# A Comprehensive Evaluation in Medical Curriculum Using the Kirkpatrick Hierarchical Approach: A Review and Update

**Mia Kusmiati** [1]

[1] Department of Medical Education, Bioethics and Humanity-Medical Faculty of Universitas Islam Bandung, Tamansari street no.22, Bandung-West java, Indonesia

Email: mia.kusmiati@unisba.ac.id

## ABSTRACT

This paper reviews the application of the Kirkpatrick hierarchical model in evaluating medical education curricula, emphasizing its role in enhancing curriculum quality through stakeholder feedback. It outlines the four levels of evaluation defined by the Kirkpatrick model: Reaction, Learning, Behaviour, and Results, each involving diverse stakeholders in assessing educational outcomes. The paper highlights the importance of multi-source feedback (MSF) methodologies, which provide holistic insights into medical students' competencies, facilitating targeted improvements. Despite the challenges associated with implementing comprehensive evaluations, such as stakeholder resistance and varying interests, the framework promotes ongoing quality enhancement in medical education, ensuring alignment between educational strategies and workforce needs. The study reinforces that effective evaluation systems are crucial for cultivating skilled, competent healthcare professionals and suggests further integration of various theoretical models to bolster educational practices and policy decisions.

**Keywords:** Evaluation, Kirkpatrick model, medical curriculum, stakeholder feedback

## Introduction

The quality of the medical program is dependent on the curriculum. Evaluation of stakeholders is an important component to ensure quality in medical education. Feedback and evaluation from stakeholders are also an essential element for the continuous improvement of quality.[1] A comprehensive evaluation is built from a common understanding of stakeholders on the most appropriate strategy to produce the expected outcome of education.[2] The evaluation of stakeholders includes four components, namely: satisfaction, evaluation, needs, and perspectives.[3]

Curriculum evaluation tools often focus on specific, compartmentalized components, leaving a gap in assessing the holistic performance of medical students. Recognizing this limitation, medical education has increasingly embraced multisource feedback (MSF) over the past decade. This shift stems from the understanding that academic performance encompasses a multifaceted skill set, including clinical competence, communication, professionalism, and teamwork – areas not always captured by traditional assessments.[4] Multi-source feedback on the curriculum, in this case, can be considered a comprehensive evaluation, with the method employed using a Kirkpatrick's hierarchy approach.

Multi-source feedback, by gathering perspectives from various sources such as peers, patients, medical teachers, and supervisors, provides a more comprehensive evaluation.[5] It offers valuable insights into a student's strengths and weaknesses, facilitating targeted feedback and personalized learning plans. This holistic approach not only promotes individual growth but also fosters a culture of continuous improvement within the medical education system.

While challenges exist in implementing and interpreting MSF, its potential to enhance the development of well-rounded and competent medical professionals is undeniable. By moving beyond fragmented assessments, medical education can better equip future physicians to meet the complex demands of modern healthcare. Multi-source feedback represents a crucial step towards cultivating a learning environment that values comprehensive development and fosters excellence in patient care.

This paper aims to explore the Kirkpatrick model approach in comprehensively evaluating the medical curriculum. The Kirkpatrick model has been widely recognized for its comprehensive methodology in assessing programs or curricula, particularly in medical education.

The Kirkpatrick model consists of four evaluation levels,[6,7] each involving different stakeholders:

1. **Reaction** - Measures participants' satisfaction with the program (student feedback).
2. **Learning** - Assesses knowledge, skills, and competency development (evaluation by faculty members).
3. **Behaviour** - Examines whether learned concepts are applied in professional practice (performance assessment in hospital settings).
4. **Results** - Determines the extent to which learning translates into improved workplace performance and patient care outcomes.

