



Applications of the Kirkpatrick Model in Post-secondary Health Sciences Education: A Scoping Review

Natasha L. Gallant¹, Elizabeth Oddone Paolucci², Chelsea L. Russill¹, Ray Jewett^{3,4}, Chantelle Recsky⁵, Katherine Ford⁶, Dina Idriss-Wheeler⁷, Victrine Tseung⁸, Hina Ansari⁹, Zeenat Ladak¹⁰, Aida Fernandes¹¹, Deborah A. Marshall^{12*}

Abstract

Background: The Kirkpatrick model is commonly used as a systematic approach to evaluate training programs, although its application to health sciences experiential learning programs is not well-established. To inform the use of the Kirkpatrick model in the evaluation of the Canadian Institutes of Health Research's (CIHR's) Health System Impact Fellowship National Cohort Training Program (HSIF NCTP), we examined its application in post-secondary health sciences programs.

Methods: Using the Joanna Briggs Institute's updated methodology for scoping reviews, we searched CINAHL, EMBASE, ERIC, MEDLINE, PsycINFO, and Web of Science for studies published from 2017 to 2023 that focused on health sciences experiential learning programs held at universities and reported on at least one level of the Kirkpatrick model (ie, reaction, learning, behavior, results). We extracted data on study characteristics and reported outcomes for each of the Kirkpatrick model levels.

Results: After deduplication, we screened 755 titles and abstracts, we reviewed 97 full texts, and we included 34 studies in our scoping review. Many studies reported outcomes at the reaction or learning levels followed by the behaviour and results levels. Across levels, despite identifying several areas of improvement, learners typically reported favourable perceptions, increased confidence and knowledge, improved performance, and organizational improvements.

Conclusion: The Kirkpatrick model is a widely used and highly adaptable evaluation model that has been successfully used to evaluate a range of post-secondary health sciences programs. Despite its wide use, evaluators using the Kirkpatrick model should use more robust methodologies to capture long-term behaviour and results associated with the programs. Future work should focus on evaluating a broader spectrum of programs such as doctoral- and post-doctoral-level experiential learning programs and underrepresented healthcare professions such as psychologists and dietitians. Integration of behaviour change and implementation science methodologies within the broader educational evaluation literature is also needed.

Registry Name and Number: Open Science Framework, <https://osf.io/5xqvg>

Keywords: Education, Evaluation, Experiential Learning, Healthcare Professions, Health Systems, Impact

Copyright: © 2026 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Gallant NL, Oddone Paolucci E, Russill CL, et al. Applications of the Kirkpatrick model in post-secondary health sciences education: a scoping review. *Int J Health Policy Manag.* 2026;15:8857. doi:10.34172/ijhpm.8857

Article History:

Received: 18 October 2024

Accepted: 30 March 2026

ePublished: 6 April 2026

*Correspondence to:

Deborah A. Marshall

Email:

damarsha@ucalgary.ca

Introduction

Post-secondary programs aim to prepare learners to use the theoretical knowledge and practical skills they have gained in the classroom to address evolving real-world challenges within a professional environment.¹ Many post-secondary programs will enhance real-world learning opportunities through experiential learning and other educational approaches (eg, project-based learning, service learning).¹ As an example, the Health System Impact Fellowship National Cohort Training Program (HSIF NCTP), funded through a grant from the Canadian Institutes of Health Research (CIHR) Institute for Health Services and Policy Research in 2019, is an experiential learning program that provides doctoral (PhD) students, post-doctoral fellows, and early career researchers the opportunity to become an embedded researcher within a health system organization. The HSIF

NCTP was designed in response to escalating complexities within the healthcare sector by equipping learners with the requisite skills to navigate and address multifaceted real-world and professional workplace challenges.² Aligned with this objective, the HSIF NCTP emphasizes cultivating a diverse set of enriched core competencies (eg, leadership, mentorship and collaboration, analysis and evaluation of health and health-related programs and policies; understanding and comparing healthcare systems and the policy-making process).² These enriched core competencies generate standard professional development milestones for a cohort of trainees and scientists from diverse disciplines such as computer science, engineering, geography, medicine, nursing, occupational therapy, pharmacy, physiotherapy, and psychology. The educational themes of the HSIF NCTP are broadly centered around project management and leadership;

patient-oriented research; equity, diversity, and inclusion; demonstrating impact through knowledge translation; building sustainability; and implementation science.²

Many evaluations of educational programs focus on measuring knowledge acquired,³⁻⁵ but health education evaluations increasingly emphasize measurement of what is learned and practical application (ie, results) in the healthcare system (eg, Competency-Based Education⁶). The evolving emphasis on effective evaluation in health sciences education is driven by a myriad of pressing challenges and unique demands faced by healthcare professionals (eg, rapid technological advancements⁷; staffing shortages⁸; emphasis on interprofessional collaboration⁹; increasing healthcare service demand¹⁰). The dynamic landscape of healthcare systems, coupled with evolving patient needs and technological advancements, underscores the importance of knowledge- and skill-based continued education for healthcare professionals to ensure optimal patient outcomes and safety.¹¹ Health sciences education is collaborative, where professionals from diverse disciplines contribute unique perspectives and expertise. This interdisciplinary environment presents both enriching opportunities and complexities within the teaching and learning environment. Amidst these dynamics, institutions are grappling with resource constraints and competing priorities.^{12,13} Educational activities are prioritized when they yield optimal outcomes aligned with learner and organizational goals and can enhance quality of patient care.¹⁴ Rigorous evaluation of educational programs is foundational for the judicious allocation of educational resources. Evaluating educational programs and interventions also ensures learners are equipped with the requisite attitudes, knowledge, and skills to contribute to the healthcare system workplace.¹⁵⁻¹⁷

Evaluation frameworks (eg, Competency-Based Education¹⁸; Kern's Six-Step Approach¹⁹; Kirkpatrick model²⁰; Miller's Pyramid²¹; Outcome-Based Education²²) have emerged to assess the effectiveness of educational programs within health sciences education. These frameworks offer mechanisms for assessing educational programs, extending beyond mere knowledge acquisition to encompass practical application and impact across domains relevant to the healthcare sector. The Kirkpatrick model stands out as a widely adopted tool for evaluating educational training programs, particularly within post-secondary health sciences and related disciplines.^{20,23-25} The model delineates four levels (ie, *reaction*, *learning*, *behaviors*, and *results*) that each serve a distinct purpose in assessing program outcomes. At the *reaction* level, immediate learner responses provide insights into engagement and satisfaction, while the *learning* level evaluates the acquisition of knowledge, skills, and attitudes. The *behavior* level assesses the practical application of acquired knowledge in real-world scenarios, ensuring observable changes in behavior, and the *results* level, which presents the greatest measurement challenge,²⁶ extends the model to encompass organizational outcomes, thus linking program success with broader organizational goals and objectives.^{20,25}

