

Educational Program and Curriculum Evaluation Models: A Mini Systematic Review of the Recent Trends

Peyman Nouraey^{1,*}, Ali Al-Badi², Mohammad Javad Riasati³, Rolou Lyn Maata⁴

¹Research Department, Gulf College, Muscat, Oman

²Centre for Postgraduate Studies & Research, Gulf College, Muscat, Oman

³Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

⁴Faculty of Computing Sciences, Gulf College, Muscat, Oman

Received May 21, 2020; Revised June 27, 2020; Accepted July 29, 2020

Cite This Paper in the following Citation Styles

(a): [1] Peyman Nouraey, Ali Al-Badi, Mohammad Javad Riasati, Rolou Lyn Maata , "Educational Program and Curriculum Evaluation Models: A Mini Systematic Review of the Recent Trends," *Universal Journal of Educational Research*, Vol. 8, No. 9, pp. 4048 - 4055, 2020. DOI: 10.13189/ujer.2020.080930.

(b): Peyman Nouraey, Ali Al-Badi, Mohammad Javad Riasati, Rolou Lyn Maata (2020). *Educational Program and Curriculum Evaluation Models: A Mini Systematic Review of the Recent Trends*. *Universal Journal of Educational Research*, 8(9), 4048 - 4055. DOI: 10.13189/ujer.2020.080930.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The present study aimed at reviewing some of the most reputable models of curriculum and program evaluation. In so doing, 63 related research papers were selected based on the pre-defined criteria. These included systematic reviews, meta-analyses, case studies, book reviews, and experiments. A curriculum is an important element which can affect the effectiveness of an educational or a pedagogical program. The main merit of a curriculum is providing the stakeholders with a transparent idea as to what must be achieved during the program conduction and whether the objectives of the program have been met. The term program evaluation was first used in the United States during the 1960s. Since then, various program and curriculum evaluation models and frameworks have been conceptualized around the world. Nearly all these models eventually have the same focus, i.e., to see whether a program meets its defined objectives or not. Some of these models have already been subject to evaluation within various educational contexts, while some have been less investigated. In the present systematic review, some reputable models of program and curriculum evaluation were discussed, while some other models were touched upon. The review was mainly divided into theoretical considerations and empirical background. The pros and cons of each model were briefly discussed. In

addition, it was shown how some models have evolved or been challenged by others. It was concluded that choosing an appropriate model of evaluation would depend on several criteria such as the context, the purpose, and the expected outcome of the evaluation.

Keywords Curriculum Evaluation, Educational Programs, Evaluation Models, Program Evaluation, Systematic Review

1. Introduction

A curriculum has a significant role in the effectiveness or failure of an educational program [1]. It is a combination of what shall be taught in an educational context along with a set of pre-defined approaches, delivery methods, assessment criteria, teaching materials and teacher education [2]. It provides the stakeholders (such as students and their parents, staff, shareholders and sponsors) with a clear idea as to what must be achieved during a program conduction period, and how things have actually progressed by the end of the program in question [1].

Curriculum assessment is a significant part in most of

the program evaluations, especially educational programs. The term program evaluation dates back to the 1960s in the United States, where elimination of racial injustice and poverty were sought by the government [3]. After evaluating several programs at that point of time, the US government decided to stop funding some of the programs due to lack of efficiency. However, some are still recognized and funded by the government [3].

The main purpose of program evaluation is clear: whether the program is effective or does it need to be revised [4]. However, some researchers have tended to categorize the possible purpose(s) of program evaluation in a systematic way. As an example, two possible purposes of program evaluation have been pointed out by Weir and Roberts [5], including evaluation for purposes of *accountability* vs. evaluation for purposes of program or project *development*. While the former type of evaluation fits within professional and contractual levels, the latter type aims at improving the program in question. Within the context of education, both professional accountability and program development are meant by the conduction of a program evaluation [5].