A comprehensive curriculum evaluation is crucial to understanding its effectiveness, including methodologies used, assessment contexts, relevant stakeholders, and their roles in different evaluation levels.[8] Given the variety of assessment methods in medical education, aligning evaluation objectives—whether bureaucratic, autocratic, or democratic—is essential.[9] The categorization of evaluation purposes based on these three aspects is grounded in political objectives and professional accountability. Evaluations that are bureaucratic in nature relate to the prevailing value systems within the government, and the information provided is intended for policymaking. Autocratic evaluation is conducted as a conditional service to the government. The evaluator provides validation of a particular policy with recommendations that are to be met by the agency. Democratic evaluation provides an information service to the community about a particular curriculum or educational program. This approach emphasizes the rights to know about the program and the responsibilities of the evaluator to offer confidentiality to the informants, involve more stakeholders in the evaluation process and provide increased opportunities for dialogue and deliberation. [9]

Conducting a comprehensive evaluation is not without challenges, particularly due to the diverse characteristics and interests of stakeholders. However, with sufficient time and commitment, potential barriers can be managed effectively. However, if sufficient time and commitment from stakeholders are dedicated throughout the evaluation process, the barriers faced are not insurmountable. One key point to emphasize is that stakeholder commitment can be fostered, as they share common goals. Stakeholder engagement is vital as their shared objectives foster commitment to the evaluation process. A comprehensive approach to curriculum or program evaluation is crucial and significant to discuss, starting from how to conduct it, what context is being assessed, who the relevant stakeholders are to be involved, when and at what level they will contribute to the object being evaluated.[8] This topic is important, considering the many evaluation methods used to assess a program, particularly in the field of medical education. The stakeholder perspective is a key factor in identifying the developmental needs of the curriculum[3] because stakeholders can provide an overview of a variety of aspects of the medical curriculum that may challenge existing practices. Stakeholders' evaluation is the process to review, analyse, and to critic the importance or value of information gathered from stakeholders regarding the program, particularly medical curriculum.

## Theoretical Framework

This evaluation approach is grounded in several theoretical models and conceptual frameworks, including:

- **Multisource Feedback Theory.** [4] Multisource feedback can be considered a 360-degree evaluation model. The 360-degree evaluation model is one of the best assessment methods that can be deployed to examine professionalism and

communication skills competencies. In line with complexity theory, a 360-degree curriculum evaluation refers to a thorough evaluation of the curriculum components and the interactions between those components.[5]

Multisource feedback is a comprehensive evaluation method that gathers performance insights about an individual, subject or educational program from a variety of sources. This approach encompasses feedback from variety stakeholders like customers or client, in addition to the individual's self-assessment.[10]

In this case, the group of stakeholders in medical education includes: students, lecturers, patients, relevant policymakers, government agencies overseeing education, and others.

Several studies have shown that multisource ratings are related to a variety of performance measures and hence provide evidence concerning the concurrent validity of multisource ratings. Other studies have examined whether different rater sources (e.g., peers, direct reports) conceptualize performance in a similar manner

- **Conceptual Framework Integrating the Logic Model and Kirkpatrick Model,**[11] highlighting comprehensive evaluation methods.

  A conceptual framework that combines the logic model and Kirkpatrick with the involvement of various stakeholders in medical education.[3] Both evaluations using the logic model approach and Kirkpatrick represent comprehensive approaches to evaluating programs or curricula. The logic model details the evidence and assumptions that underpin the complex pathway from interventions to demand management impact. Logic models aim to uncover the theories of change, identifying the assumptions that link interventions to both short- and long-term outcomes.[12,13]

  Evaluating curriculum effectiveness requires a robust framework that considers various dimensions, from resource allocation to student learning. Combining the Logic Model and Kirkpatrick's Model offers a powerful synergy, yielding a comprehensive assessment spanning inputs, processes, and outcomes. The Logic Model, focusing on the program's theory of change, allows for a structured examination of inputs (resources, personnel), activities (instructional strategies), outputs (program participation, completion), and outcomes (short-term and long-term impact). This provides a clear roadmap of how the curriculum is intended to function.[12]

  Kirkpatrick's Model, on the other hand, provides a hierarchical framework for evaluating the process and impact. Level 1, Reaction, assesses participant satisfaction. Level 2, Learning, measures gains in knowledge, skills, and attitudes. Level 3, Behaviour, evaluates the transfer of learning to practice. Finally, Level 4, Results, assesses the ultimate impact on organizational goals or student achievement.