The popularity of the Kirkpatrick model stems from its

comprehensive nature and adaptability, making it a valuable resource for evaluating program effectiveness and alignment with organizational goals. The model's hierarchical structure posits a reinforcing relationship between its levels whereby outcomes at one level inform the next. For instance, favorable reactions are postulated to contribute to improved learning outcomes, thereby reinforcing the model's relevance for institutional effectiveness and strategic planning.^{20,25} This integrated approach facilitates the identification of successful programs and potential unintended consequences of curriculum changes, ensuring strategic alignment. The model's evolution, exemplified by the introduction of the New World Kirkpatrick Model in 2016, underscores its responsiveness to evolving demands, ensuring continued efficacy in evaluating and highlighting organizational impact across various programs.²⁰ The Kirkpatrick model also has limitations. It assumes a linear progression across the four levels, potentially overlooking the dynamic and iterative nature of the education process, where feedback loops influence each level.^{26,27} The model's emphasis on quantitative data may neglect the intricacies of educational outcomes, and its assumption of causal relationships between educational initiatives and outcomes may overlook contextual variables, such as prior knowledge or the organization, industry, or cultural environments.^{26,28} The model may inadequately address the evaluation of informal learning or on-the-job experiences, both of which significantly contribute to skill development.²⁸ Variations in methodological approaches to its application have also led to limitations in the model's effectiveness.²⁰ Despite these limitations, the Kirkpatrick model remains the preferred option due to its practical applicability and versatility. The dynamic nature of healthcare systems focused on achieving optimal patient outcomes, necessitates rigorous evaluation of educational programs, highlighting the critical role of frameworks like the Kirkpatrick model.

Purpose and Research Questions

We conducted a preliminary search of Cochrane Database of Systematic Reviews, Google Scholar, JBI Evidence Synthesis, Open Science Framework, and PROSPERO. Several systematic and scoping reviews of the use of the Kirkpatrick model across several disciplines (eg, business, nursing, physics) have been published,²⁹⁻³¹ but none of these reviews examined the ways in which the Kirkpatrick model was specifically used to evaluate post-secondary health sciences programs. Thus, our purpose was to examine how the Kirkpatrick model is used in evaluating post-secondary health sciences programs to identify and analyze knowledge gaps. By identifying these gaps, we aimed to guide practices in evaluation studies. More specifically, our aim was to answer the following research questions: (1) What are the characteristics of studies using the Kirkpatrick model to evaluate post-secondary health sciences programs? (2) In which way has the Kirkpatrick model been applied to evaluate post-secondary health sciences programs? (3) What outcomes are associated with each of the Kirkpatrick model levels?

Methods

We conducted this scoping review using the Joanna Briggs Institute's updated methodology for scoping reviews³² and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR).³³ We chose a scoping review methodology because the primary purpose of this review was to consolidate existing literature, reveal patterns across existing literature, and identify gaps in the current literature, which was appropriate for our purpose given the anticipated diverse study designs and types of evidence that we would identify.³² The inherent adaptability of scoping reviews was seen as highly accommodating to the diverse study designs and types of evidence that were anticipated.³² We carried out our scoping review in several phases: (1) development of our review protocol and research questions, (2) systematic search for relevant studies, (3) study selection based on pre-established eligibility criteria, (4) data extraction, and (5) interpretation of results from the extracted data. Given the broader aims of scoping reviews,³² we did not perform a risk of bias or methodological quality assessment on the included studies. We registered our scoping review protocol with Open Science Framework (<https://osf.io/5xqvqg>).

Search Strategy

We developed a systematic search strategy with support from a librarian at the University of Regina (CB). We used the following search terms (ie, CINAHL, EMBASE, ERIC, MEDLINE, PsycINFO, Scopus): “Kirkpatrick evaluat*,” “Kirkpatrick model*,” “exp education, premedical” or “exp education, professional” or “exp schools, health occupations” or “exp universities,” “universit* or colleg* or postsecondar* or school* or educat* or residen*.” We completed the final search strategy was run on February 20, 2023 and February 21, 2023 (See [Supplementary file 1](#)).

Eligibility Criteria

Time Period. We included studies published in 2016 onwards and excluded studies published before 2016. We only considered studies published in 2016 onwards to only include studies using the most recent edition of the Kirkpatrick model.²⁰ The most recent edition of the Kirkpatrick model comprises new constructs within each level, including engagement and relevance for the reaction level, confidence and commitment for the learning level, required drivers for the results level, and leading indicators for the impact level.

Language. We included studies published in English and excluded studies published in any language besides English. Evidence suggests that language constraints do not significantly impact review outcomes³⁴; thus, our decision to exclude non-English studies presents minimal risk of bias.

Publication Method. We included peer-reviewed journal articles and books published through a scholarly press and excluded journal articles that were not peer reviewed, books published through a non-scholarly press, theses and dissertations, and conference proceedings.

Study Design. We included any empirical study using

original data (eg, quantitative, qualitative, mixed methods, or case studies) and excluded non-empirical studies (eg, opinion pieces, commentaries) or studies not using original data (eg, reviews).

Study Population. We included studies with data on learners and/or educators and excluded studies with data on non-learners and non-educators.

Education Setting. We included studies examining programs within for-credit accredited university settings at the bachelor's, master's, doctoral, post-doctoral, post-graduate, or residency (eg, MD, PharmD) levels and excluded studies examining programs within non-credit, non-accredited, or non-university settings (eg, college diploma, post-graduate certification, Journey person certificate). The decision to focus on programs offering academic credit(s) aligns with our goal to ascertain the functions of the Kirkpatrick model when integrated into accredited university settings at all levels with structured courses and assessments.

Educational Fields. Our selection of educational fields was guided by the International Standard Classification of Education 2011.³⁵ Educational fields that were eligible included education, humanities, social and behavioral science, journalism and information, business and administration, law, life sciences, physical sciences, mathematics and statistics, computing, engineering and engineering and trades, architecture and building, health, social services, environmental protection, and security services. Educational fields that were ineligible included manufacturing and processing, agriculture, forestry and fishery, veterinary, personal services, and transportation services.

Evaluation Framework. We included studies that reported on at least one level of the Kirkpatrick model (ie, *reaction, learning, behavior, results*)²⁰ and excluded studies that did not report outcomes related to at least one level of the Kirkpatrick model. It was acceptable for studies to apply evaluation models in addition to the Kirkpatrick model.

Study Selection

We uploaded all studies identified through our search strategy into Covidence.³⁶ After Covidence automatically deduplicated the retrieved studies, two reviewers (BDR, CLR) independently screened the titles and abstracts and reviewed the full texts for the first ten studies to pilot the screening and reviewing processes. Following this pilot, the two reviewers discussed and reached a consensus regarding any changes to the screening and/or reviewing protocols. Next, the same two reviewers independently screened titles and abstracts and reviewed full texts for eligible studies. One of the reviewers (CLR) did a manual citation search of all eligible studies that resulted in the identification of nine additional studies that were considered at the screening stage. While reviewing full texts, the two reviewers documented reasons for exclusion. At the screening and reviewing stages, the two reviewers resolved disagreements through consensus. As shown in [Figure 1](#), our search strategy yielded 1423 studies, of which 668 were duplicates, with 34 studies included in our scoping review.