To date, several researchers have tended to shed light upon the concept of program evaluation [6-13]. There are more than 50 models of curriculum evaluation available [14]. This variance might be related to differences in evaluation philosophies [15]. Some of these models have constantly been investigated by other researchers and are more reputable in the field, while some have been less studied. Accordingly, the present work aimed at reviewing some of the most well-known program and curriculum evaluation models, focusing on how and when they were conceptualized, what their objectives were and whether they have been challenged by other researchers in the field. The main merit of the present systematic review was aiding the researchers in choosing the right model for their evaluations, which may include educational programs, training courses, organizational workshops, etc. One of the issues raised in the literature was the failure of evaluations due to inappropriate selections of evaluation models. We have provided a brief overview of the most reputable models in question, while touching upon some newer models and frameworks of evaluation. In general, the present work aimed at answering the following research questions:

- 1) What are some of the most reputable program and curriculum evaluation models?
- 2) What specific areas do these models cover?
- 3) What are some of the pros and cons of each model?

2. Research Method

2.1. Design

Having a mixed-methods design, the present systematic

review delved into both qualitative and quantitative research conducted. The underlying reason was to include both theories and applications in the field of program and curriculum design and development.

2.2. Data Collection Instruments

The instruments included research papers (open access and non-open access) from several reputable indexing/abstracting databases, including ScienceDirect, PubMed, and Google Scholar. Different criteria were considered for inclusion of the research articles, including relevance to the searched keywords, impact factor and CiteScore of the research journal, *h* index, and the number of citations of each paper.

2.3. Data Collection Procedure

To collect the data, first, some (N=63) research articles were downloaded and categorized based on their subject areas. These included literature reviews and meta-analyses (N=37), case studies (N=12), book reviews (N=9) and experiments (N=5). The criteria to categorize each paper under a certain research approach were being a) highlighted in the title of the paper, b) mentioned in the abstract, c) determined in the methodology, d) categorized by the database, and e) brought up by search engines. Then, each paper was reflected in the present work in two self-standing sections, namely theoretical considerations, and empirical background.

3. Theoretical Considerations

In line with the objectives of the present review, this section aimed at providing a brief overview of some of the most well-known program and curriculum evaluation models. The main criterion which led to such labeling was the usage frequency of each model within the research conducted in the related realm of evaluation. In other words, these models have repeatedly been utilized by researchers to evaluate various types of programs and curricula.

3.1. The CIPP Model

The CIPP model (Figure1) is a comprehensive, detailed model of evaluation which focuses on four main areas within a program. These include *context*, *input*, *process*, and *product*. Stufflebeam [13] has defined the components of the model as followings:

- 1) *Context evaluation*: What needs to be done? It assesses needs, assets, and problems within a defined environment.
- 2) *Input evaluation*: How should it be done? It assesses competing strategies and the work plans and budgets of the selected approach.

- 3) *Process evaluations*: Is it being done? They monitor, document, and assess program activities.
- 4) *Product evaluation*: Did it succeed? In the latest checklist, the product evaluation part is divided into *impact*, *effectiveness*, *sustainability*, and *transportability* evaluations:
 - a) *Impact evaluation* assesses a program’s reach to the target audience.
 - b) *Effectiveness evaluation* assesses the quality and significance of outcomes.
 - c) *Sustainability evaluation* assesses the extent to which a program’s contributions are successfully institutionalized and continued over time.
 - d) *Transportability evaluation* assesses the extent to which a program could successfully be adapted and applied elsewhere.

Based on this model, a program may go on as it is, if the evaluation checklist confirms its efficiency in achieving its objectives; otherwise, there would be a need for innovative changes and updates, in case shortcomings are observed.