  Integrating these models allows for a more holistic view. The Logic Model provides the structural framework for identifying key elements,[12] while Kirkpatrick's Model provides the instruments for measuring the effectiveness of each stage. For example, the Logic Model might identify "engaging activities" as a critical activity, while Kirkpatrick's

Level 1 (Reaction) could gauge participant engagement with those activities. Level 2 can then assess whether engagement translated into actual learning gains aligned with the intended outcomes defined by the Logic Model. Ultimately, Levels 3 and 4 of Kirkpatrick's Model assess the real-world impact, validating the long-term effectiveness of the curriculum as envisioned in the Logic Model.[7]

By combining these two established evaluation methodologies, curriculum developers can gain a deeper understanding of the curriculum's strengths and weaknesses, leading to evidence-based improvements and ensuring it effectively achieves its intended goals.

- **Elements of a program or curriculum as objects to be evaluated.** In the context of curriculum elements, we recognize components such as learning content, teaching strategy, methods of assessment, and evaluation processes.[14] The curriculum is a sophisticated blend of teaching strategies, the content of the program, learning outcomes, learning experiences, assessment methods, learning environments, and individual student's learning style, timetable, and program of work.[6] All these components can be continuously assessed by various educational stakeholders comprehensively through the Kirkpatrick level assessment approach[7] which emphasizes a 360-degree assessment model for evaluating professionalism and communication competencies.

- **Educational Evaluation Philosophy.** Tavakol[15] which defines evaluation as a structured process aimed at enhancing educational practices through systematic assessment and judgment. Evaluation in medical education serves to describe and analyse programs, policies, or procedures to improve teaching strategies and learning outcomes. An ideal evaluation process involves multiple phases, including baseline evaluation to identify strengths and weaknesses, formative evaluation for curriculum feasibility, summative evaluation to assess outcomes, and impact evaluation for long-term effects.[8] This holistic approach ensures that all elements of education are continuously refined and adapted to meet evolving academic and professional demands. Curriculum evaluation can be conducted for formative or summative purposes. Formative evaluation fulfils a role in improvement, accountability (measurement of results or efficiency), development, or modification and is conducted as the curriculum is evolving, while evaluation as a summative take places generally to assess the outcomes, the merit, or worth of a particular curriculum.[9]

- **Program Evaluation Types** which categorize evaluation approaches into three distinct types:[2]
  1) Reductionist or Linear Perspective — This model emphasizes a direct cause-and-effect relationship between variables. It is primarily used for formative evaluations where the stability and predictability of a program are central. This approach is particularly effective when evaluating well-defined and structured educational components that do not undergo frequent changes.

2) System Theory – Unlike the reductionist model, the system theory approach views the educational program as an interconnected and dynamic entity. This model considers external influences, feedback mechanisms, and interactions between different curriculum components. It is particularly useful when evaluating curriculum adaptability and integration across multiple disciplines.

3) Complexity Theory – This model recognizes that educational programs exist within an evolving and unpredictable environment. It considers nonlinear interactions, emergent properties, and adaptability within the curriculum. Complexity theory is best suited for evaluating programs that require ongoing modifications, interdisciplinary collaboration, and innovative educational strategies.

By understanding these distinct evaluation approaches, educators and policymakers can select the most appropriate model that aligns with their program's objectives, ensuring a balanced and effective assessment process

These theories support a structured and multi-perspective evaluation approach for medical education curricula. Curriculum evaluation as a formative could provide feedback into the decision-making process at the beginning, during the planning, and implementation phases of curriculum development. Curriculum evaluation as a summative, some of them have developed by Logic model,[13] Kirkpatrick model,[2,16] and context-input-process-product (CIPP) model.[2,9] However, there are other forms of curriculum evaluation in terms of responsive evaluation, portrayal evaluation, transactional evaluation, illuminative evaluation, and holistic evaluation. Responsive evaluation was deemed as an attitude more than a model is a perspective that responds to the teaching and learning program to be evaluated and constitutes a search for quality. This approach is more responsive to issues in the program identified by different people and represents a further development, particularly methodologically, toward a more process-oriented approach.[9]

Transactional evaluation and portrayal evaluation are concepts typically found in educational and evaluative contexts, particularly concerning the assessment of learning outcomes or performance. Transactional evaluation focuses on the interactions between educators (or evaluators) and learners. It assesses how the educational processes, including teaching methods and learner engagement, facilitate or hinder learning. While portrayal evaluation refers to how learning outcomes, performances, or projects are presented and represented. In other words, Portrayal Evaluation focuses on the representation and communication of what has been learned. Both evaluation methods can be integrated to provide a comprehensive evaluation of educational outcomes, giving insight into both the learning journey and the effectiveness of students' communication skills. Illuminative evaluation is a valuable approach for exploring the complexities of educational programs and processes. By focusing on context, using qualitative

methods, and involving stakeholders, it provides insights that can inform improvement and understanding of program effectiveness.[9,17]