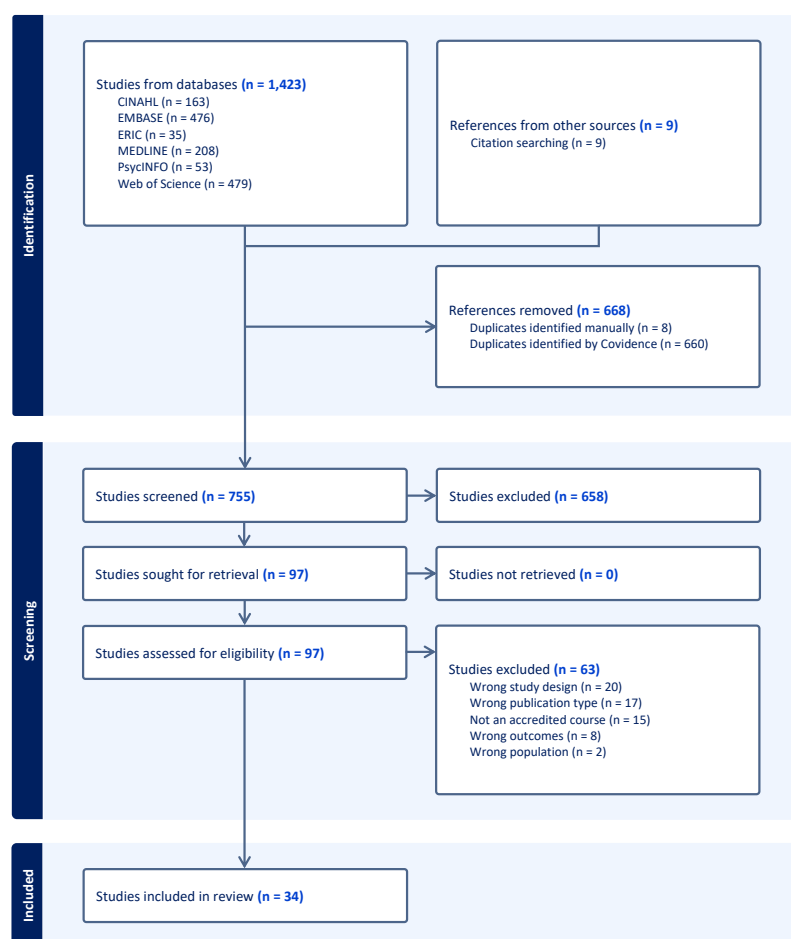


Figure 1. PRISMA-ScR Flow Diagram. Abbreviation: PRISMA-ScR, Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews.

Data Extraction

Based on our scoping review's research questions and with feedback from other authors, one of the authors (RJ) created a data extraction process in Google Forms. To pilot the data extraction process, two independent reviewers (CLR, BDR) extracted data from the first 10 studies. Following this pilot, the two reviewers discussed and reached a consensus regarding any changes to the data extraction process. One reviewer (CLR) extracted data for the remaining studies. Extracted data included authors, year of publication, study design, sample (learners or educators with, if appropriate, comparator groups), study purpose, education level, education field, education content, and the incorporation and definition of experiential learning and/or other educational approaches. We defined experiential learning as the process of learning through experience and through reflecting on this experience. We considered hands-on and related learning approaches as a form of experiential learning if it also encompassed learners reflecting on their work.³⁷ Additionally, educational approaches that addressed the National Cohort Training Program core competencies were extracted. Evaluation models employed beyond the Kirkpatrick model, relevant study outcomes, and any programming adjustments proposed or implemented based on Kirkpatrick model outcomes were also extracted.

Results

Study Characteristics

We present a summary of study characteristics in [Table](#).

Year of Publication. The number of studies published increased each year with the peak number of studies published in 2021 ($n = 9$, 27%)^{42,45,46,52,55,60,63,64,71} and 2022 ($n = 9$, 27%).^{39,44,48,54,57,58,66,67,69} Only two (6%) studies were published in 2023,^{41,61} but we only ran the search strategy only a couple of months into 2023.

Country. Studies were predominantly conducted in the United States ($n = 14$, 41%).^{38,42,44,47-49,53,54,58,59,61-64} Studies conducted in other countries, including Australia,^{50,56,69} Canada,^{55,60} India,^{51,57,68} Italy,⁶⁶ Spain,^{43,67} Sweden,⁴⁰ Switzerland,⁴⁶ Taiwan,⁶⁵ The Netherlands,⁷⁰ Turkey,³⁹ United Arab Emirates,^{45,52,71} and United Kingdom,⁴¹ ranged from 1 to 3 studies.

Study Design. Most studies involved cross-sectional³⁸⁻⁵² ($n = 15$, 44%) or cohort⁵³⁻⁶³ ($n = 13$, 38%) designs. Outside of these common designs, three studies (9%) employed a mixed methods design,⁶⁶⁻⁶⁸ one study (3%) was a retrospective mono-centered design,⁶⁹ another study (3%) was a case series,⁷⁰ and a final study (3%) was a controlled trial.⁷¹

Education Level. The sample included in each study was predominantly studying at the undergraduate level ($n = 12$, 35%)^{39,42,45,46,51,52,57,60,67-69,71} or PhD or other residency levels

Table. Study Characteristics

Study	Year	Country	Study Design	Education Level	Education Field	Education Content	Sample
38	2018	United States	Cross-sectional	PhD/Doctorate	Health sciences	Pharmacy	Learners, educators
39	2022	Turkey	Cross-sectional	Undergraduate	Health sciences	Medical education	Learners
40	2017	Sweden	Cross-sectional	PhD/Doctorate	Health sciences	Implementation science	Learners, comparator group
41	2023	United Kingdom	Cross-sectional	Continuing education	Clinical skills	Medical education	Learners
42	2021	United States	Cross-sectional	Undergraduate	Social sciences	Adaptive leadership	Learners
43	2020	Spain	Cross-sectional	Master's	Health sciences, clinical skills	Palliative care	Learners, comparator group
44	2022	United States	Cross-sectional	PhD/Doctorate	Health sciences, clinical skills	Medical education	Learners
45	2021	United Arab Emirates	Cross-sectional	Undergraduate	Health sciences	Medical education	Learners
46	2021	Switzerland	Cross-sectional	Undergraduate	Clinical skills	Medical education	Learners
47	2019	United States	Cross-sectional	Master's	Health sciences	Public health	Learners
48	2022	United States	Cross-sectional	PhD/Doctorate	Health sciences, clinical skills	Medical education	Learners
49	2020	United States	Cross-sectional	PhD/Doctorate	Clinical skills	Medical education	Learners
50	2019	Australia	Cross-sectional	Continuing education	Health sciences	Interprofessional clinical health service redesign	Learners
51	2020	India	Cross-sectional	Undergraduate	Health sciences, clinical skills	Medical education	Learners, educators
52	2021	United Arab Emirates	Cross-sectional	Undergraduate	Health sciences, clinical skills	Medical education	Learners, comparator group
53	2017	United States	Cohort	Master's	Health sciences, clinical skills	Nursing	Learners
54	2022	United States	Cohort	PhD/Doctorate	Health sciences, clinical skills	Medical education	Learners, comparator group
55	2021	Canada	Cohort	Graduate or second-entry undergraduate	Health sciences	Interprofessional pain management	Learners
56	2016	Australia	Cohort	PhD/Doctorate	Health sciences	Interprofessional healthcare	Learners, comparator group
57	2022	India	Cohort	Undergraduate	Health sciences, clinical skills	Medical education	Learners, educators
58	2022	United States	Cohort	PhD/Doctorate	Health sciences	Medical education	Learners
59	2018	United States	Cohort	PhD/Doctorate	Health sciences	Medical education	Learners, educators, others
60	2021	Canada	Cohort	Undergraduate	Health sciences	Pharmacy	Learners, educators
61	2023	United States	Cohort	Master's	Health sciences, clinical skills	Interprofessional geriatrics	Learners
62	2019	United States	Cohort	PhD/Doctorate	Health sciences	Interprofessional medication therapy management	Learners
63	2021	United States	Cohort	PhD/Doctorate	Health sciences, clinical skills	Medical education	Learners, educators, others
64	2021	United States	Cohort	PhD/Doctorate	Health sciences	Medical education	Learners
65	2018	Taiwan	Cohort	Master's	Library or information science	Information organization	Learners, comparator group
66	2022	Italy	Mixed methods	Master's	Clinical skills	Nursing	Learners
67	2022	Spain	Mixed methods	Undergraduate	Health sciences	Nursing	Learners, educators
68	2020	India	Mixed methods	Undergraduate	Health sciences	Medical education	Learners, comparator group
69	2022	Australia	Retrospective mono-centred	Undergraduate	Health sciences	Dentistry	Learners
70	2020	The Netherlands	Case series	Master's	Health sciences	Interprofessional quality and safety in patient care	Learners
71	2021	United Arab Emirates	RCT	Undergraduate	Health sciences	Medical education	Learners, comparator group