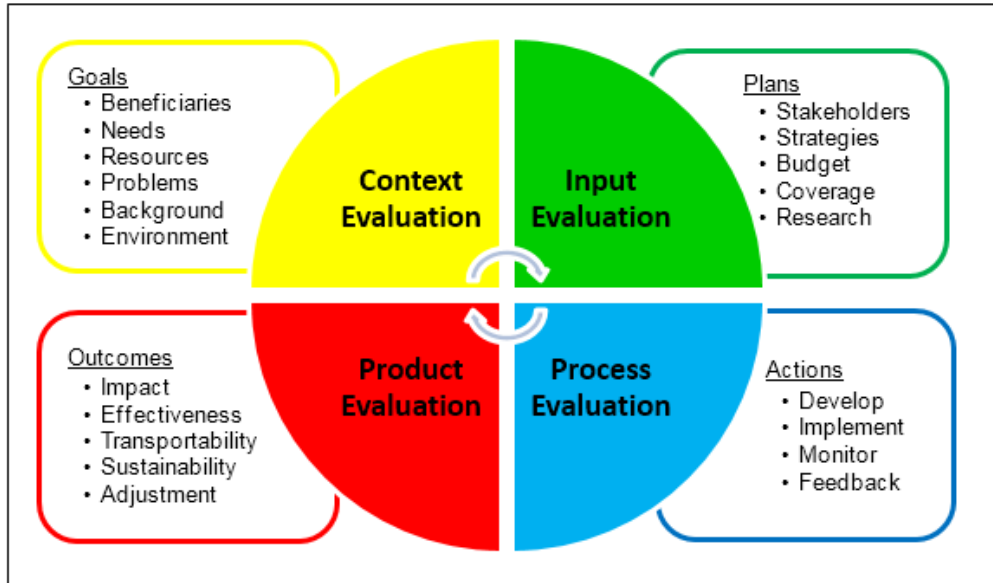


Figure 1. The CIPP model by Stufflebeam [13]

Table 1. Philips’ [18] Model of Learning Evaluation

| Levels of Evaluation | Measurement Focus | Time frame |
|----------------------|--|--|
| Level 1: Reaction | Participant reaction to a learning program | Conclusion of program |
| Level 2: Learning | Degree to which participants acquired new knowledge, skills or attitudes | Conclusion of program or within 6 to 8 weeks after |
| Level 3: Behavior | Degree to which participants applied back-on-the-job what was learnt | 2 to 12 months |
| Level 4: Results | Degree to which targeted business outcomes were achieved | 9 to 18 months |
| Level 5: ROI | Degree to which monetary program benefits exceed program costs | 9 to 18 months |

3.2. The Four-Level Model of Learning Evaluation

Kirkpatrick's [16] four-Level model of learning evaluation (also known as a framework), a detailed model of program evaluation, was first introduced in the 1950s. Since then, it has gone under several revisions; however, the main concepts of the model (i.e., the four main levels of evaluation) have remained intact so far [17]. According to Kirkpatrick and Kirkpatrick [16], these four levels of evaluation are:

- a) *Reaction*: what participants think and feel about the program;
- b) *Learning*: the increase in the knowledge and/or skills of participants, as well as the change in their attitudes; evaluation at this level occurs during the period of program conduction through either a knowledge demonstration or various types of tests and assessments.
- c) *Behavior*: positive and effective transfer of knowledge, skills, and/or attitudes of participants from one level to another; this occurs as a post-training evaluation, usually through observations of managers, lecturers, or supervisors.
- d) *Results*: the final results occurred because of attendance, participation, implementation of program objectives in real-life situations, etc.