## Methods

This work employs a literature review and analysis of existing research on item development, drawing primarily from the author's previously published tools and findings in peer-reviewed journals. It synthesizes conclusions derived from these publications to offer a unique perspective informed by both established literature and the author's direct research experience. The study represents a perspective piece grounded in scholarly inquiry and complemented by practical research insights.

The theoretical framework presented is grounded in fundamental philosophies concerning evaluation – its types, methods, and approaches – and a comprehensive understanding of curriculum, particularly within medical education. This foundation acknowledges that effective medical education necessitates a robust evaluation system. The chosen framework emphasizes alignment between learning objectives, teaching strategies, and assessment methods, ensuring that evaluations accurately measure competency acquisition and identify areas for improvement in both student performance and curriculum design. Ultimately, this theoretical grounding aims to promote continuous quality enhancement within the medical education program, fostering the development of competent and well-rounded physicians.

The success of any educational program, particularly in the complex and demanding field of medicine, hinges on a robust and well-defined evaluation process. This process, encompassing the types, methods, and approaches employed, must be firmly grounded in a philosophical understanding of both evaluation principles and the specific nuances of medical curricula. This essay will outline the theoretical framework underpinning a comprehensive evaluation system designed to assess and improve medical education programs.

At its core, this framework is built upon the understanding that evaluation is not merely a summative exercise, judging the final outcomes. Instead, it is a continuous and iterative process, integral to the development and refinement of the curriculum itself. This aligns with a formative evaluation philosophy, emphasizing ongoing data collection and analysis to inform curriculum adjustments and ensure alignment with learning objectives and societal needs. This philosophy recognizes that medical knowledge and practice are constantly evolving, requiring curricula to be equally dynamic and responsive to change.

## Analysis

The term of evaluation, assessment and appraisal are often used interchangeably. These terms are different in the context of its use. The evaluation focuses on the design, implementation, improvement, or outcomes of a program rather than an individual. While the term assessment refers to the process of individual measurements, such as learners' ability.[18] In other words, evaluation is about reviewing, analysing, and judging the

importance or value of the information gathered by all these assessments. Assessment different from evaluation, they have an aspect in measurement. Assessment is a strategy or method chosen to congregate information needed in order-making judgement or decision.[2,3]

Evaluation, at its core, transcends mere measurement and enters the realm of nuanced judgment. It represents a synthesis of two crucial elements: perceptions of value and estimations of goal achievement.[19] Perceptions of value encompass the underlying principles, ethical considerations, and societal norms that influence our judgment. These values provide the framework for determining what is deemed desirable or beneficial, ultimately influencing the criteria used to evaluate success. Simultaneously, estimations of goal achievement provide empirical grounding to the evaluative process. This involves collecting and analysing data to determine the extent to which pre-defined objectives have been met.[9]

A comprehensive evaluation approach integrates multiple perspectives, including various evaluation models, methodologies, and objectives. Reimann & Schober's framework[8] provides a holistic understanding of evaluation by addressing key aspects:

1. **What should be evaluated?**[15] Here we can take one aspect of a whole of program, for instance, taking subject the curriculum. It could be a part of curricula such as structure curriculum, teaching methods, assessment method and learning outcome. A statement of learning outcome refers to someone/graduates must be able to do something.
2. **What is the frame of reference?** In this regard, we could use a model to evaluate the curriculum, such as the logic model, CIPP model, Kirkpatrick'model,[16,20] reductionist model.[21]
3. **What perspectives should be considered?** It could be the key performance indicator for quality curriculum, namely: acquired expertise (a result of OSCE, accurately of diagnosis), actual performance in the medical school (Grade point average of graduates) and in the workplace (professional behaviour and competence development.[22,23]
4. **Who are the evaluation participants?** In terms of stakeholders, there are some stakeholder groups of medical education, namely: medical students, lecturers, graduates, patients and doctor supervisor.[8]

   A truly comprehensive evaluation approach considers not only the individual components but also their interactions within the broader educational framework. Evaluation subjects may include specific courses, faculty development programs, competency-based training, or entire curricular structures. Contribution analysis[24] can help measure curriculum impact while addressing conflicts of interest among key stakeholders. Furthermore, evaluation should be designed to adapt to the needs of the evolving educational landscape, ensuring that stakeholder perspectives align with institutional objectives.

