MEASURING THE EFFECTIVENESS OF UNIVERSITY PROGRAMMES BASED ON EVALUATION MODELS: A META-ANALYSIS

Pengukuran Keberkesanan Program Universiti Berdasarkan Model Penilaian: Satu Meta Analisis

*Ranita Manap[a], Norziah Othman[a],
Siti Nurashidah Roslan[a],Khairunnisa Ismail[a],
Aimi Fadzirul Kamarubahrin[b]

[a]Kolej Universiti Islam Antarabangsa Selangor
[b]Universiti Sains Islam Malaysia

*ranita@kuis.edu.my

Abstract

This article presents an extensive systematic review and meta-analysis on the effectiveness of university programmes based on different evaluation models. This study was conducted on primary sources published between year 2000 to 2018 using systematic searches from online databases including EBSCOhost, CENGAGE Learning, Cochrane Library, Clinical Key, ProQuest, ScienceDirect, and Scopus. There were, twenty-eight studies found on measuring the effectiveness of university programmes using different evaluation models. From these studies, the most common evaluation models used are the Context Input Process Product (CIPP) model and the Kirkpatrick model. It is concluded that the effectiveness of a programme needs to be improved through the implementation of optimising training designs, redefining training roles, management commitment, attention to individual and use of ongoing evaluation.

Keywords: evaluation, model, effectiveness, programme.

Abstrak

Artikel ini membentangkan kajian sistematik terperinci dan meta-analisis mengenai keberkesanan program universiti berdasarkan model penilaian yang berbeza. Kajian ini dijalankan berdasarkan sumber-sumber yang diterbitkan di antara tahun 2000 hingga 2018 dengan menggunakan carian sistematik dari pangkalan data dalam talian termasuk *EBSCOhost, CENGAGE Learning, Cochrane Library, Clinical Key, ProQuest, ScienceDirect,* dan *Scopus*. Terdapat dua puluh lapan kajian yang mengukur keberkesanan program universiti menggunakan model penilaian yang berbeza. Hasil kajian menunjukkan, model penilaian yang paling biasa digunakan ialah model Konteks Input Proses Produk (KIPP) dan model Kirkpatrick. Kesimpulannya, keberkesanan program perlu diperbaiki menerusi mengoptimumkan reka bentuk latihan, mentakrifkan semula peranan latihan, komitmen pengurusan, perhatian kepada individu dan penggunaan penilaian yang berterusan.

Kata kunci: penilaian, model, keberkesanan, program.

INTRODUCTION

Globalization, liberalization and information technology have necessitated the development and empowerment of higher education to cultivate human resources who are holistically educated, well-informed, adequately skilled and well-rounded. Education programmes has been undergoing continuous improvement ever since the idea of programme evaluation effectiveness was introduced (Barr et al., 2006). Evaluation programs are often seen as an obligation for institutions that wish to identify the benefits and weaknesses of the implemented activities.

The evaluation model of programme effectiveness especially in the universities is the measurement of student's knowledge, skills and behavioral patterns as a result of the programme provided. Evaluation programme should be conducted in a thorough and sustainable way, involving assessment on the learning process and outcomes. One of important factors that contribute to the achievement of educational objectives is the learning process itself. Furthermore, on-going evaluation and assessment (on the learning process and the outcomes) also play a role in encouraging the teaching staffs to improve the quality of learning process.

In fact, evaluation is one of the main components in education system. Evaluation provides not only a description or information on the students' achievement or mastery of the learnt materials, but also a feedback to the educational programme itself. Programme effectiveness evaluation is conducted as part of a decision-making process when it comes to the students' mastery of the materials after they are engaged in the teaching-learning process. In addition, the evaluation model is also useful to determine whether the programme strategy or approach is appropriate.

In order to realize how evaluation model as an effective pedagogical tool can significantly contribute in measuring the effectiveness of a university program, an exploration of its implementation needs to be carried out. Thus, one important aspect is to understand the underlying factors that contribute to a successful pairing formation. This research aims to explore the practice of using an evaluation model of programme effectiveness in a university. The researchers applied meta-analysis (systematic literature review) in assessing the existing evaluation models of programme effectiveness in universities. The key contribution of this paper is to summarise the findings of empirical studies on programme effectiveness in universities by using meta-analysis.

Researchers present the meta-analysis results by investigating evidence of recurring patterns that can be used to understand the current state-of-the-art of research in evaluation model when applied in programme effectiveness in university context. The researchers carry out a meta-analysis (Stanley 2001), to observe the different factors which can explain the variations in the results of these studies. To do this, the researchers select 28 studies, 4 evaluation model which were carried out in different countries, focusing on the programme effectiveness in universities. This article is organized as follows. Section 2 presents a debate in the literature on the evaluation model of programme effectiveness in universities. In section 3, the methodology of the study is presented used. Section 4 presents the model and estimation method. Section 5 presents the results and interpretation. Finally, section 6 presents the conclusion.

