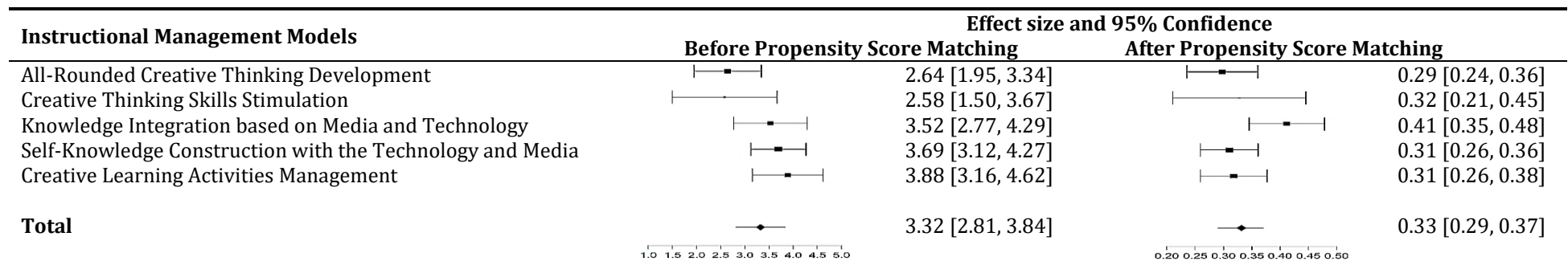
Instructional Management Models	Effect size and 95% Confidence	
	Before Propensity Score Matching	After Propensity Score Matching
All-Rounded Creative Thinking Development	2.64 [1.95, 3.34]	0.29 [0.24, 0.36]
Creative Thinking Skills Stimulation	2.58 [1.50, 3.67]	0.32 [0.21, 0.45]
Knowledge Integration based on Media and Technology	3.52 [2.77, 4.29]	0.41 [0.35, 0.48]
Self-Knowledge Construction with the Technology and Media	3.69 [3.12, 4.27]	0.31 [0.26, 0.36]
Creative Learning Activities Management	3.88 [3.16, 4.62]	0.31 [0.26, 0.38]
Total	3.32 [2.81, 3.84]	0.33 [0.29, 0.37]

Figure 3. Forest Plot of Effects Sizes of Studies by the Variable of Instructional Management Model

Before and After the Propensity Score Matching

In addition, the results of considering publication bias from the dispersion of the data with funnel plot before propensity score matching (Figure 4) showed that most of the research effect sizes were positively distributed within the triangle, and some were outside the triangular frame. Furthermore, the data were asymmetrically distributed from the total mean effect size. It shows that the instructional management model has a high effect on publication bias. After the propensity score matching (Figure 5), it was found that most of the effect sizes were positively distributed within the triangle and symmetrically distributed close to the center line, that was the total mean effect size. This indicates that the instructional models have a low effect on publication bias.
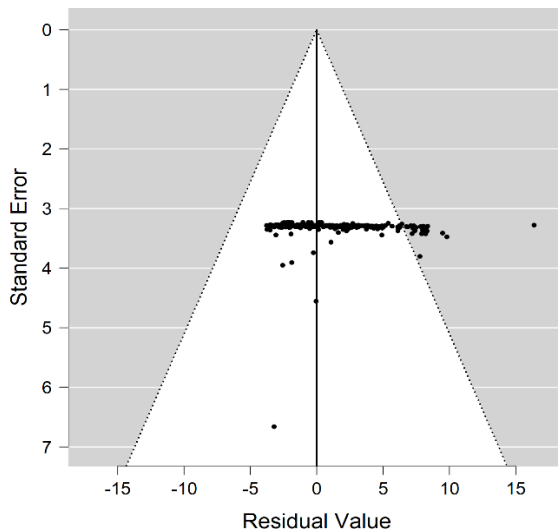
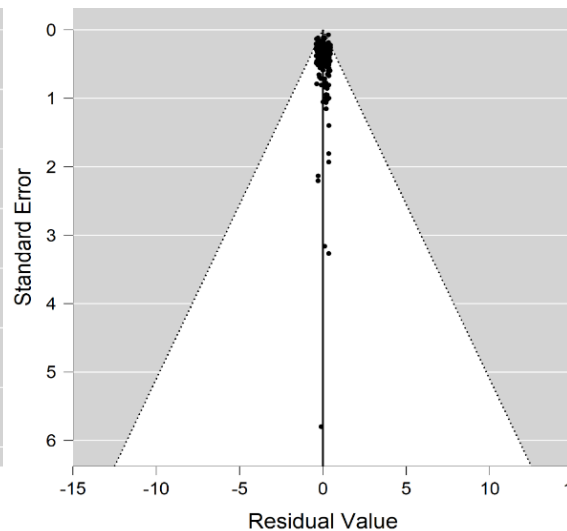