By utilizing multiple evaluation approaches, including both qualitative and quantitative measures, a well-rounded assessment can be conducted. The integration of diverse evaluation frameworks allows for an in-depth understanding of curriculum effectiveness, ensuring that the evaluation process is not only systematic but also reflective of real-world educational challenges.

Different evaluation approaches should align with evaluation objectives and expected outcomes. Logic models offer systematic evaluation by visually mapping required resources, activities, target groups, and intended outcomes, facilitating stakeholder communication and understanding.[20] However, this approach can be time-consuming and costly.

The Kirkpatrick model, widely used for program evaluation, comprises four levels:[16,25]

1. **Reaction** - Participants' affective responses and satisfaction. It includes collecting input regarding the content, delivery, materials, and general experience of the program. This feedback can be obtained via surveys, questionnaires, and conversations.
2. **Learning** - Quantifiable indicators of knowledge, skills, and attitudes acquisition. This involves measuring learning outcomes through tests, assessments, skill demonstrations, and observations.
3. **Behaviour** - Application of learned skills in professional practice. Assessing behaviour change typically involves continuous observations, feedback from supervisors and peers, and various other approaches to measure how effectively the program, content is being applied in practice.
4. **Results** - The highest level of evaluation; long-term impact on organizational goals. This level looks at the broader impact of training on the organisation's goals and objectives.

A logic model is a tool used to develop and evaluate programs in a systematic manner. In this regard, a logic model is highly appropriate to evaluate the curriculum because it has visual depictions outlining resources required to support program activities, services to be delivered, the target groups that may be surveyed or interviewed, and the intended outcomes and outputs a program proposed. Another strength from the logic model evaluation can open the lines of communication between all stakeholders involved in an educational program, implementation, and evaluation. Thus, it can contribute to creating a common understanding for a diverse group of educational stakeholders to effectively discuss a target issue and possible routes to its solution and knowing that solution has been successfully achieved.[20] However, implementing the logic model can be a lengthy process depending upon the number of stakeholders involved and more costly than other evaluation frameworks.

While the Kirkpatrick model of curriculum evaluation is also developed to evaluate a program. Kirkpatrick's model of four levels criteria is the most popular approach to the evaluation of program or training. There have 4 levels of program outcome, namely: reaction, learning, behaviour, and results. Level one-reaction includes the assessment of participants' affective responses to the

quality (e.g. satisfaction with the instructor) or the relevance of training. Level two-learning measures are quantifiable indicators of the learning that has taken place during the course. Level three-behaviour outcomes address either the extent to which knowledge or skills gained in training are applied on the job. Level four-results are intended to provide some measure of the impact that training or program has had on organizational goals and objectives [16]

From the preceding explanation, it can be concluded that evaluation of Logic model can be assumed as a qualitative phase in an exploratory sequential design-mixed method. Whilst the evaluation model of Kirkpatrick Hierarchy is a quantitative phase to generalize what aspect to be found in the qualitative phase.

**Critique of the Kirkpatrick's Model**
An article written by Reio et al.[26] reveals several points, including:

- Assumptions: Critics argue that the levels are often viewed as hierarchically linked, leading professionals to ignore lower levels and jump to "Results" without confirming "Learning" or "Behaviour."

- Causality: The model implies causality between levels, which lacks empirical support; positive reactions do not necessarily lead to improved performance.

- Measurement Challenges: Many organizations do not utilize all four levels due to complexity and lack of resources. Most focus predominantly on Levels 1 and 2.