LITERATURE REVIEW
Evaluation
Evaluation is a continuous process which serves as the basis for all activities in a good learning process. Evaluation refers to assessment of a specified programme. It is a systematic process for gathering and interpreting information in order to assess the implementation of objectives (Fitzpatrick et al., 2004). In general, evaluation is defined as a systematic process to determine the value of something (objective, activity, decision, performance, process, individual, or even an object) based on certain criteria. In the learning context, evaluation is defined as a systematic process to determine the achievement level based on the specified learning objectives. The National Study Committee on Evaluation (Stark and Thomas,1994) suggested that evaluation is the process of ascertaining the decision of concern, selecting appropriate information, and collecting and analysing information in order to report summary data useful to decision makers in selecting among alternatives. This is confirmed by Griffin and Nix (1991) suggesting that measurement, assessment, and evaluation are hierarchical. The comparison of observation with the criteria is the

measurement; the interpretation and description of the evidence is the assessment and the judgment of the value or implication of the behaviour is the evaluation.

Evaluation is an integral part of activities in university programmes. A lecturer can apply a variety of evaluation methods to assess the students' level of achievement. These include tests, assignments, oral questions, observations during teaching-learning session, and maintaining a portfolio. The activities are conducted not only to determine the students' grades but also to improve the quality of a university programme. Meanwhile, programme evaluation is defined as a process of gathering information for the purposes of programme planning (McNamara, 1998). According to Fitzpatrick et al., (2004) one's view on valuation produce a direct impact on the type of evaluation activity made, either for curriculum evaluation, corporate training programme or others. Furthermore, Brinkerhoff (1989) stated that the evaluation of the implemented programme should demonstrate the effectiveness of the programme and also the significant relationship between training with skills development, human performance improvement and organisational effectiveness. Brinkerhoff et al., (1983) stated that evaluation has various meanings and definitions, depending on the purpose of the evaluation. This diverse definition is also derived from the different standards used when evaluating. According to Tyler (1950), evaluation is a process of determining the extent to which the objectives of a programme can be achieved. In addition, Alkin (1969), Cronbach (1963) and Stufflebeam et al., (1971) agreed that evaluation is the process of gathering and using information to make decisions. Meanwhile, Suchman (1967) stated that evaluation is an attempt to identify whether a programme carried out within an organisation or institution succeeds in achieving its goals and objectives.

Maradapi (2000) suggested that there are seven elements of learning evaluation; 1) focusing the evaluation, 2) designing the evaluation, 3) collecting information, 4) analyzing and interpreting, 5) reporting information, 6) managing evaluation, and 7) evaluating evaluation. The definition shows that in the early phases, an evaluator must first determine the focus and design of an evaluation. The objective of evaluation is to obtain accurate and objective information on a programme, which has been planned and implemented in the previous phases. The information may be obtained from the process of programme implementation, impacts or results, and efficiency. The results of evaluation determine whether the programme is successful, whether it is going to be continued or it is going to be used as a basis for the next programme. A number of evaluation models have been developed and widely used as strategies or guidelines in the implementation of university programme effectiveness such as Kirkpatrick's Evaluation Model (Kirkpatrick Four Levels Evaluation Model), 2) CIPP Evaluation Model (Context, Input, Process, Product), 3) Stake's Evaluation Model (Model Countenance), 4) Hammond Model and others.

However, the most common models used by researchers are the Kirkpatrick's Evaluation Model and CIPP model.

Model

A model is a conceptual picture of an activity that can demonstrate the relationship between the various elements involved in a programme. The CIPP evaluation model (Stufflebeam et al., 1971) provides a theoretical framework that can guide the determination of the overall quality and merit of a programme. The CIPP model requires consideration of multiple aspects of a programme, including input from representative stakeholders, to conduct a comprehensive assessment. These aspects are assessed via four main evaluations (Context, Input, Process, and Product), which collectively provide data to assess the overall programme. The CIPP model has been utilised for large-scale analyses of education programmes (Azizi, 2001; Megan and Patricia, 2017). The CIPP model is used to evaluate the curriculum of *Kemahiran Hidup* in secondary schools in Malaysia. This model has successfully identified several weaknesses and strengths of a variable in each dimension. The results of the study have been used to make the decision to continue the curriculum implementation of the subject and some recommendations for future development (Azizi, 2001). Megan and Patricia (2017) studied the CIPP model to assess the quality and merit of a nursing education programme. They identified the missing content, programme strengths, and curricular redundancies within the program. Academic nursing programmes engage in multiple evaluation efforts, such as analysing licensure pass rates, verifying coverage of all requisite essential elements and accreditation standards within the coursework, and assessing faculty expertise (Billings and Halstead,2015; Commission on Collegiate Nursing Education, 2013). When used appropriately, the CIPP model serves as a valuable guide for in-depth curriculum evaluation.