Figure 4. Before Propensity Score Matching   *Figure 5. After Propensity Score Matching*

Discussion

The mentioned research findings show that the propensity score matching analysis plays an essential role in meta-analysis. As a result, the assumption in the overall picture expects to see a change in suitable methods for creative thinking development from all existing research. After adjusting the effect size value, the researcher found three significant issues for discussion. The first issue is the analysis of the propensity scores to develop the students' creative thinking by utilizing the mean effect size ($\bar{d}$=3.483) as the criterion for categorizing the high and low effect groups. Estimated by propensity score matching, where the dependent variable is the effect size of the instructional management model (n=400). As the independent variable is an attribute of research related to the instructional management model consisting of 26 variables, these attribute variables satisfy the conditions of the propensity score matching analysis, i.e., 1) Intervention depends on the basic attributes of the research; 2) Every unit of the variable must have an opportunity to take part in experimental research (probability must not be zero) since all research attribute variables are significant to the creative thinking development (Dani et al., 2018; Marshall & Paul, 1999; Ulger, 2019). It can be considered from several common variables, having a balancing score, and the distribution of the characteristics of the common variables as being similar (Rosenbaum & Rubin, 1983). However, it was found that the logistic regression analysis results in higher propensity scores for the low effect group (n=256) than for the high effect group (n=144). This might be because most of the effect size was lower than the mean effect size. Additionally, in the analysis of propensity scores, the amount of data in the analysis is not too small, and the data must be similar. When the data were compared, the distribution of the propensity scores for the low and high effect size groups shows that the data is distributed similarly in both groups, thus enabling the matching among the two sample groups to compare to achieve equality.

The second issue is that according to the study of the effect of research attribute variables on the creative thinking effect size between before and after the propensity scores adjustments led, the researcher to be somewhat surprised. After propensity score matching analysis, it resulted in controlling the effect of more than 12 variables, indicating that the propensity score eliminated the differences in the extraneous variables of the samples with low and high effect sizes. Nevertheless, the researcher found an essential observation that only variables in the group used in the research were still high and statistically significant at the .05 level (p-value=0.034), indicating that such variables still affected the effect size. This may be because the research group variable in each group differed significantly from the sample size. The effect size was very high, and it was statistically different from zero at the .001 level. According to the results, it may be linked to critical issues of research quality. Most of the research had the quality at good level (n=343). This may be due to the lack of coverage of research in many areas. The methodology was inconsistent with the initial agreement, selection of appropriate samples for research, research process, and analysis of results. Moreover, most of the research lacked rigorous internal validity, such as randomization, blinding, protocol adherence, and external validity. Such problems may

cause the publication bias contained in this research which was consistent with Alinaghi and Reed (2018), Bom and Rachinger (2019) argue that high-risk patients tend to be high-risk patients, resulting in poor methodological quality). It may be due to the low cost of research and a tendency to conduct research that is not rigorous. Therefore, the cause of publication bias may be due to different research values (Borenstein, 2009).