Abbreviations: PhD, Doctor of Philosophy; RCT, randomized controlled trial.

Note. Two studies did not clearly report on the education level of all study participants^{56,62}

(eg, MD, PharmD; $n = 12$, 35%).^{38,40,44,48,49,54,56,58,59,62-64} Fewer studies evaluated master's students ($n = 7$, 21%),^{43,47,53,61,65,66,70} continuing education ($n = 2$, 6%),^{41,50} or graduate and second-entry undergraduate trainees ($n = 1$, 3%).⁵⁵ Two studies (6%) did not clearly report the education level of all participants in their sample.^{56,62}

Educational Field. More than half of the studies incorporated health sciences curricula ($n = 18$, 53%),^{38-40,45,47,50,55,56,58-60,62,64,67-71} with four studies (12%) focused on clinical skills.^{41,46,49,64} Several studies addressed a combination of these topics ($n = 10$, 29%),^{43,44,48,51-54,57,61,63} while others covered topics related to social sciences ($n = 1$, 3%)⁴² or library and information sciences ($n = 1$, 3%).⁶⁵

Educational Content. Half of the curricula were medical education ($n = 17$, 50%),^{39,41,44-46,48,49,51,52,54,57-59,63,64,68,71} while three studies (9%) exclusively assessed educational programming for nursing students.^{53,66,67} Several studies adopted an interprofessional approach, of which one study (3%) evaluated an implementation science curriculum⁴⁰ and another study ($n = 1$, 3%) examined principles of adaptive leadership.⁴² Other interprofessional studies were oriented towards students across healthcare professions, and included themes such as clinical health service redesign ($n = 1$, 3%),⁵⁰ pain management ($n = 1$, 3%),⁵⁵ palliative care ($n = 1$, 3%),⁴³ geriatrics ($n = 1$, 3%),⁶¹ public health ($n = 1$, 3%),⁴⁷ and quality and safety in patient care ($n = 1$, 3%).⁷⁰ One study (3%) explored interprofessional healthcare among allied health profession students⁵⁶ and another ($n = 1$, 3%) focused on medication therapy management for multiple professions.⁶²

Finally, one study (3%) evaluated a curriculum developed for dentistry students,⁶⁹ while another study assessed an information organization course ($n = 1$, 3%).⁶⁵

Sample. Most studies primarily focused on outcomes related to learners ($n = 27$, 79%)^{39-50,52-56,58,61,62,64-66,68-71} (of which eight studies [24%] included a comparator group)^{40,43,52,54,56,65,68,71} and five studies (15%) assessed both learners and educators.^{38,51,57,60,67} In addition, two studies (6%) involved individuals in supporting roles, such as clinical support staff, community partners, and program coordinators.^{59,63}

Evaluation Methodologies for Kirkpatrick Model Levels

We present a summary of the inclusion of each of Kirkpatrick model levels along with evaluation methodologies used in each level in Figure 2. Of note, two studies (6%) in this review did not report how the chosen strategies were applied to each level of the model, resulting in unclear reported outcomes for each level. These studies similarly conducted interviews with participants ($n = 1$, 50%)⁶⁶ or assessed outcomes through practical or written assessments or surveys ($n = 2$, 100%).^{53,66} Notably, another study combined analyses of behavior and results, making it unclear how to disentangle the findings from each level ($n = 1$, 3%).⁴¹

Level One: Reaction. Most studies assessed the initial reaction and engagement in the evaluated educational programs ($n = 30$, 88%),^{38-52,54,56-65,67-69,71} and primarily used self-report surveys ($n = 27$, 90%).^{38-43,45-52,54,56-58,60-65,68,69,71} Some studies conducted interviews with participants ($n = 3$, 10%),^{44,59,60} while others used formal practical or written