Meanwhile, Syamsu (2013) studied the implementation of Kirkpatrick's evaluation model in the learning of initial value and boundary condition problems. The study stated that Kirkpatrick's Evaluation Model for the learning programme of initial values and boundary condition problems is very effective. Kirkpatrick is an expert of training programme evaluation in the human resource development context. The evaluation model developed by Kirkpatrick is known as Kirkpatrick's Four Level Evaluation Model. According to Kirkpatrick (1998) the evaluation of the training programme effectiveness involves four different levels, namely, Level 1 Reaction, Level 2 Learning, Level 3 Behaviour, and Level 4 Outcome.

Effectiveness

A nation's greatest asset is its intellectual resource cultivated through the process of education. A high quality education system will be able to create individuals who are able to function intellectually and productively in society. Therefore, it is very important to have a world class quality education system in order to ensure the development of a country (Nurfaradilla et al., 2010). According to Sopoaga et al., (2017), the value and effectiveness of a programme is for improving the participation and academic success in higher education. Besides that, teacher's effectiveness also varies significantly when different statistical methods are used (Briggs and Domingue, 2011; Newton et al., 2010; Rothstein, 2007). Effectiveness is an effective description, actions, change and so forth (Dictionary of the Third Edition Board, 2007). It produces results in which a change occurs in the mind of a person or the organisation. In this research, effectiveness is intended as a result of the programme implementation in universities.

METHODOLOGY

A meta-analysis is defined as a process of identifying, assessing and interpreting all available research evidence with the purpose of answering specific research questions (Kitchenham and Charters, 2007). It is a tool that aims to produce a scientific summary of the evidence in a particular area, in contrast to the "traditional" narrative review (Pettivrew and Roberts, 2006). The researchers adopted the procedures of Kitchenham and Charters (2007). The intended inclusion criteria were limited to the following characteristics shown in Table 1:

Table 1: Summary of Inclusion Criteria

Item	Description
Year	This item described the 'year' of the articles in this study.
Country	The item described the 'region' studied in the articles.
Methodology	The item described the 'methodology' adopted in the articles.
Model	The model described the evaluation model used to measure the effectiveness in the previous studies.
Dimensions	Dimension recognized specific studies related to the present study.

Meta-Analysis

The term meta-analysis comes from Glass (1976), which is defined as the statistical analysis of the results of individual studies, with the aim to integrate them. According to Willig (1985) and Greene (1998), meta-analysis has provided the best published sources of integrated evidence thus far; however meta-analysis focused primarily on studies done before year 1985. According to Stanley and Jarrell (1989), meta-analysis is an analysis of the empirical analysis which attempts to explain the differences in results between studies. Pignon and Poynard (1993) defined meta-analysis as the use of statistical techniques for the synthesis of a set of separate but similar experiments. Meta-analysis is a simultaneous analysis of a set of studies addressing the same question, in order to obtain the information that none of these studies taken singly could provide and explain the differences in the results of these studies. The first meta-analysis has been realised in the medical field. The objective was to reduce the costs of experimental studies, which often led to different results. This method has rapidly spread in other areas of research such the environment, marketing and social sciences.

Identification of Relevant Literature

A systematic literature search was employed in seven online databases and grey literature sources to search for articles which were relevant to the study objectives. Those seven online databases were EBSCOhost, CENGAGE Learning, Cochrane Library, Clinical Key, ProQuest, ScienceDirect, and Scopus. Researchers' also manually searched targeted paperback journals as well as dissertation reports since some periodicals and publications are not available online. The strategy researchers used to construct the search strings was as follows (Kitchenham and Charters, 2007; Mendes, 2005):

- Derive major terms used in the review questions (i.e. based on the population, intervention, outcome and context);
- List the keywords mentioned in the articles (primary studies) studied by the researchers;
- Search symposium and alternative words. Researchers have also consulted a subject librarian to seek further advice in the proper use of the terms;
- Use Boolean OR to incorporate alternative spellings and synonyms;
- Use Boolean AND to link the major terms from population, intervention and outcome.

The complete search string initially used to search the literature was as follows:

(*Evaluation Model*) AND (*Effectiveness Program in University*)

Petticrew and Robert (2006) highlighted that the two major issues in conducting meta-analysis search are the sensitivity and specificity of the search. Sensitivity refers to a search that retrieves a high number of relevant studies. Specificity causes the search to retrieve a minimum number of irrelevant studies. In preliminary search, researchers retrieved a very small number of articles when using the complete search string defined above. For instance, Google engine searches retrieved only ten and five articles respectively. Therefore, the researcher sought the opinion of a subject librarian regarding the appropriate use of search strings and her advice was that the researchers should use a much simpler string than the one defined in the protocol to enable the retrieval of more results. The researchers used the keywords "Evaluation Model" and "Programme Effectiveness in University" which resulted in a higher number of studies retrieved from various online databases. The primary search process involved the use of 11 online databases: ScienceDirect, Google Scholars, EBSCOhost, ProQuest, PQDT, ISI Web of Science, Sage Publications, ISI Proceedings, SpringerLink, Scopus and SciHub.