3.3. Philips' Model of Learning Evaluation

Philips' [18] model of evaluation (Table 1), also known as a model of learning evaluation, is considered as a complementary model to Kirkpatrick's four-level model [16] by adding a fifth level of evaluation to it, i.e., Return on Investment (ROI). In brief, Philips' [18] model focuses on how to collect data, isolate the effect of training vs. other factors, and account for more benefits. According to Philips [18], having determined a learning program's business impact at Kirkpatrick's fourth Level [16], the impact in question may be turned into monetary terms and subsequently compared with the total cost of the program through ROI calculation. According to Philips' [18] model, the ROI is calculated based on the following formula:

$$ROI\% = \frac{\$benefit\ of\ training - \$cost\ of\ training}{\$cost\ of\ training}$$

3.4. Summative and Formative Evaluations

The terms summative and formative evaluations were first introduced by Scriven [19], each aiming at reflecting the distinctions between issues related to implementation on the one hand, and evaluation of a program's effectiveness on the other. These terms have constantly been the subject of discussion in his later work [20-22]. For Scriven [22], formative evaluation was primarily associated with program design and implementation analyses. On the other hand, summative evaluation would deal with whether a program had achieved its stated and/or

intended aims and objectives.

Although Scriven's [22] model gained significance among the evaluators, it has gone under some critiques by scholars and researchers in the field. As an example, Chen [23] challenged Scriven's [22] taxonomy by highlighting the fact that evaluations may be summative and formative, simultaneously. In fact, Chen [23] proposed a framework with two main evaluation purposes (i.e., assessment and improvement), as well as two program stages (i.e., process and outcome). Chen [23] argued that while assessment might be summative, improvement could be formative, and evaluations aiming at improving or assessing a program could focus on achievement and objective implementation. Therefore, it might be a mixture of both summative and formative evaluations and may not purely belong to one category. Despite all the challenges, Scriven's [22] taxonomy has remained very popular among the evaluators [24].

3.5. Shifting to Competency-Based Models of Evaluation

Health sector is one of the most critical and challenging sectors which requires careful evaluation when it comes to different programs. Buker and Niklason [6], referring to Association of University Programs in Health Administration (AUPHA), have pointed out that almost 1 out of 4 programs does not fully meet the certification and/or accreditation criteria. Therefore, Buker and Niklason [6] proposed the following six steps in making program and curriculum evaluations more effective:

- 1) Evaluating the program's mission and aligning it with the preferred outcome;
- 2) Mapping the program curriculum to the adopted set of competencies;
- 3) Mapping the competencies to course objectives by using a defined educational model, framework or taxonomy;
- 4) Designing the measures of competency mastery by using a mixture of summative and formative evaluation strategies;
- 5) Compiling and reviewing the results of summative assessment to ensure that student learning is efficient and leading into competency; and,
- 6) Developing an action plan to initiate change whenever necessary in order to improve the program curriculum and increase its competency.

3.6. Similarities among Evaluation Models

Program evaluation models developed by different researchers in the field have sometimes been very similar to one another in terms of their evaluation criteria. For instance, two very similar program evaluation models within the realm of education were introduced in 1992. Table 2 compares these two models in terms of their evaluation criteria:

Table 2. A Comparison among Two Similar Program Evaluation Frameworks (adopted from [25])

| Item | Rea-Dickins and Germaine [4] | Alderson and Beretta [26] |
|------|-----------------------------------|---|
| 1 | Who/Why? | Stakeholders / purpose |
| 2 | What | Content |
| 3 | Criteria | ----- |
| 4 | When | Timing |
| 5 | Procedures | Method |
| 6 | Who/Where to get information from | Sources of information |
| 7 | Who to involve | Stakeholders, staff, those to be evaluated, those who are the sources of data |
| 8 | How to manage evaluation | Project framework |
| 9 | What to do with the information | Implementing an evaluation |
| 10 | Constraints | Factors the absence of which causes restrictions on the progress |

3.7. Role of Other Factors in Program Evaluation

One of the significant concepts upon which some evaluation models and theories have been conceptualized is the concept of timing. Timing has constantly played a crucial role in how program evaluation models were formed. For example, McDavid and Hawthorn [24] have categorized program evaluation into two mainstreams, namely a) *ex ante* evaluation (done prior to implementation of the program) and b) *ex post* evaluation (done after program implementation). For Henning [27], there are three stages related to timing in a program evaluation, including a) *prior-to-program implementation*, b) *during-program delivery*, and c) *following-program execution*.