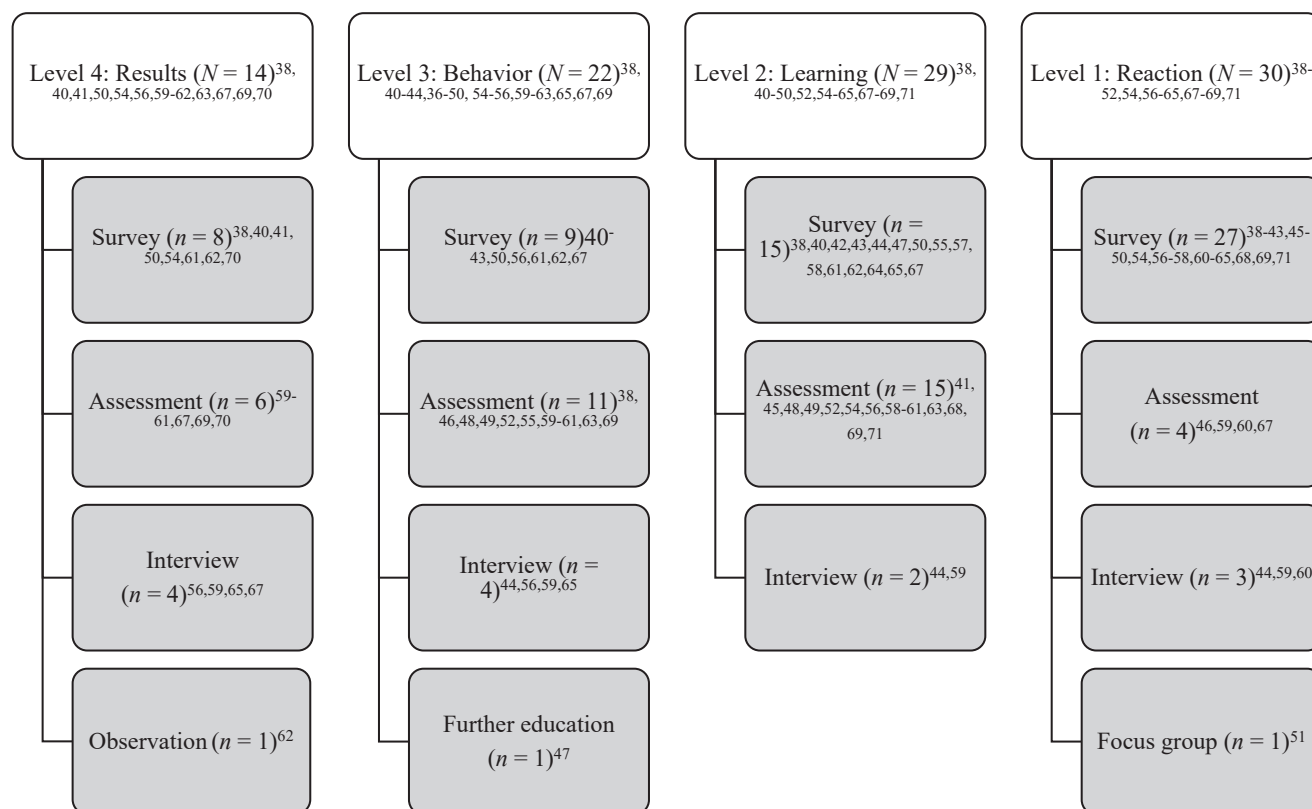


Figure 2. Number of Studies Using Specified Evaluation Methodologies Across Each of the Kirkpatrick Levels. *Note.* Some studies employed multiple evaluation methodologies per Kirkpatrick level.

assessments ($n=4$, 13%).^{46,59,60,67} One study facilitated a post-course group discussion involving all faculty members engaged in the program ($n=1$, 3%).⁵¹

Level Two: Learning. The second level of the model was assessed in almost all studies ($n=29$, 85%).^{38,40-50,52,54-65,67-69,71} Of these, approximately half administered surveys to assess participant learning ($n=15$, 52%),^{38,40,42,43,46,47,50,55,57,58,61,62,64,65,67} while others evaluated performance through written or practical assessments ($n=15$, 52%).^{41,45,48,49,50,54,56,58-61,63,68,69,71} Two studies (7%) used participant interviews to address the second level of the model.^{44,59}

Level Three: Behavior. Compared to levels one and two, fewer studies evaluated the behaviors level ($n=22$, 65%).^{38,40-44,46-50,54-56,59,60-63,65,67,69} Similar to the previous two levels, behaviors were frequently evaluated via written or practical assessments ($n=11$, 50%)^{38,46,48,49,54,55,59-61,63,69} or self-report surveys ($n=9$, 41%).^{40-43,50,56,61,62,67} Select studies conducted interviews with participants ($n=4$, 18%),^{44,56,59,65} while one demonstrated this level through students' pursuit of additional learning outside the course ($n=1$, 5%).⁴⁷

Level Four: Results. The results level was the least frequently reported facet of the model, with just under half of the studies including it as an outcome ($n=14$, 41%).^{38,40,41,50,54,56,59-62,65,67,69,70} Several of these studies measured impact through self-report surveys administered to participants in the training program ($n=7$, 50%)^{38,40,41,50,51,61,70} or organizational participants (ie, the patients of healthcare professions students; $n=1$, 7%)⁶² and slightly fewer used formal written or practical assessments ($n=6$, 43%).^{59-61,67,69,70} Four studies (29%) incorporated interviews with participants for this level.^{56,59,65,67} One study conducted interviews with organizational informants and reviewed documents related to program outcomes (eg, newspaper articles on successful projects; $n=1$, 7%),⁵⁶ and one study employed informal observation of the organizational environment following program implementation ($n=1$, 7%).⁶²

Outcomes Associated with Kirkpatrick Model Levels

Level One: Reaction. Across all studies that evaluated health sciences education above the undergraduate level (ie, continuing education, master's, PhD or other doctorate programs), participant reactions demonstrate predominantly positive perceptions of the curricula, despite reporting several areas for improvement. For instance, students frequently regarded the courses as valuable, applicable, and satisfying ($n=14$, 41%),^{38,40,43,44,47,48,50,53,54,58,59,61,63,64} reported that the content and formats were engaging ($n=2$, 6%),^{38,53} and that the program would support their future work ($n=1$, 3%).⁵⁰ Some students appreciated sessions that encouraged collaboration and interaction with their colleagues ($n=2$, 6%)^{56,62} and preferred sessions that facilitated active learning ($n=1$, 3%).⁶² Others commented on the value of learning ($n=4$, 12%) from their professors and mentors,^{38,40,59,61} as well as the enhancement of skill competence and confidence following the curriculum ($n=1$, 3%).⁴⁸ Studies that examined the perspectives of faculty or support staff yielded similar outcomes, as they also emphasized the importance of the mentor-mentee relationship ($n=1$, 3%),⁵⁹ perceived

improvements in students' skills and knowledge ($n=3$, 9%),^{38,59,63} and found the courses engaging and useful ($n=1$, 3%).⁵⁹ Suggestions for improvement frequently pertained to the structure, format, and timing of the course. Participants expressed concerns about the course timing, tight timelines, and the demanding nature of the curriculum ($n=3$, 9%),^{38,50,54} and prolonged gaps between sessions ($n=1$, 3%).⁵⁸ Additionally, other students noted a lack of support and clarity from instructors about course material and assignments ($n=3$, 9%).^{47,58,59} Faculty and support staff expressed a need for improved processes in pre-work preparation (eg, grading, class presentations) and grading uniformity ($n=1$, 3%)³⁸ and mentorship support and communication ($n=1$, 3%).⁵⁹ Several studies noted participants' desire for increased opportunities to apply their learnings in a real-world setting ($n=2$, 6%).^{40,53}

Level Two: Learning. All studies that assessed learning outcomes reported positive results for students, albeit to varying degrees. Most of these studies reported that the curriculum enhanced students' learning and knowledge ($n=12$, 35%).^{38,40,43,44,47,48,50,53,54,56,62,65} Several studies reported variable effects across their curricula. For instance, while Pfeifle et al found significant effects on learning within their program, they also noted high baseline scores, which may have limited the observable effects.⁶¹ In some cases, there was variability in skill acquisition, where students displayed increased confidence or knowledge, but this was not always reflected in their performance ($n=1$, 3%),⁵⁸ nor was it perceived as such by faculty ($n=1$, 3%).⁶³ Concerns about information retention and application were also highlighted. In these studies, some students expressed concerns about maintaining acquired knowledge over time ($n=1$, 3%).⁵⁸ This concern was echoed by community partners who suggested implementing practices to promote deeper learning ($n=1$, 3%).⁵⁹