The selection of online databases was based on researcher's knowledge of databases that index evaluation model of effectiveness program in university, researchers were aware of and the list of available online databases subscribed by the university. Khan et al., (2003) recommended searching multiple databases to cater for as many citations as possible to avoid bias review. Thus, researchers also searched Mozilla Firefox and Microsoft Edge website using similar keywords (i.e. evaluation model of programme effectiveness in universities). The online Google Scholar was used to search full text of articles. More than twenty- eight studies were identified and reviewed. Researchers experience in literature search supports the suggestion by Kitchenham et al., (2007) that it is important to identify a list of relevant online databases to facilitate the process. Upon completion of the primary search phase, the identification of relevant literature continued with the secondary search phase. During this search phase, all the references in the papers identified from the primary sources were reviewed. If all papers were found to be suitable, it was added to the existing list of studies qualified for the synthesis.

Selection of Studies
The researchers' inclusion criteria aimed to only include evaluation model of programme effectiveness in universities As such, the literature search only covered studies published within the period of year 2000 to 2018. The detailed inclusion criteria comprises; (i) studies on evaluation model of programme effectiveness in

universities and (ii) studies on measuring the evaluation model of programme effectiveness in universities.

The main exclusion criterion comprises evaluation model papers not targeted on programme effectiveness. In addition, the following exclusion criteria were also applied (i) papers presenting claims by authors without supporting evidence; (ii) papers solely on evaluation models; (iii) papers involving university programmes; and iv) papers not written in English.

Data Extraction and Study Quality Assessment
To facilitate the data extraction, a process form was designed and used to gather evidence relating to the research questions, as well as to measure the quality of the primary studies. When designing the study quality checklist, the researchers adopted some of the questions proposed in the literature (Leedy and Omrod, 2005; Petticrew and Roberts, 2006; Spencer et al., 2003; Crombie, 1996; Fink, 2005; Greenhalgh, 2000). The checklist comprises seven general questions (see Table 2) to measure the quality of both quantitative and qualitative studies using a ratio scale: Yes = 1 point; No = 0 points; Partially = 0.5 point. The resulting total quality score for each study ranged between 0 (very poor) and 7 (very good). In order to validate the data extraction process, a random sample comprising of 20% from the total number of primary studies had their data extracted by the other researchers in a review meeting.

Table 2: Study Quality Checklist

No.	Item	Answer
1.	Was the article refereed? (Leedy and Ormrod, 2005)	Yes/No
2.	Were the aims of the study clearly stated? (Crombie, 1996; Greenhalgh, 2000).	Yes/No/Partially
3.	Were the study participants or observational units adequately described? For example, students programming experience, year of study etc. (Petticrew and Roberts, 2006; Greenhalgh, 2000)	Yes/No/Partially
4.	Were the data collections carried out very well? For example, discussion of procedures used for collection and how the study setting may have influenced the data collected. (Greenhalgh, 2000; Fink, 2005; Petticrew and Roberts, 2006; Spencer et al., 2003)	Yes/No/Partially

5.	Were potential confounders adequately controlled for in the analysis? (Fink, 2005).	Yes/No/Partially
6.	Was the approach to and formulation of the analysis well conveyed? For example, description of the form of the original data, rationale for choice of method/tool/package. (Spencer et al., 2003; Greenhalgh, 2000; Fink, 2005).	Yes/No/Partially
7.	Were the findings credible? For example, the study was methodologically explained so that we can trust the findings; findings/conclusions are resonant with other knowledge and experience. (Petticrew and Roberts, 2006; Spencer et al., 2003; Greenhalgh, 2000).	Yes/No/Partially

RESULTS AND DISCUSSIONS

Studies based on Research Approach

**Figure 1: Studies based on Research Approach**
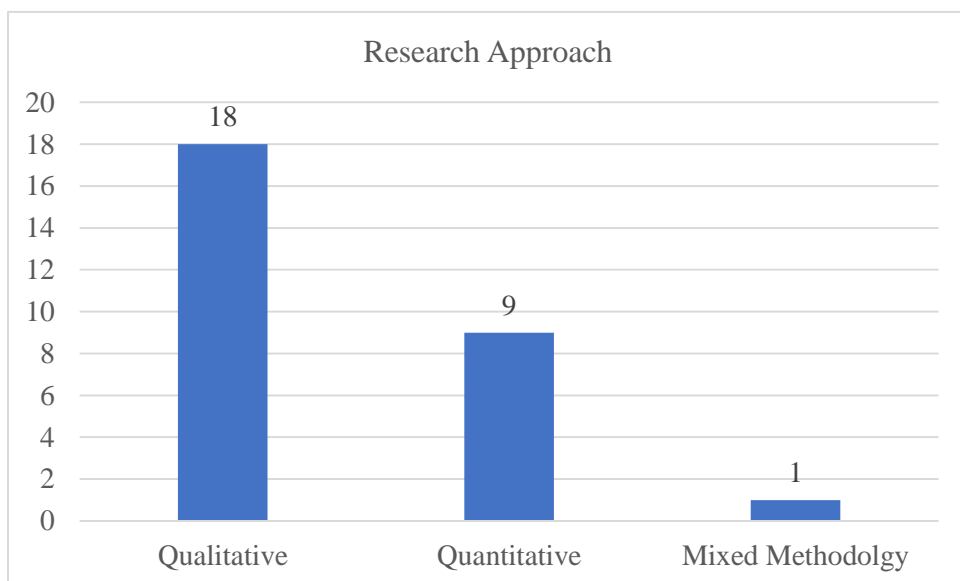
The exploration of research is a ground-breaking strategy that authors must take after to guarantee the fulfilment of research goals. Some portion of the exploration research approach is either a mixed-method, qualitatively, or quantitatively methodology. Figure 1 demonstrates that majority of the investigation on the evaluation model of effectiveness programmed in university were conducted using qualitative methodology (64%), following by quantitative (32%) and mixed methodology approach (4%).