Another important factor in the formation of curriculum and program evaluation models is the purpose of evaluation. Based on this factor, McNamara [10] has divided evaluation into three categories, including a) *goals-based*, b) *process-based*, and c) *outcomes-based*. In another taxonomy developed by Rossi *et al.* [3], five aspects of program evaluation have been identified, including:

- a) Needs assessment, which aims at examining the problem which the program addresses;
- b) Program theory, or the conceptual framework of the program;
- c) Process analysis, which aims at evaluating the program implementation;
- d) The impact of evaluation, which aims at highlighting the effect of the program; and,
- e) Cost-benefit or cost-effectiveness analysis, which aims at assessing the program's effectiveness and efficiency with insights from the costs and benefits.

In addition to these two factors, roles of certain

designations on curriculum and program evaluation have been discussed by researchers. For example, the following roles in the process of curriculum and program evaluation have been proposed [28]:

- a) Students: The primary and the most important source of information to check the implementation, effectiveness and needs analyses required;
- b) Teachers: Having the role of transacting the curriculum in the class, as well as being a part of it, teachers are known to have a considerable share in its evaluation process;
- c) Subject experts: Form a discipline point of view, subject experts could provide a considerable deal of useful information to contribute to the process of evaluation and its implementation;
- d) Curriculum experts: The responsibilities of such experts may include enriching the evaluation process with ways to develop a curriculum and introducing modern methods to evaluate it;
- e) Policy makers: Due to their position, policy makers have a clearer idea about how the program is being implemented and whether it is meeting its targets;
- f) Community: This is where the final product of an educational context (i.e., trained or educated people) would have interactions with; therefore, it could considerably reflect the effectiveness and efficiency of the curriculum or the program;
- g) Dropouts sample: Students who have dropped out from a particular course could provide invaluable information about the misconceptions, reasons or factors which have led them to make such a decision;
- h) Employers and entrepreneurs: Similar to the community, these groups of people could also reflect the strengths and weaknesses of a curriculum or a program.

4. Empirical Background

Despite the significant role of curricula in the development of the nursing profession and nurses, there is no widely accepted or consistent approach in curriculum development, redesign and renewal in this field. This gap motivated Jager *et al.* [29] to identify the current curriculum redesign and renewal practices through creation of an aggregated logic model. Data were collected through qualitative, quantitative, and non-research literature in both English and French languages. The outcome of Jager *et al.*'s [29] study, referred to as the Ottawa Model for Nursing Curriculum Renewal, included information on the context, process and outcomes of the curricula renewal process, as well as when and how to conduct the evaluation on the curricula.

The CIPP model [13] has repeatedly been used by researchers to conduct evaluations in various contexts. Kavgaoglu and Alci [14] carried out a case-study of a call

center within a reputable telecommunication sector in Turkey based on the CIPP model [13]. The main purpose of their study was to evaluate the competence-based curricula designed by means of internal funding through the CIPP model. Participants of the study were 622 call center agents serving in three different regions during 2014 and 2015. Data were collected and analyzed based on both gap analysis and Structural Equation Modelling (SEM). Findings out of the CIPP evaluation model [13] scorings revealed significant differences among gender, education sphere and education level of the participants. It was also found that the male participants made higher scorings than the females in some aspects of the questionnaire designed based on the CIPP model [13]. Similarly, higher scorings were observed by high-school graduates as compared to high-school and upper secondary-school students.