The third issue is comparing the effect size on instructional methods that developed creative thinking between before and after adjustment with the propensity scores. This led the researcher to the important findings in line with the assumptions set because the instructional model for creative thinking development had changed from the past. Nevertheless, the finding analysis of the random effect model revealed that the effect size differed from zero (Q_a=8.68) and has residual as zero (Q_b=34134.92***). The variance of the effect size was high (τ^2=10.83). The percentage variance of the effect size was high (I^2=99.21%), showing that each study had a statistically significant difference in the effect of creative thinking and the effect of those studies. This makes the conclusions of the research still not reliable ($\bar{d}$=3.88). After adjustment of the propensity scores, it was found that the knowledge integration-based media and technology model had the greatest positive effect on creative thinking development ($\bar{d}$=0.41). The fixed effects model analysis indicated that the effect size was not different from zero (Q_a=7.51) and had a residual of zero (Q_b=294.48). No variance in the effect size was found and showed that there were no statistically significant differences between those studies. From such research results, it can be confirmed that the knowledge integration-based on media and technology model resulted in the best effect on the students' creative thinking development. This is because the instructional model is characterized by instructional management that emphasizes integrating knowledge in the subject content and connecting with real life and applying media and technology to promote the students' creative thinking. According to the preliminary statistical survey, it was found that the knowledge integration-based on media and technology model had a positive effect on instruction at the primary school level ($\bar{d}$=4.54) for the art ($\bar{d}$=5.06) course and followed the principles of Torrance's creative thinking development theory approach ($\bar{d}$=2.77). Creativity skills were measured in all Torrance: originality, fluency, flexibility, elaboration ($\bar{d}$=5.29). There were one to six instructional management plans to apply ($\bar{d}$=4.08). Duration per plan was less than one hour ($\bar{d}$=3.68). They were taught for more than four weeks ($\bar{d}$=4.21). Moreover, the creative thinking development duration included more than 31 hours ($\bar{d}$=4.51). The instructional processes employed leading, instruction, and conclusion ($\bar{d}$=2.98). Evaluation and assessment were test ($\bar{d}$=3.90). Instructional materials were textbooks ($\bar{d}$=3.34). The classroom was the main location ($\bar{d}$=3.77). This aligns with the creative thinking development theory that focuses on enhancing the brain's ability to see the relationship of things around them, resulting in learning and understanding. Then, it becomes a reaction to an imaginary thought leading to inventing new things to meet the needs or solve problems in daily life (Dani et al., 2018; Royston & Reiter-Palmon, 2019; Taylor & Kaufman, 2021). This is supported by behaviorism theory: creative thinking is a behavior that results from learning by emphasizing the importance of positive reinforcement and responses to specific stimuli or situations (Moore, 2011). It is also the brain process in which the relationship with the surrounding stimuli causes new ideas or the creation of new things (Algahtani, 2017). Moreover, it is also in line with humanism theory, explaining that creative thinking is inherently human. The only person who can bring out creative thinking has self–actualization and is himself completely, has independent thinking, self-awareness, and self-satisfaction, and uses his abilities to the fullest potential. To be able to express creative thinking must depend on favorable conditions or atmosphere, or a creative atmosphere that results in psychological safety and stability of the mind, desire to crystalize with ideas, and to be open to new experiences (Soudien, 2019), so that the development of creative thinking can cultivate and foster higher creative thinking with an appropriate instructional management model (Bush, 1978; Knowles et al., 1998; William, 1979).