The following sections present the number of studies found between 2003 and 2018 based on the year of study, country of Journals/Thesis/Report/Conference and evaluation models.

Studies based on Years

<h3 style="text-align:center">Figure 2: Studies based on Years</h3>
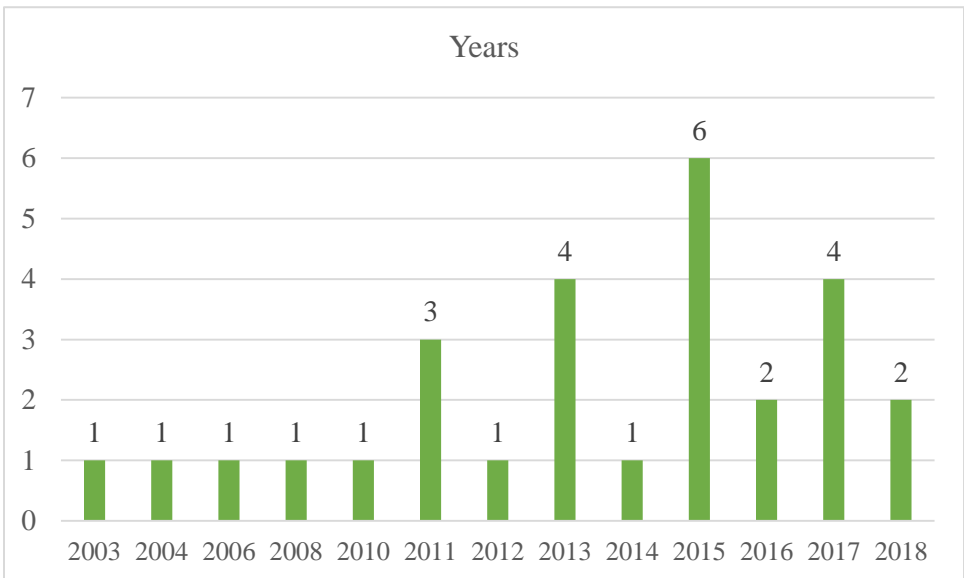
In figure 2, the findings revealed that the highest number of studies on programme effectiveness is six articles in the year 2015. This is follows by four articles in 2003 and 2017, three articles in 2011, two articles in 2015 and 2018, and only one article each year between 2003 and 2010, in 2012 and 2014. The number of articles found between the period of 2003 and 2018 indicated that there has been development in this sort of research since 2013/2015.

**Studies based on Country**

Figure 3: Studies based on Country
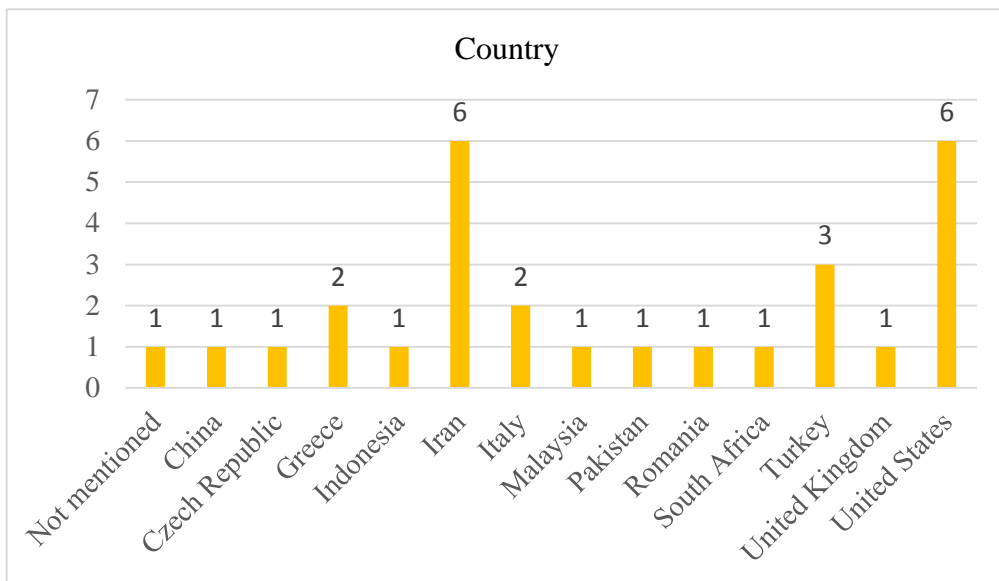
The results indicated in Figure 3 indicated that most of the studies on evaluation programme focused on the United States of America and Islamic Republic of Iran, 6 studies respectively. This is followed by Turkey (3), Italy (2) and Greece (2). Meanwhile there was only 1 study each focused on other country like China, Czech Republic, Indonesia, Malaysia, Pakistan, Romania, South Africa and the United Kingdom.