Another study utilizing the CIPP model [13] was conducted by Karimnia and Kay [30], who carried out an empirical investigation to evaluate the undergraduate curriculum at B.A level within the field of Teaching English as a Foreign Language (TEFL) in a university context. In so doing, five universities were selected through cluster sampling, and 20 students were randomly asked to fill in a questionnaire designed based on the CIPP model [13]. Findings revealed that most of the students would prefer revisions on the TEFL program along with its materials being taught at universities. In addition, students argued that instructors had better focus on teaching specific modules, concentrating more on the students' learning strategies. In addition to students' responses, some semi-structured interviews with instructors were conducted, the results of which called for fundamental changes in the TEFL curriculum.

As mentioned earlier, some of the evaluation models have repeatedly been used by researchers, which could possibly reflect the reputability, as well as the popularity of these models. An example includes an empirical study conducted by Hamemoradi *et al.* [17] on the effectiveness of on-the-job training courses in a major gas company. The study was conducted using the CIPP model [13], the four-level model of learning evaluation [16] and Philips' [18] model of learning evaluation. Participants were 291 employees randomly selected and asked to fill in a survey questionnaire based on the three aforementioned models. Findings of the study revealed that the training courses in question lacked effectiveness in terms of reactions, learning, behavior, results, procedure and relevant output. It was also highlighted that the program was effective only in two of its predefined elements, namely input and profitability factors.

Al-Mamari [31] aimed at reviewing the academic standards for the General Foundation Program (GFP), a pre-requisite program consisting of general English language, information technology, general study skills and basic, pure and applied mathematics modules [32]. It was argued that the standards in question only provided minimum requirements to be attained, and in fact, the

program did not consider the possible challenges derived from its implementation. Al-Mamari [31] highlighted some of these challenges, including *a) low proficiency in English for higher education intake, b) four areas of learning, c) entrance requirements and placement tests, d) exit requirements, and, e) resources.* Al-Mamari [31] discussed a number of issues and challenges that prevented the GFP to meet its final end, i.e., quality enhancement [33].

Al-Mahrooqi's [34] study aimed at highlighting the shortcoming of the GFP. Data were collected from 100 tertiary education students through a qualitative questionnaire. Findings of Al-Mahrooqi's [34] study revealed that despite the large budget and investment by the government, the GFP did not provide the desired outcomes. Lack of effectiveness in teachers, inappropriate curriculum, lack of interest in students, lack of exposure to English outside the classrooms, lack of support by parents, a poor school system, as well as peer-group discouragement was among the most addressed challenges reported by the participants.

Forouzandeh *et al.* [25] carried out a large-scale study to evaluate the TEFL program at master's level in nine major universities. The evaluation of the official curriculum developed first in 1987 used in these universities was done through the CIPP model [13]. Sixty-eight M.A students, 34 instructors and 9 administrators participated in the study. Data were collected through three sets of questionnaires, interviews, and written responses. Findings of Forouzandeh *et al.* [25] revealed no agreement among the participants with special reference to the program's main objectives. In addition, it was found that the implemented curriculum was partially compatible with the official one. The participants also urged the need for the reform in the program's delivery, revision of the official curriculum, and screening system reconsideration.

A quick search of the literature reveals several studies conducted in the field of program and curriculum evaluation. These studies focus on both theories and applications, and therefore, various designs could be found. From quantitative to qualitative designs, from literature reviews to experiments, from case studies to reports, all tend to highlight the core issues found in program and curriculum evaluation and development. There are also various domains which have constantly been investigated, including programs of nurse education, teacher education, language education, etc.

5. Conclusions

Several program and curriculum evaluation models have been proposed by scholars and researchers in the field. Some of these models have gone under investigation, while others have been less studied. Similarly, some have been subject to criticisms and have constantly been challenged. However, the aim of all these evaluation models is ultimately figuring out whether or not a program or a

curriculum has met its objectives. Recently, there has been a shift in how program evaluation models are conceptualized, designed and manifested. In fact, researchers are no longer after the sole cause-and-effect relationship between what is expected from and the actual outcomes of a program. They rather tend to prioritize concepts such as adequacy, effectiveness, efficiency, value, and competency [3]. This has led to more complexity in the field of program and curriculum evaluation [35].