## Alternative Models
The writing also reviews other evaluation models that have emerged, which are often adaptations or expansions of Kirkpatrick's framework. These include:

- Brinkerhoff's Six-Stage Model: Emphasizes the importance of pre-training assessment and comprehensive evaluation. It consist a six-stage evaluation model encompassing the following stages: 1) goal setting, 2) program design, 3) program implementation, 4) immediate outcomes, 5) intermediate or usage outcomes, and 6) impacts and value assessment. This model has similarities to Kirkpatrick's framework, but it goes further by adding two initial phases that facilitate a formative assessment of training needs and design. Brinkerhoff's model offers a robust and sequential approach to training evaluation, expanding upon Kirkpatrick's foundational four-level model. Brinkerhoff's stages 3 and 4 correspond to Kirkpatrick's "reaction" and "learning" levels respectively, focusing on participant experience and knowledge acquisition. More significantly, Brinkerhoff refines the higher levels, dedicating stage five to evaluating the transfer of learned skills to the workplace, akin to Kirkpatrick's "behaviour" level. Stage six then assesses the ultimate organizational value derived from the training, directly paralleling Kirkpatrick's "results" level.[7] The key strength of Brinkerhoff's model lies in its emphasis on sequential progression; each stage builds upon the successful completion of the preceding one, creating a more comprehensive and interconnected evaluation framework. This structured approach ensures a clear line of sight between training delivery and tangible organizational impact, providing valuable insights for continuous improvement and return on investment.[25]

- CIPP Model (Context, Input, Process, Product): Focuses on ongoing evaluation throughout the training process rather than only post-training outcomes. The CIPP Model is a comprehensive evaluation framework developed in the 1960s by Daniel Stufflebeam. [27] It aims to improve accountability in educational programs and has been adapted across various sectors. This model emphasizes evaluation for improvement rather than proof and supports both formative and summative evaluations, guiding decision-making.[28] This model is useful for assessing the quality of education in university. The context encompasses the university's goals, objectives, history, and background. Inputs pertain to the materials, time, physical resources, and human resources required for the school to function effectively. The process involves all teaching and learning activities, while the product emphasizes the quality of these activities and their impact and benefits for society.[27,29]

## Implications
This evaluation approach has significant implications for medical education practice, policy formulation, and research. It provides valuable insights for education managers and policymakers, enabling informed decision-making for curriculum improvement.

Challenges may arise from stakeholder resistance, particularly when changes disrupt established norms. However, fostering a culture of continuous improvement and shared commitment can mitigate these barriers. Institutions must cultivate a healthy organizational culture that values constructive feedback, acknowledges contributions, and promotes growth.

An effective evaluation system follows a structured quality management cycle:[30,31]
1. **Planning** - Identifying key evaluation aspects.
2. **Implementation** - Conducting thorough evaluations, even in segments.
3. **Assessment** - Ensuring alignment between evaluation execution and objectives.
4. **Decision-making** - Taking appropriate actions based on evaluation outcomes while mitigating risks.

A well-structured evaluation cycle ensures adequate input, efficient processes, and effective outcomes. The implications of this perspective on best practices in the field of medical education, educational management policies, or even advanced research can certainly provide insights and contributions to best practices as well as feedback to policymakers through contribution analysis. Educational management benefits significantly when they manage and decide on the forms to be taken in addressing managerial issues.

Although barriers to evaluation implementation may arise, challenges from stakeholders present a unique obstacle in managing the evaluation of a program,

especially when ideal conditions are hindered by their reluctance to change due to being in a comfort zone. Once again, commitment is essential to building a framework of togetherness accompanied by a spirit of self-improvement and an abundance mentality. An educational institution or organization must grow and develop with a healthy organizational culture, rich in values, appreciation for hard work and achievements, as well as recognition of contributions.

The assessment results must be communicated to the relevant stakeholders as part of ongoing improvement and strengthening for follow-up actions. Through the Kirkpatrick hierarchy approach, we attempt to conduct a comprehensive evaluation of various assessed aspects/components. A good quality management cycle begins with planning what will be evaluated, followed by conducting a thorough evaluation process, even if its implementation may be segmental, then checking whether the execution aligns with our initial evaluation objectives, and finally taking adequate actions to make the best decisions among the available options while mitigating risks associated with our actions. A good quality cycle will provide adequate input, practice effective and efficient processes, and produce the right outcomes.

## Conclusion

This paper highlights the importance of the Kirkpatrick hierarchical model in conducting comprehensive evaluations in medical education. By integrating multiple theoretical perspectives and evaluation frameworks, stakeholders can achieve a holistic understanding of curriculum effectiveness. Ensuring stakeholder commitment and fostering a culture of continuous improvement are crucial in overcoming evaluation challenges and enhancing the overall quality of medical education.