Studies based on Journals (Thesis (PhD, Masters, Bachelor, Journals)

### Figure 4: Studies based on Journals

As shown in Figure 4, Procedia-Social and Behavioural Sciences Journal published
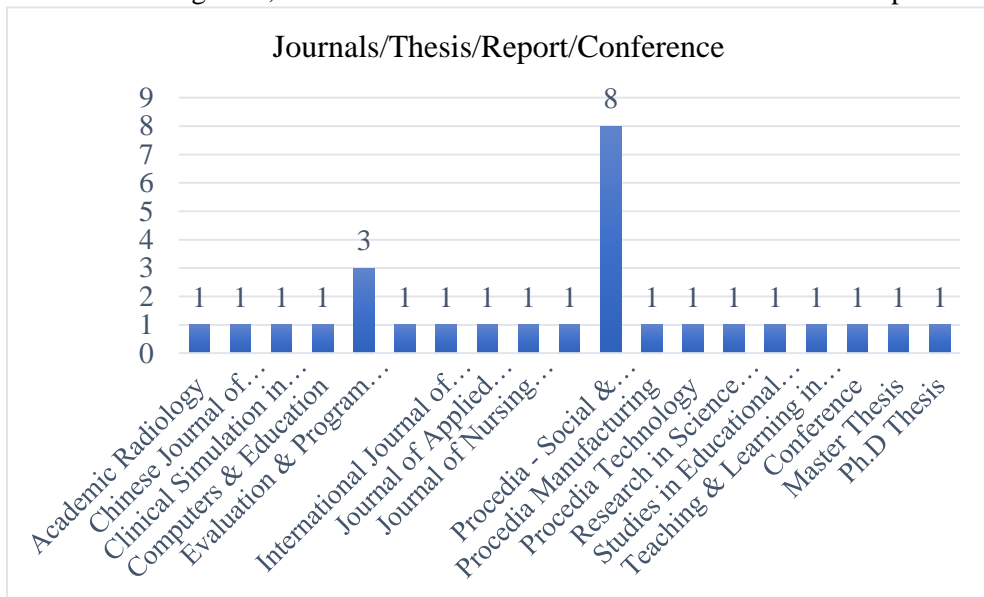
the highest number of studies (eight articles) on evaluation model of programme effectiveness in universities. The second highest number of articles on evaluation model of programme effectiveness in universities is published by the Evaluation and Program Planning Journal. Meanwhile the other articles were published by (1 each) journals of various social sciences disciplines including master and philosophy doctorate theses.

Studies based on Evaluation Model

Figure 5: Studies based on Evaluation Model
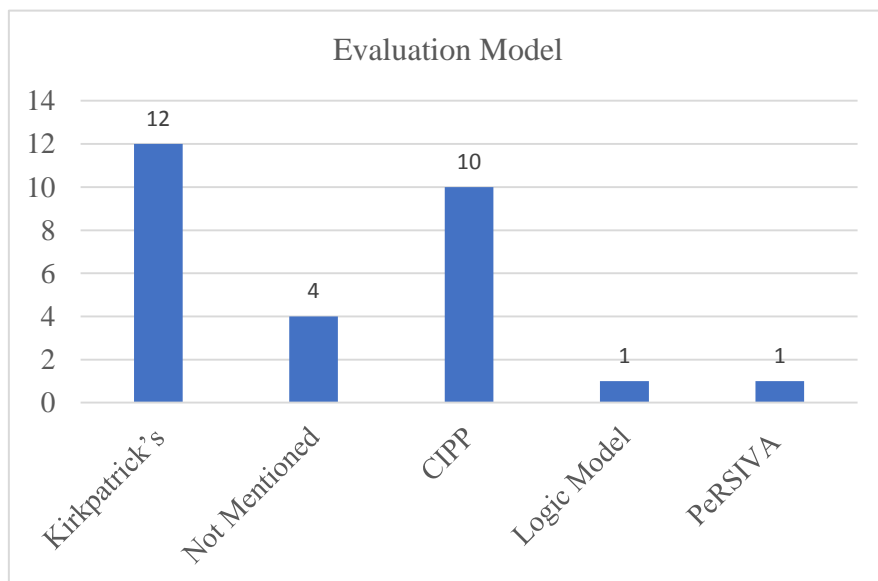
Evaluation is a process that makes it possible to judge quality and it is of great importance. (Pazargadi and Azadi Ahmadabadi, 2008). It can be performed on any activity that has an aim. Through evaluation of an educational programme, the rate of compatibility and conformity of the programme with individuals and societal needs is revealed, the capability of the methods and tools is specified, and the effective factors in programme development are clarified. (Leverenz, 2009).