Conclusion

According to this research, the researcher summarized the findings of three significant findings according to the objectives as follows: 1) Analysis of the research propensity scores to develop the students' creative thinking: the results of the logistic regression analysis resulted in propensity scores for 256 subjects with low effect size ($\bar{d}$=1.345, SD=0.960) and 144 subjects with high effect size ($\bar{d}$=7.284, SD=0.277). The propensity score range was divided into three periods (Q_1-Q_3). When comparing the distributions of propensity scores for both groups, there were the same distributions of 0.00 to 1.00 scores for both groups. Thus, the propensity score matching method can be applied; 2) The study of the attribute variables effect of the research on the effect size of creative thinking before and after the propensity score adjustment: It was revealed that after the propensity score adjustments, twelve variables with the statistical significance at the .05 level were probabilistic for the attributes of the research. The variables included manufacturing research institutions, year of publication, educational institutions, curriculum, creative thinking indicators, instructional materials, types of research, research objectives, research groups, research protocols, statistics used in research, quality of research; and 3) The comparison of the effect size between instructional methods to develop creative thinking before and after before and after propensity score adjustments: it showed that the knowledge integration-based on media and technology model has the best effect on the students' creative thinking development. Therefore, it can be concluded that the propensity scores can help to eliminate the effect of research attributes variables from instructional models improving the students' creative thinking.

Recommendations

The research mentioned above has recommended three crucial issues for innovation development to promote the students' creative thinking. The first issue is that the knowledge integration-based on media and technology model should be designed to emphasize connecting knowledge of subject content to the real-life or surrounding environment of learners. An environment is provided to encourage creative learning. Media and technology are used to assist in instructional activities such as STEM education instruction, STEAM education instruction, syndetic instruction, interdisciplinary instruction, blended instruction, etc. Such an instructional model will reflect the benefits of the instruction at the primary school level for the art. The teachers should follow the principles of Torrance's creative thinking development theory approach to measure four components: originality, fluency, flexibility, elaboration. Six instructional management plans should be taught. Duration per plan should be less than one hour and should be taught in more than four weeks. Moreover, instructional processes should include leading, instruction, and conclusion. The test should be used for evaluation and assessment. Instructional materials are textbooks along with new media and technologies. Moreover, the classrooms should be managed creatively. The second issue, there should be to study the research attributes variables that are advantageous for developing creative thinking in instructional management or designing additional learning activities such as learning management plans, the process of organizing learning activities, measuring creative thinking skills, number of quizzes, etc. to benefit research and teachers who are interested in developing the students' creative thinking in their context. The last issue is that the propensity score matching should be a useful tool for meta-analysis since it can help control the extraneous variables well. It can eliminate the extraneous variables that affect the research results. Moreover, it also helps in reducing publication bias and reducing Type I Error. As a result, applying the propensity score matching is very suitable for the research type. Meta-analysis leas the research results are complete Therefore, the researcher supports the idea of applying the propensity score adjustment in controlling extraneous variables before analyzing by meta-analysis.

Limitations

The application of the propensity score matching as a tool to support analytical studies led to some issues that were eliminated: 1) The propensity score matching only prevents bias from important variables related to creative thinking. If unrelated extraneous variables were analyzed for the propensity scores, the study results would be inaccurate, and 2) The propensity score matching is applicable where the distribution of propensity scores is equal or similar between experimental and control groups. If this is not the case, the propensity scores cannot remove their effect from the study results.

Authorship Contribution Statement

Itsarangkul: Conceptualization, design, data acquisition, data analysis/interpretation, statistical analysis, drafting manuscript, critical revision of manuscript, and writing. Damrongpanit: Consulting research, data analysis, statistical analysis, editing/reviewing, supervision, final approval, technical or material support, and securing funding.

References

Algahtani, F. (2017). Teaching students with intellectual disabilities: Constructivism or behaviorism? *Educational Research and Reviews, 12*(21), 1031-1035. https://doi.org/10.5897/ERR2017.3366

Alinaghi, N., & Reed, W. R. (2018). Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? *Research Synthesis Methods, 9*(2), 285-311. https://doi.org/10.1002/jrsm.1298

Arikan, S., Vijver, F. V. d., & Yagmur, K. (2018). Propensity score matching helps to understand sources of DIF and mathematics performance differences of Indonesian, Turkish, Australian, and Dutch students in PISA. *International Journal of Research in Education and Science, 4*(1), 69-81. https://doi.org/10.21890/ijres.382936