Level Three: Behavior. At this level, reported outcomes indicated positive behavioral changes among participants. Results from course surveys demonstrated that many students either already had or expressed increased confidence in applying their acquired knowledge ($n=6$, 18%).^{40,43,47,50,53,59} Analysis of examinations or assignments demonstrated behavioral changes pre- to post-curriculum ($n=1$, 3%),³⁸ as well as the achievement of course objectives ($n=1$, 3%).⁶¹ Evaluation by instructors and observation of skill implementation suggested increased competency ($n=2$, 6%)^{48,59} and perceived value of the achieved skills ($n=1$, 3%).⁴⁴ In one study, a comparison of standardized exam scores between a three-year and four-year cohort revealed no significant differences, suggesting the program's effectiveness compared to the pre-intervention course ($n=1$, 3%).⁵⁴ Similarly, Sieplinga et al reported comparable results based on competency ratings by clinical support staff and attendings, showing no significant differences before and after the curricular intervention ($n=1$, 3%).⁶³ Additionally, surveys in one study indicated that patients were satisfied and had positive experiences with student teams ($n=1$, 3%),⁶² while another study suggested that implementation of the new program successfully expanded and strengthened

multidisciplinary learning experiences ($n=1$, 3%).⁵⁶

Level Four: Results. Studies frequently reported positive impacts of the training; however, identifying and implementing long-term impacts posed challenges. Some students perceived the course as complementing their learning ($n=1$, 3%)³⁸ and helping them obtain and improve skills ($n=3$, 9%).^{59,61,70} Supporting staff and supervisors likewise identified the benefits of the program on students and their respective organizations ($n=2$, 6%).^{54,59} One study assessed results through participants' sharing of major projects (eg, conference presentation, publication, distributed outside their institution) and found that only 8% of participants did not share their work ($n=1$, 3%).⁵⁰ Another study distributed surveys to program directors to compare workplace-based performance outcomes between three-year and four-year medical degree students. The returned surveys demonstrated no significant differences between the cohorts, signifying that the two programs were comparable ($n=1$, 3%).⁵⁴ Craig et al found that there was some level of organizational improvements, such as continued use of a student project (eg, newly developed procedures in local healthcare services) following program cessation, in just over half of the interprofessional learning teams ($n=1$, 3%).⁵⁶ In addition, Schussel et al identified several organizational benefits of their program, including expanding allied professions' roles, hiring a team of students following the program, and implementing additional staff training ($n=1$, 3%).⁶² Despite indications of improvement, ascertaining the actual benefits of the programs are limited. Some students noted uncertainty about the long-term sustainability of programs ($n=1$, 3%).⁷⁰ or the authors did not collect patient outcomes or organizational data ($n=1$, 3%).⁶¹ Alternatively, Carlford et al reported that while students believed the course content applied to their research, some mentioned it may not be as valuable for those in non-research-based settings ($n=1$, 3%).⁴⁰

Other Evaluation Models

Seven ($n=7$, 21%) of the studies integrated evaluation models in addition to the Kirkpatrick model, including the analyze, design, develop, implement, and evaluate (ADDIE) model;⁴⁵ context, input, process, and product (CIPP) model⁶⁴; Freeth/Kirkpatrick model^{56,61,62}; and the strengths, weaknesses, opportunities, and threats (SWOT) model.^{44,53}

Educational Approaches

As shown in Figure 3, five studies (15%) explicitly incorporated experiential learning into their curriculum in some format,^{42,50,60,66,70} such as through the implementation of specialized projects ($n=1$, 3%)⁷⁰ or practical training sessions ($n=1$, 3%).⁶⁶ Several studies explicitly incorporated other established educational approaches including constructivist ($n=2$, 6%)^{48,55} or socio-constructivist learning theory ($n=1$, 3%)⁶³ and the Gibbs model of reflection ($n=1$, 3%).⁷¹ Other explicitly incorporated educational approaches comprised of integrated problem ($n=1$, 3%)⁴⁰ or project-based ($n=1$, 3%)⁵⁰ learning, service learning ($n=1$, 3%),⁶⁷ and active learning ($n=3$, 9%).^{51,58,69} Among the studies that explicitly incorporated experiential learning or other educational approach ($n=14$, 41%),^{40,42,48,50,51,55,58,60,63,66,67,69,71} approximately half did not directly define the concept. Instead, many discussed its implementation broadly without directly describing the concept ($n=8$, 57%).^{42,50,51,58,60,66,69,71} Nonetheless, six (43%) of the studies provided a clear definition and explanation of their respective framework.^{40,48,55,63,67,70}

A large proportion of the studies' curricula described components of experiential learning without explicitly labelling them as being experiential ($n=15$, 44%).^{38,41,43-46,49,52,53,56,59,61,62,64,68}

The curriculum of five studies (15%) did not fit our criteria for experiential learning or other educational approaches.^{39,47,54,57,65}

Enriched Core Competencies

As shown in Figures 4 and 5, studies frequently incorporated elements aligning with the recommended enriched core competencies outlined by the CIHR.² A large proportion of these studies incorporated elements related to Research and Analytic Skills ($n=22$, 65%).^{38,40,42-44,46-48,50,53,55,56,58-61,62,64,67,69,71} Even more studies incorporated elements related to Professional Skills ($n=27$, 79%).^{44,47,50,53,56,58-60,64,67,70} Finally, 12 (35%) studies emphasized Change Management and Implementation,^{40,42,47,50,53,55,56,58-60,64,70} and six (18%) studies described Networking.^{40,44,53,55,59,64}

Summary of Findings

This scoping review examined the application of Kirkpatrick model in post-secondary health sciences education across each of the evaluation levels (ie, *reaction*, *learning*, *behaviors*, and

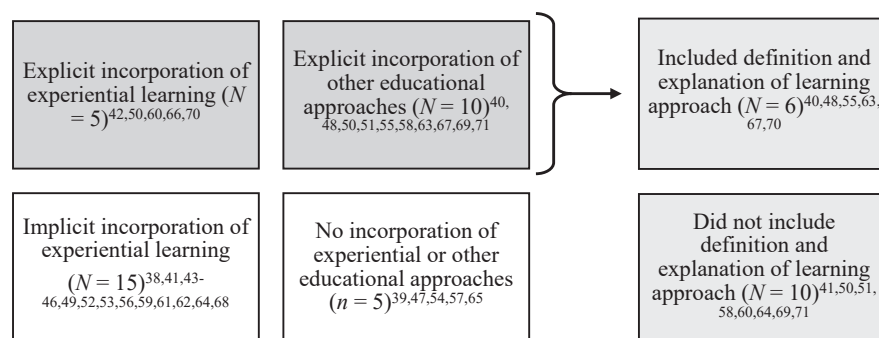


Figure 3. Number of studies incorporating and/or defining experiential or other educational approaches. *Note.* One Study Employed Experiential and Other Educational Approaches.⁵⁰