In conclusion, choosing a certain model of evaluation depends on several criteria that need to be considered by the evaluator(s) in advance. Some examples may include the context of the program (e.g., educational, administrative, organizational, etc.), the evaluation objectives (e.g. whether ROI should be considered as part of the evaluation or not), the evaluation risk factors, and other criteria. All of these would in turn affect the type and the nature of the evaluation model to be selected. For instance, Philips' [18] model would be a suitable choice for evaluating the effectiveness and efficiency of a series of workshops conducted as a professional development program in an organization, as it calculates whether or not the ROI has been successful, while it may not work properly when utilized within other contexts. Therefore, choosing the appropriate evaluation model must be done based on the aims and objectives of the evaluation. Similarly, it would be better to use reputable models, as they have been continuously challenged and investigated by others in the field. Finally, it would be important to consider all aspects of the model. For example, Kirkpatrick & Kirkpatrick [16] have argued that level 3 is the "forgotten level" in many evaluations, and therefore neglected widely, as professionals usually have a lot of control over the first two levels and executives are interested in level 4 (p.83). Others have also reported similar issues [36]. Considering these suggestions may lead to a more effective and efficient evaluation.

Acknowledgments

The research leading to these results has received funding from The Research Council (TRC) of the Sultanate of Oman under the Block Funding Program (Agreement No. TRC/BFP/GULF/01/2019).

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

- [1] Drakos, J., *Ideas for developing a personal EFL teaching*

curriculum. The Internet TESL Journal, XI (3), 2005: p. 1 – 3.

- [2] Medgyes, P. and M. Nikolov, *Curriculum development in foreign language education: The interface between political and professional decisions*, in *The Oxford handbook of applied linguistics*. 2000.
- [3] Rossi, P.H., M.W. Lipsey, and H.E. Freeman, *Evaluation: A Systematic Approach*. 1993.
- [4] Rea-Dickins, P. and K. Germaine, *Evaluation: Oxford University Press*. Hong Kong, 1992.
- [5] Weir, C. and J. Roberts, *1994: Evaluation in ELT*. Oxford: Blackwell. 1994.
- [6] Buker, M. and G. Niklason, *Curriculum Evaluation & Improvement Model*. Journal of Health Administration Education, 2019. **36**(1): p. 37 – 55.
- [7] Hall, C.E., *Toward a model of curriculum analysis and evaluation—Beka: A case study from Australia*. Nurse education today, 2014. **34**(3): p. 343 – 348.
- [8] Horan, M., et al., *A model for curriculum evaluation*. Journal of Nursing Education, 1984. **23**(7): p. 319 – 322.
- [9] McCormick, R. and M. James, *Curriculum evaluation in schools*. 2018: Routledge.
- [10] McNamara, C., *Field guide to nonprofit program design, marketing and evaluation*. 2006: Authenticity Consulting.
- [11] McNeil, R.C., *A Program Evaluation Model: Using Bloom's Taxonomy to Identify Outcome Indicators in Outcomes-Based Program Evaluations*. Journal of adult education, 2011. **40**(2): p. 24 – 29.
- [12] Norman, P.J. and S.A. Sherwood, *Using internal and external evaluation to shape teacher preparation curriculum: A model for continuous program improvement*. The New Educator, 2015. **11**(1): p. 4 – 23.
- [13] Stufflebeam, D.L., *The CIPP model for evaluation*, in *International handbook of educational evaluation*. 2003, Springer. p. 31 – 62.
- [14] Kavgaoglu, D. and B. Alci, *Application of Context Input Process and Product Model in Curriculum Evaluation: Case Study of a Call Centre*. Educational Research and Reviews, 2016. **11**(17): p. 1659 – 1669.
- [15] Worthen, B.R. and J.R. Sanders, *Educational evaluation: Alternative approaches and practical guidelines*. 1987: Longman Pub Group.
- [16] Kirkpatrick, D. and J. Kirkpatrick, *Evaluating training programs: The four levels*. 2006: Berrett-Koehler Publishers.
- [17] Hamemoradi, M., A. Khorasani, and K. FathiAjargah, *Evaluation of the effectiveness of the on-the-job training courses in NIGC based Kirkpatrick, Phillips and CIPP models*. Human Resource Management in The Oil Industry, 2014. **5**(19): p. 103 – 130.
- [18] Phillips, J.J., *Return on investment in training and performance improvement programs*. 2012: Routledge.
- [19] Scriven, M., *The Methodology of Evaluation*. In. Tyler, RW, Gagne, RM, Scriven, M.(ed.): *Perspectives of Curriculum*