Bom, P. R. D., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods, 10*(4), 497-514. https://doi.org/10.1002/jrsm.1352

Borenstein, M. (2009). *Introduction to meta-analysis*. Biostat, Inc. https://doi.org/10.1002/9780470743386

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Longman. https://bit.ly/3AEE2vy

Bush, M. (1978). Preliminary considerations for a psychoanalytic theory of insight: Historical perspective. *International Review of Psycho-Analysis, 5*(1), 1-13. https://bit.ly/3P28XGn

Card, N. A. (2012). *Applied meta-analysis for social science research*. A Division of Guilford Publication, Inc. https://doi.org/10.1080/10705511.2013.824795

Ciğerci, F. M. (2020). Primary school teacher candidates and 21st century skills. *International Journal of Progressive Education, 16*(2), 157-174. https://doi.org/10.29329/ijpe.2020.241.11

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://bit.ly/3A8r1cS

Cooper, S., Endacott, R., & Chapman, Y. (2009). Qualitative research: Specific designs for qualitative research in emergency care. *Emergency Medicine Journal, 26*(11), 773-776. https://doi.org/10.1136/emj.2008.071159

Dani, D., Hallman-Thrasher, A., & Litchfield, E. (2018). Creative assessments. *Science Teacher, 85*(5), 46-53. https://doi.org/10.2505/4/tst18_085_05_46

Dowdy, A., Tincani, M., & Schneider, W. J. (2020). Evaluation of publication bias in response interruption and redirection: A meta-analysis. *Journal of Applied Behavior Analysis, 53*(4), 2151-2171. https://doi.org/10.1002/jaba.724

Fitzmaurice, G. (2006). Confounding: Propensity score adjustment. *Nutrition, 22*(11-12), 1214-1216. https://doi.org/10.1016/j.nut.2006.08.015

Gajda, A., Beghetto, R. A., & Karwowski, M. (2017). Exploring creative learning in the classroom: A multi-method approach. *Thinking Skills and Creativity, 24*, 250–267. https://doi.org/10.1016/j.tsc.2017.04.002

Gray, E., Pasta, D. J., Norris, S., & O'Leary, A. (2017). Effectiveness of triple therapy with direct-acting antivirals for hepatitis C genotype 1 infection: Application of propensity score matching in a national HCV treatment registry. *BMC Health Services Research, 17*(1), Article 288. https://doi.org/10.1186/s12913-017-2188-1

Hedges, L. V., & Vevea, J. L. (1998). Fixed and random effects models in meta-analysis. *Psychological Methods, 3*(1), 486-504. https://doi.org/10.1037/1082-989X.3.4.486

Herdem, D. Ö. (2019). A comparison of self-leadership characteristics of the students of department of fine arts and the others "The case of Gazi University". *Universal Journal of Educational Research, 7*(1), 198-205. https://doi.org/10.13189/ujer.2019.070125

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(1), 205-224. https://doi.org/10.3102/01623737027003205

In'am, A., & Sutrisno, E. S. (2021). Strengthening students' self-efficacy and motivation in learning mathematics through the cooperative learning model. *International Journal of Instruction, 14*(1), 395-410. https://doi.org/10.29333/iji.2021.14123a

Justus, J. R., Falbe, K., Manuel, A. K., & Joseph, L. B. (2014). A step-by-step guide to propensity score matching in R. *Practical Assessment, Research & Evaluation, 19*(18), Article 18. https://doi.org/10.7275/n3pv-tx27

Karatas, K., & Zeybek, G. (2020). The role of the academic field in the relationship between self-directed learning and 21st century skills. *Bulletin of Education and Research, 42*(2), 33-52. https://bit.ly/3nP9y2c

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education, 16*(2-3), 57-66. https://doi.org/10.4219/jsge-2005-473