Evaluation of university programme effectiveness is an increasing demand in obtaining information on academic quality, which contributes to accountability among authorities and affects universities ranking (Neyazi and Arab, 2016).

Figure 5 above shows that 12 studies identified with the Kirkpatrick evaluation model. This is followed by 10 studies adopting the CIPP evaluation model. One study adopted the Logic model and PeRSIVA which is a combination of a well-known evaluation method of Kirkpatrick and the layered evaluation framework is found one study. While four studies did not mention the evaluation model adopted.

CONCLUSION

This paper described a meta-analysis of empirical studies of evaluation model of university programme effectiveness. A total 28 primary study are used in this meta-analysis and two major evaluation models measuring university programme effectiveness were identified. This paper presents clearly, the importance of an evaluation model in measuring university programme effectiveness. Perhaps, a rigorous evaluation model will help the university management in making effective decisions by providing better and meaningful programmes.

An evaluation models helps providers to make decisions on strategic planning in the continuity of a programme. Full commitment from the university's top management is required. If these types of aspirational programmes are to effect change and become sustainable in the long run. Continuous evaluation and improvement of the inappropriate and fairly appropriate indicators are essential to enhance the quality of the programmes. Thus, this meta-analysis study provides informative contents, as well as in extant literature, supports its flexibility and utility to explore the effectiveness of a university programmes using different models. Conclusively, well planned programmes for improving the effectiveness of university programmes is essential.

ACKNOWLEDGEMENT

REFERENCES

Alkin, M. C. (1969). Evaluating theory development. *Evaluation Comment, 2*, 2-7.

Azizi, Y. (2001). *The using of model context, input, process and products (CIPP) in learning programmes assessment*. Paper presented at the International Conference on Challenges and Prospects in Teacher Education. University of Technology Malaysia.

Barr. H & Ross,F. (2006). Mainstreaming interprofessional education in the United Kingdom: A position paper. *Journal of Interprofessional Care, 20*(2):96-104.

Billings, D.M., & Halstead, J. A. (2015). *Teaching in nursing: A guide for faculty.* (5th ed.). St. Louis, Missouri: Saunders.

Briggs, D., & Domingue, B. (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District teachers by the Los Angeles Times. Boulder*, CO: National Education Policy Centre.

Brinkerhoff, R. O., Brethower, D. M., Hluchyj, T., & Nowakowski, J. R. (1983). *Programme evaluation: A practitioner's guide for trainers and educators*. The Netherlands: Kluwer Academic Publishers Group.

Brinkerhoff, R. O. (1989). Using evaluation to transform training. Special Issue: *Evaluating Training Programmes in Business and Industry*, *44*, 5-20.

Cambridge Dictionary. (2018). Retreived from https://dictionary.cambridge.org/dictionary/english/effective

Commission on Collegiate Nursing Education. (2013). *Standards for accreditation of baccalaureate and graduate nursing programmes*. Retrieved from http://www.aacn.nche.edu/ccne-accreditation/Standards-Amended-2013.pdf.

Creswell, J.W. (2003). *Research design qualitative, quantitative and mixed method approaches*. New York: SAGE Publications.

Crombie, I.K. (1996). *The Pocket Guide to Appraisal*. London:BMJ Publishing Group [p1].

Cronbach, L. J. (1963). Course improvement through evaluation. *Teacher College Record*, *64*, 672-683.

Davis, L.M., Robert,B., Jennifer, L.S., Jessica, S., & Jeremy, N. V. M. (2013). *Evaluating the effectiveness of correctional education: A meta-analysis of programs that provide 12 education to incarcerated adults*. Santa Monica, CA: RAND Corporation retrieved from https://www.rand.org/pubs/research_reports/RR266.html. Also available in print form.

Elsevier, B.V. (2008). *Scopus overview: What is it?* [Online]. Available: http://www.info.scopus.com/about/.

Fink, A. (2005). *Conducting research literature reviews: From the internet to paper*. New York: Sage Publication Inc.

Fitzpatrick, J. L., Saners, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternatives, approaches, and practical guidelines*. (3rd ed.). Boston: Pearson Education Inc.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10),3-8.

Greene,J.P.(1998). *A meta-analysis of the effectiveness of bilingual education. Claremont,* CA: Thomas Rivera Policy Institute.

Greenhalgh, T. (2000). *How to read a paper: The basics of evidence-based medicine*. (6th ed.) London[k2]:BMJ Publishing Group.