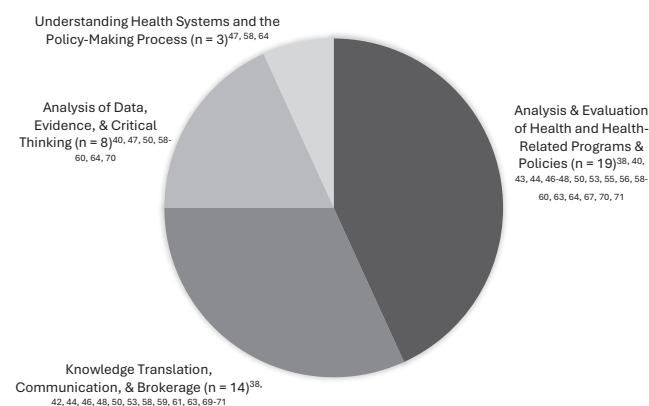


Figure 4. Number of Studies Incorporating Enriched Core Competencies Involving Research and Analytic Skills.

results). Our first aim was to determine the characteristics of studies using the Kirkpatrick model to evaluate post-secondary health sciences programs. Most studies were conducted in the United States. The focus was primarily on undergraduate and professional students in fields like medicine and nursing, with less emphasis on graduate-level master's and doctoral education. The educational content varied, with a focus on medical education, including specialized curriculum such as clinical skills and interprofessional programs.

Our second aim was to understand the way in which the Kirkpatrick model has been applied to evaluate post-secondary health sciences programs. Our findings reveal a predominant reliance on self-report surveys. Most studies used cross-sectional or cohort designs, with some using mixed methods. Although not always explicitly defined, experiential learning was a significant focus of most studies. Some studies also incorporated additional evaluation models alongside the Kirkpatrick model (ie, ADDIE,⁷² CIPP,⁷³ Freeth/Kirkpatrick,⁷⁴ SWOT⁷⁵).

Our third aim was to identify outcomes of post-secondary health sciences programs associated with each of the Kirkpatrick model levels. The main findings of our scoping review suggest positive perceptions and outcomes associated with post-secondary health sciences programs under evaluation. Participants generally held favorable views of the curricula, citing their value, applicability, and engagement. Suggestions for improvement at the *reaction* level tended to focus on course structure and format. Learning outcomes were generally positive and perceived to be enhanced, although variations were observed across different subtopics within the curricula. Participants also reported positive *behavioral* changes, including increased confidence and skill competency. However, challenges were encountered in identifying and measuring the resulting *behavior* changes and long-term *results* of the programs.

Discussion

A notable pattern across the included studies was the reduced emphasis on Levels 3 (behavior) and 4 (results) relative to Levels 1 and 2.⁷⁶ Although this has traditionally been attributed to practical barriers such as resource demands or methodological complexity,^{76,77} our findings suggest that

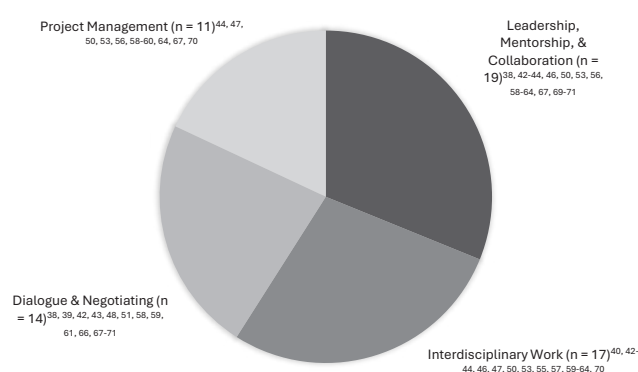


Figure 5. Number of Studies Incorporating Enriched Core Competencies Involving Professional Skills.

conceptual and contextual mechanisms may also contribute to this pattern. Within the New World Kirkpatrick Model, meaningful movement from learning (Level 2) to behavior change (Level 3) depends on the presence of “required drivers,” including organizational reinforcement, supervisory support, and opportunities to apply newly acquired skills.⁷⁸ Similarly, the ability to demonstrate results (Level 4) rests on the identification of “leading indicators” that serve as early signals of broader organizational impact. When these drivers and indicators are not intentionally built into training programs – or are not feasible to measure within academic timelines – evaluators may be limited in their ability to assess higher-level outcomes.⁷⁶

This mechanism-oriented interpretation aligns with broader evaluation and implementation science literature, which emphasizes that changes in professional behavior are shaped by the interaction between individual learning, contextual enablers, and organizational readiness.^{79,80} Consistent with this perspective, many programs in our review reported strong learning outcomes yet lacked corresponding structures in the workplace that would facilitate, reinforce, or monitor behavior change. Similarly, attempts to measure organizational results were constrained by challenges in accessing system-level data, attributing outcomes to a single educational intervention, and following learners long enough to observe sustained change. These contextual influences illustrate why higher-level outcomes are less frequently captured, even when programs appear to be well designed and positively received.⁸⁰

Taken together, these findings highlight the importance of designing evaluations that make mechanisms visible. Incorporating indicators such as learner confidence and commitment at Level 2, documenting environmental and supervisory supports at Level 3, and identifying feasible short-term organizational markers at Level 4 may enhance the quality and completeness of Kirkpatrick-aligned evaluations.⁸¹ In turn, these refinements can bridge the gap between what learners *know* and what they *do* in practice, supporting a more layered understanding of educational impact across levels.^{79,80}

Strengths

Our scoping review has several strengths. First, our use of the scoping review methodology allowed us to capture

a wide range of studies relevant to evaluation of post-secondary health sciences programs, ensuring the inclusivity and breadth of our review, while also applying rigorous screening and selection procedures and meticulous data extraction approaches. Second, careful synthesis and analysis of the 34 eligible studies using the scoping review methodology allowed us to identify key themes, patterns, and gaps in broader literature. Third, the use of a scoping review methodology allowed for the inclusion of diverse study designs, methodologies, and educational contexts. Our review findings also highlighted several strengths of using the Kirkpatrick model to evaluate post-secondary health sciences programs. First, the Kirkpatrick model was used to evaluate a wide range of health sciences programs, demonstrating its adaptability across disciplines. Second, a considerable number of studies adopted a cross-sectional or cohort design. These methodologies are well-suited for assessing prevalence and relationships, either at a single point or over an extended timeframe, enabling an exploratory examination of training outcomes. A portion of the studies used a mixed-methods design, which, despite its increased complexity and demands in implementation, offers a comprehensive analysis by integrating quantitative and qualitative data. The review includes one study each of a retrospective mono-centered design, a case series, and a controlled trial. While the controlled trial is represented singularly, it brings a valuable experimental dimension to the predominantly observational nature of the studies included. Third, the explicit or implicit incorporation of experiential learning elements across the included studies implies a dynamic educational landscape, where educators are seen drawing from multiple strategies to enrich the learning environment. These findings suggest that educational programs in the health sciences are actively incorporating a broad spectrum of skills and competencies. This integration should equip learners to make meaningful contributions to learning healthcare systems. The studies placed significant emphasis on research and analytic skills, such as analyzing and evaluating healthcare programs and effectively communicating knowledge. Finally, the focus on interdisciplinary collaboration, effective communication, and adaptability highlights a forward-looking approach to professional development within the healthcare sector.