## Limitations of the Kirkpatrick Model

1. Propensity Towards Lower Levels: Evaluators often focus on the initial levels of the model (reaction and learning), neglecting the higher levels (behaviour and impact), which are crucial for comprehensive evaluation. This practice, while prevalent in the business context, is contradictory in higher education, where many participants have not yet entered the workforce, making behaviour and impact assessments difficult.
2. Rigidity: The model's structured approach does not account for various contextual factors influencing educational outcomes. Critics argue that this rigidity can fail to capture multidimensional outcomes and underlying reasons for observed effects, thus oversimplifying effectiveness.
3. Paucity of Evidence on Causal Relationships: There is insufficient evidence demonstrating causal connections between the levels of the model. While the framework assumes that reactions lead to learning, which then influences behaviour and ultimately impacts the organization, empirical studies often reveal weak or non-existent correlations among these levels. This inadequacy undermines the model's effectiveness as a comprehensive evaluation tool.[32]

Despite its widespread use, the Kirkpatrick Model presents limitations in higher education settings, focusing too much on lower levels, showing rigidity, and lacking evidence for causal relationships among its levels. To enhance its application, the author recommends integrating contextual inputs, employing robust observational methods, and analysing causal links quantitatively.

## References

1. Busari JO. Comparative analysis of quality assurance in health care delivery and higher medical education. *Adv Med.educ Pract*. 2012;3:121-127.

2. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE Guide No. 67. *Med Teach*. 2012;34(5). doi:10.3109/0142159X.2012.668637

3. Ruhe V, Boudreau JD. The 2011 Program Evaluation Standards: A framework for quality in medical education programme evaluations. *J Eval Clin Pract*. 2013;19(5):925-932. doi:10.1111/j.1365-2753.2012.01879.x

4. Oktay C, Senol Y, Rinnert S, Cete Y. Utility of 360-degree assessment of residents in a Turkish academic emergency medicine residency program. *Turkish J Emerg Med*. 2017;17(1):12-15. doi:10.1016/j.tjem.2016.09.007

5. Kusmiati M, Sanip S, Bahari R. The Development of a 360-degree Evaluation Model of Medical Curriculum with the Kirkpatrick Hierarchy Approach. *Educ Med J*. 2024;16(1):93-115. doi:10.21315/eimj2024.16.1.7

6. Sanchez-Reilly S, Ross JS. Hospice and palliative medicine: Curriculum evaluation and learner assessment in medical education. *J Palliat Med*. 2012;15(1):116-122. doi:10.1089/jpm.2011.0155

7. Heydari MR, Taghva F, Amini M, Delavari S. Using Kirkpatrick's model to measure the effect of a new teaching and learning methods workshop for health care staff. *BMC Res Notes*. 2019;12(388):1-5. doi:10.1186/s13104-019-4421-y

8. Spiel C, Schober B, Reimann R. Evaluation of curricula in higher education: Challenges for evaluators. *Eval Rev*. 2006;30(4):430-450. doi:10.1177/0193841X05285077

9. Klenowski V. Curriculum Evaluation: Approaches and Methodologies. *Int Encycl Educ Third Ed*. Published online 2009:335-341. doi:10.1016/B978-0-08-044894-7.00069-5

10. Donnon T, Al Ansari A, Al Alawi S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: A systematic review. *Acad Med*. 2014;89(3):511-516. doi:10.1097/ACM.0000000000000147

11. Ruhe V, D Boudreau. The 2011 Program Evaluation Standards : a framework for quality in medical education programme evaluations. *J Eval Clin Pract ISSN 1365-2753*. Published online 2012:1-8. doi:10.1111/j.1365-2753.2012.01879.x

12. Baxter SK, Blank L, Woods HB, Payne N, Rimmer M, Goyder E. Using logic model methods in systematic review synthesis: Describing complex pathways in referral management interventions. *BMC Med Res Methodol*. 2014;14(1):1-9. doi:10.1186/1471-2288-14-62

13. Cooksy LJ, Gill P, Kelly PA. The program logic model as an integrative framework for a multimethod evaluation. *Eval Program Plann*. 2001;24(2):119-128. doi:10.1016/S0149-7189(01)00003-9