Griffin, P., & Nix, P. (1991). *Educational assessment and reporting*. Sydney: Harcout Brace Javanovich Publisher.

Khan,K.S., Kunz, R., Kleijnen,J., & Antes,G. (2003). *Systematics reviews to support evidence-based medicine*. London: The Royal Society of Medicine Press Ltd.

Kirkpatrick, D.L. (1998). *Evaluating training programmes: The four levels*. San Francisco: Berretp-Koehler Publisher, Inc.

Kitchenham, B.A., & Charters, S. (2007). *Procedures for performing systematic literature reviews in software engineering*, EBSE Tech. Report Version 2.3, EBSE-2007-01, Software Engineer, Group, Keele University of Durham, UK.

Laura Keeler, R., Lee, K., Kim, P., Katie, D. M., & Jill, A. H. (2018). Effectiveness of the wheelchair skills training program: A systematic review and meta-analysis. *Disability and Rehabilitation. Assistive Technology[p3] Journal,14*(4),391-409. DOI: 10.1080/17483107.2018.1456566.

Leedly, P.D., & Ormrod, J.E. (2005). *Practical research planning and design*. (8th ed.). United Kingdom: Pearson Merril Prentice Hall.

Leverenz, L. (2009). *Allied health education program accreditation - what does it mean?* Retrieved from http://www.studyoverseas.com/alliedhealthcare/articles/healthed.html.

Mardapi, D. (2000). *Evaluasi pendidikan*. In a paper presented at the National Education Convention. Jakarta: Jakarta State University, 19-23 September

McNamara, C. (1998). *Basic Guide to Program Evaluation*. https://www.google.com.my/search?q=McNamara%2C+C.+(1998).-+Basic+Guide+to+program+evaluation.

Megan, L., & Patricia, C. (2017). *Using the CIPP model to assess nursing education program quality and merit. Teaching and learning in nursing* xxx (2017) xxx–xxx. (Article available in sciencedirect.com).

Mendes, E. (2005). *A systematic review of web engineering research*. In International Symposium on Empirical Software Engineering. 498-507.

Narges,N., & Mohammad Arab. (2016). Evaluation of selected faculties at Tehran University of medical sciences using CIPP model in students and graduates point of view. *Evaluation and Program Planning* http://dx.doi.org/10.1016/j.evalprogplan.2016.06.013

Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modelling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives, 18*(23), 1-27.

Nurfaradilla,N., Siti Norhidayah, R., Mohammad Iskandar, S., Kasmah, A.B., Lilia, H.,& Sharifah Nor, P.(2010). Effectiveness of school wide enrichment model (MPSS) in Mara Junior Science College. *Procedia Social and Behavioral Sciences, 7(C),* 43-49.

Pazargadi, M., & Azadi Ahmadabadi, G. (2008) *Quality and quality assessment in universities and higher education institutions* . Tehran: Boshra.

Petticrew, M., & Roberts,H. (2006). *Systematic reviews in the social sciences a practical guide*. United States: Blackwell Publishing.

Pignon, J., & Poynard, T. (1993). [Meta-analysis of therapeutic trials. Principles, methods and critical reading]. *La Revue du praticien, 43*(18): 2383-2386.

Rothstein, J. (2007). *Do value-added models add value? Tracking, fixed effects, and causal inference*. CEPS Working Paper No. 159. Cambridge, MA: National Bureau of Economic Research.

Sopoaga, F., Kokaua, J., Van der Meer, J., Lameta-Huro, Malia., Zaharic, Tony., Richards, Rose., & Inder, Marie. (2017). *Evaluating the effectiveness of a programme for improving the participation and academic success of an underrepresented minority group in New Zealand*. Evaluation and Program Planning http://dx.doi.org/10.1016/j.evalprogplan.2017.06.002.

Spencer,L., Ritchie,J., Lewis,J., & Dillon,L.(2003). *Quality n qualitative evaluation: A framework for assessing research evidence*. Government Chief Social Researchers' Office. London.

Stanley, T. D., & Jarrell, S. B. (1989). Meta-Regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys, 3*(2), 161-170.

Stark, J.S., & Thomas, A. (1994). *Assessment and programme evaluation*. Needham Heights: Simon & Schuster Custom Publishing.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, H. D., Merriman, H. O., & Provus,M.M. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.

Suchman, E. A. (1967). *Evaluative research*. New York: Russel Sage Foundation Talmage.

Syamsu, Q.B. (2013). The implementation of kirkpatrick's evaluation model in the learning of initial value and boundary condition problems. *International Journal of Learning & Development*, *3*(5), 74-88.

Tyler, R.W. (1945). General statement on evaluation. *Journal of Educational Research*, *35*, 492-501.

Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press. Wallace.

Willig, A.C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*(3), 269-318.

Wohlin, C., Runeson, P., Host, M., Ohlsson,M.C., Regnell, B., & Wesslen.A. (2000). *Experimentation in software engineering: An introduction*. Boston: Kluwer Academic Publisher.