Kirikkaleli, D., Ertugrul, H. M., Sari, A., Ozun, A., & Kiral, H. (2021). Quality of education and technological readiness: Bootstrap panel causality analysis for Northern European countries. *Scandinavian Journal of Educational Research, 65*(2), 276–287. https://doi.org/10.1080/00313831.2019.1705892

Knowles, M. S., Holton, E. F., & Swanson, R. A. (1998). *The adult learner: The definitive classic in adult education and human resource development*. Gulf Publishing. https://doi.org/10.1007/978-1-4419-1428-6_1022

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. SAGE. http://www.gbv.de/dms/ilmenau/toc/59825062X.PDF

Marshall, M. J., & Paul, R. R. (1999). Invited commentary: Propensity score. *American Journal of Epidemiology, 150*(4), 327-333. https://doi.org/gfgrh2

Moore, J. (2011). Behaviorism. *Psychological Record, 61*(3), 449-463. https://doi.org/10.1007/BF03395771

National Institute of Educational Testing Service. (2020). *Ordinary national educational test 2020*. https://bit.ly/3Cgnqcq

Özyurt, H., & Özyurt, Ö. (2020). Reflections of design-based research approach on learning experience of visual programming course. *Journal of Pedagogical Research, 4*(1), 12-21. https://doi.org/10.33902/JPR.2020057981

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Royston, R., & Reiter-Palmon, R. (2019). Creative self-efficacy as mediator between creative mindsets and creative problem-solving. *Journal of Creative Behavior, 53(*4), 472-481. https://doi.org/10.1002/jocb.226

Shah, B., & Gustafsson, E. (2021). Exploring the effects of age, gender, and school setting on children's creative thinking skills. *Journal of Creative Behavior, 55*(2),546-553. https://doi.org/10.1002/jocb.480

Soudien, C. (2019). The significance of new humanism for education and development. *Prospects: Quarterly Review of Comparative Education, 47*(4), 309-320. https://doi.org/10.1007/s11125-018-9440-2

Sturmer, T., Joshi, M., Glynn, R. J., Avorn, J., Rothman, K. J., & Schneeweiss, S. (2006). A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of Clinical Epidemiology, 59*(5), 437-447. https://doi.org/10.1016/j.jclinepi.2005.07.004

Taylor, C. L., & Kaufman, J. C. (2021). Values across creative domains. *Journal of Creative Behavior, 55*(2), 501-516. https://doi.org/10.1002/jocb.470

Thoemmes, F., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research, 46*(1), 90-118. https://doi.org/10.1080/00273171.2011.540475

Thoemmes, F. J., & West, S. G. (2011). The use of propensity scores for nonrandomized designs with clustered data. *Multivariate Behavioral Research, 46*(3), 514–543. https://doi.org/10.1080/00273171.2011.569395

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons. https://bit.ly/3Rkas4i

Ulger, K. (2019). Comparing the effects of art education and science education on creative thinking in high school students. *Arts Education Policy Review, 120*(2), 57-79. https://doi.org/10.1080/10632913.2017.1334612

Wang, G. (2020). On the application of cooperative learning in college English teaching. *International Education Studies, 13*(6), 62-66. https://doi.org/10.5539/ies.v13n6p62

William, W. R. (1979). On behavioral theories of reference. *Philosophy of Science, 46*(2), 175-203. https://doi.org/10.1086/288861

Yu, H. S., Hwang, J. E., Chung, H. S., Cho, Y. H., Kim, M. S., Hwang, E. C., Oh, K. J., Kim, S. O., Jung, I. S., Kang, T. W., Kwon, D. D., Park, K., Ryu, S. B., Jung, S. H., Hur, Y. H., Noh, J. H., Kim, M. K., Seo, I. Y., Kim, C. S., . . . Cheon, J. (2017). Is preoperative chronic kidney disease status associated with oncologic outcomes in upper urinary tract urothelial carcinoma? A multicenter propensity score-matched analysis. *Oncotarget, 8*(39), 66540-66549. https://doi.org/10.18632/oncotarget.16239