14. Prideaux D. Curriculum development in medical education: From acronyms to dynamism. *Teach Teach Educ*. 2007;23(3):294-302. doi:10.1016/j.tate.2006.12.017

15. Tavakol M, Gruppen LD, Torabi S. Using evaluation research to improve medical education. *Clin Teach*. 2010;7(3):192-196. doi:10.1111/j.1743-498X.2010.00383.x

16. Bates R. A critical analysis of evaluation practice : the Kirkpatrick model and the principle of beneficence. *Eval Program Plann*. 2004;27(2004):341-347. doi:10.1016/j.evalprogplan.2004.04.011

17. Stake RE. Excerpts from: "Program evaluation, particularly responsive evaluation." *J MultiDi Isc Eval*. 2011;7(15):183-201. doi:10.1016/0886-1633(91)90025-S

18. Bleaklay A, Browne J, Ellis K. Quality in Medical Education. In: Swanwick T, ed. *Understanding Medical Education: Evidence, Theory and Practice*. second. John Wiley & Sons; 2014:47-57. doi:10.1007/s40037-014-0113-4

19. Scriven M, Paul M. *Critical Thinking Defined*. Critical Thinking Conference; 1992.

20. Csiernik R, Chaulk P, McQuaid S, McKeon K. Applying the Logic Model Process to Employee Assistance Programming. *J Workplace Behav Health*. 2015;30(19 September):306-323. doi:10.1080/15555240.2014.999078

21. Carraccio CL, Englander R. From flexner to competencies: Reflections on a decade and the journey ahead. *Acad Med*. 2013;88(8):1067-1073. doi:10.1097/ACM.0b013e318299396f

22. Tavakol M, Gruppen L. Using evaluation research to improve medical education. *Clin Teach*. 2010;7:192-196.

23. Kusmiati M, Bahari R, Sanip S, Hamid NAA, Emilia O. The development of an evaluation tool to assess professional behavior and clinical competencies from the graduates' perspective. *Korean J Med Educ*. 2020;32(1):1-11. doi:10.3946/kjme.2020.148

24. Biggs JS, Farrell L, Lawrence G, Johnson JK. A practical example of Contribution Analysis to a public health intervention. *Evaluation*. 2014;20(2):214-229. doi:10.1177/1356389014527527

25. Ambu-Saidi B, Fung, C.Y., Turner, K., Lim, A.S.S. A Critical Review on Training Evaluation Models: A Search for Future Agenda. *J Cogn Sci Hum Dev*. 2024;10(1):142-170. doi:10.33736/jcshd.6336.2024

26. Reio TG, Rocco TS, Smith DH, Chang E. A Critique of Kirkpatrick's Evaluation Model. *New Horizons Adult Educ Hum Resour Dev*. 2017;29(2):35-53. doi:10.1002/nha3.20178

27. Stufflebeam DL. The CIPP Model for Evaluation. In: *International Handbook of Educational Evaluation*. ; 2003:31-62. doi:10.1007/978-94-010-0309-4_4

28. Darma IK. The effectiveness of teaching program of CIPP evaluation model. *Int Res J Eng IT Sci Res*. 2019;5(3):1-13. doi:10.21744/irjeis.v5n3.619

29. Aziz S, Mahmood M, Rehman Z. Implementation of CIPP Model for Quality Evaluation at School Level: A Case Study. *J Educ Educ Dev*. 2018;5(1):189. doi:10.22555/joeed.v5i1.1553

30. Chen Y, Li H. Research on Engineering Quality Management Based on PDCA Cycle. *IOP Conf Ser Mater Sci Eng*. 2019;490(6). doi:10.1088/1757-899X/490/6/062033

31. Bunglowala A, Asthana N. A Total Quality

Management Approach in Teaching and Learning Process. *Int J Manag.* 2016;7(5):223-227. http://www.iaeme.com/MasterAdmin/uploadfolder /IJM_07_05_021/IJM_07_05_021.pdf%0A

32. Cahapay M. Kirkpatrick Model: Its Limitations as Used in Higher Education Evaluation. *Int J Assess Tools Educ.* 2021;8(1):135-144. doi:10.21449/ijate. 856143