- Evaluation*. (Eds.): Book The Methodology of Evaluation. In. Tyler, R.W., Gagne, R.M., Scriven, M. (Ed.): Perspectives of Curriculum Evaluation, Rand McNally, Chicago, 1967.
- [20] Scriven, M., *Beyond formative and summative evaluation*. Evaluation and education: At quarter century, 1991. 10(Part II): p. 19 – 64.
- [21] Scriven, M., *Types of evaluation and types of evaluator*. Evaluation practice, 1996. 17(2): p. 151 – 161.
- [22] Scriven, M., *A summative evaluation of RCT methodology: an alternative approach to causal research*. Journal of multidisciplinary evaluation, 2008. 5(9): p. 11 – 24.
- [23] Chen, H.-T., *A comprehensive typology for program evaluation*. Evaluation practice, 1996. 17(2): p. 121 – 130.
- [24] McDavid, J.C., I. Huse, and L.R. Hawthorn, *Program evaluation and performance measurement: An introduction to practice*. 2018: Sage Publications.
- [25] Forouzandeh, E., A. Riazi, and F. Sadighi, *TEFL program evaluation at master's level in Iran*. 2008.
- [26] Alderson, J.C., A. Beretta, and M.H. Long, *Evaluating second language education*. 1992: Cambridge University Press.
- [27] Henning, G., *A guide to language testing: Development, evaluation, research*. 1987.
- [28] Sabharwal, V., et al., *Unit-2 Considerations for Curriculum Planning*. 2018, IGNOU.
- [29] Jager, F., et al., *The Ottawa model for nursing curriculum renewal: An integrative review*. Nurse Education Today, 2020: p. 104344.
- [30] Karimnia, A. and E. Kay, *An Evaluation of the Undergraduate TEFL Program in Iran: A Multi-Case Study*. International Journal of Instruction, 2015. 8(2): p. 83 – 98.
- [31] Al-Mamari, A.S. *General foundation program in higher education institutions in Oman national standards: Implementation and challenges in Oman Quality Network Regional Conference*. 2012.
- [32] Carroll, M. *The Case for Foundation Program Standards and Accreditation (Keynote Address) in Proceedings of the symposium Towards National Standards for Foundation Programs*. 2007.
- [33] Carroll, M., S. Razvi, and T. Goodliffe. *Using foundation program academic standards as a quality enhancement tool*. in INQAAHE, United Arab Emirates. Retrieved from <http://www.inqaahe.org/main/events-and-proceedings/inqaahe-2009-conference>. 2009.
- [34] Al-Mahrooqi, R., *A student perspective on low English proficiency in Oman*. International Education Studies, 2012. 5(6): p. 263 – 271.
- [35] Brewer, E.W., *Evaluation models for evaluating educational programs*, in *Assessing and evaluating adult learning in career and technical education*. 2011, IGI Global. p. 106 – 126.
- [36] Dorri, S., Akbari, M., and Sedeh, M. D., *Kirkpatrick evaluation model for in-service training on cardiopulmonary resuscitation*. Iranian Journal of Nursing and Midwifery Research, 2016. 21(5): p. 493 – 497.