Limitations

This scoping review was not without limitations. First, the scope and generalizability of the findings are limited, primarily stemming from the narrow focus on specific post-secondary health sciences programs. This narrow focus, particularly on nurses and doctors in terms of educational content, may inadvertently overlook the diverse educational needs of other health sciences programs, potentially constraining the generalizability of our conclusions across the broader spectrum of health sciences programs. Second, restricting the eligibility of our scoping review to post-secondary health sciences programs, rather than opening up our eligibility criteria to healthcare professionals already in the workforce, may limit the applicability of our findings exclusively to this

specific cohort. Third, our focus on the Kirkpatrick model to the exclusion of other evaluation models (eg, ADDIE, CIPP) did not allow us to compare the suitability of these different models in evaluating post-secondary health sciences programs.

Furthermore, with regards to our scoping review findings, several limitations were highlighted. First, our evaluation of educational outcomes reveals notable differences in assessing levels three (ie, *behavior*, 65% of eligible studies) and four (ie, *results*, 41% of eligible studies) compared to levels one (ie, *reaction*, 88% of eligible studies) and two (ie, *learning*, 85% of eligible studies), indicating a potential oversight in assessing the long-term impacts and practical applications of educational initiatives, which are critical for understanding the full spectrum of experiential learning's benefits. Measuring *behaviors* or *results* may prove challenging due to several interrelated factors. Assessing the long-term effects of educational programs on individuals and organizations requires longitudinal studies, which are resource-intensive and time-consuming to conduct. Determining causality between educational programs and broader results, such as individual and organizational improvements, often involves accounting for multiple confounding variables. Thus, attributing observed results solely to educational programs amidst other concurrent initiatives or external factors adds complexity to the evaluation process. Second, our analysis of the identified studies revealed an overreliance on self-report surveys as the primary measure for evaluating each level. This indicates a heavy reliance on the perspectives of learners and educators to assess impact. While these perspectives are valuable, they inherently possess subjectivity and may introduce bias, potentially distorting the true impact of educational interventions.⁸² Third, our scoping review highlights the lack of studies examining program outcomes for post-doctoral fellows and very few studies examining outcomes for doctoral students in the health sciences, and this limitation is critical as findings derived from undergraduate programs may not generalize to doctoral and postdoctoral health sciences programs. Fourth, while the inclusion of interprofessional programs reflects the collaborative nature of healthcare, our findings suggest underrepresentation of certain healthcare professions within these curricula (eg, physical therapists, psychologists, social workers, dietitians). Fifth, a preponderance of studies originating from the United States means our review findings more readily generalize to this cultural context. Finally, theoretical limitations emerged from the absence of a standardized definition of experiential learning. The lack of a unified definition of experiential learning poses challenges in assessing its results and mechanisms within health sciences education.

Because this review focused exclusively on studies using the Kirkpatrick model, our recommendations are intentionally situated within that framework. This approach reflects the model's continued prominence in health sciences education and the need to strengthen how it is currently applied, rather than to compare it with alternative evaluation models.⁸³ At the same time, the Kirkpatrick model has recognized limitations,

including its assumed linearity and limited attention to contextual factors. For this reason, our recommendations emphasize ways to enhance Kirkpatrick-aligned evaluations by incorporating concepts from behavior change and implementation science, such as required drivers, leading indicators, and contextual enablers.⁸⁴ These additions offer practical strategies to improve evaluation practice while preserving the scope and intent of this review. Future research would benefit from comparative work examining how the Kirkpatrick model performs relative to other evaluation frameworks in capturing mechanisms or change and longer-term outcomes.^{83,85}

Future Directions

Several future directions for our scoping review findings and scoping reviews more broadly merit consideration. First, future research needs to adopt longitudinal designs and include comparison groups to enhance the methodological rigor and validity of findings. These approaches could help determine causal relationships and observe educational outcomes for different groups over time for post-secondary professionals. Second, additional evaluation methods, such as interviews with participants or practical and written assessments, are needed to diversify the range of evaluation methods used. These supplementary measures enhance the depth and richness of the evaluation process, capturing not only participants' perceptions but also their experiences and practical understanding of the educational content. Third, given the context-specific nature of our findings, it is essential for subsequent research to diversify the scope of programs evaluated, including various countries, cultural contexts, disciplines, educational levels, educational systems, and evaluation models. Fourth, we need clear definitions to facilitate a shared understanding of experiential learning among educators, researchers, and partners, promoting consistent and effective implementation of evaluation strategies. It also ensures more effective collaboration and knowledge exchange between partners. Fifth, given the collaborative nature of healthcare service delivery, future studies should explore an interprofessional approach to both education and evaluation, fostering interdisciplinary collaboration and enhancing the effectiveness of educational initiatives in post-secondary health sciences programs. Finally, theoretical advancements are also warranted, particularly regarding the integration of behavior change⁸⁶ and implementation science⁸⁷ literature into the conceptualization of the Kirkpatrick model. This integration would enable a more refined understanding of impacts across different evaluation levels.

This scoping review was not without limitations. First, the scope and generalizability of the findings are limited, primarily stemming from the narrow focus on specific post-secondary health sciences programs. This narrow focus, particularly on nurses and doctors in terms of educational content, may inadvertently overlook the diverse educational needs of other health sciences programs, potentially constraining the generalizability of our conclusions across the broader spectrum of health sciences programs. Second, restricting

the eligibility of our scoping review to post-secondary health sciences programs, rather than broadening our eligibility criteria to healthcare professionals already in the workforce, may limit the applicability of our findings exclusively to this specific cohort. Third, our focus on the Kirkpatrick model to the exclusion of other evaluation models (eg, ADDIE, CIPP) did not allow us to compare the suitability of these different models in evaluating post-secondary health sciences programs.

Conclusion

Collectively, our findings contribute to a more comprehensive understanding of applications of the Kirkpatrick model in accredited academic settings among interdisciplinary domains that support the healthcare system. Our findings revealed a range of applications of the Kirkpatrick model across various educational contexts with an emphasis on evaluating *reaction* and *learning* outcomes. While these findings provide insight into the immediate impacts of educational programs, a notable gap remains in evaluating *behaviors* and long-term *results*, which are crucial for program evaluation and subsequent improvements. Moreover, our results highlighted limited representation of certain healthcare professions and the absence of explicit definitions for experiential learning, suggesting areas for future research and practice development. Moving forward, educators and researchers are encouraged to adopt more robust evaluation methodologies in applying the Kirkpatrick model, including longitudinal designs, to better capture the full spectrum of outcomes associated with post-secondary health sciences education programs. Addressing these limitations can lead to improved quality and effectiveness in graduate-level health sciences education, ultimately advancing healthcare delivery and patient outcomes.

Acknowledgments

We would like to acknowledge Cara Bradley, Dr. John Archer Library & Archives, University of Regina, for her work on the initial and full search strategy for this manuscript. We also thank Briana De Roo for her assistance in screening titles and abstracts, reviewing full texts, and extracting data.

Disclosure of artificial intelligence (AI) use

Not applicable.

Ethical issues

Not applicable.

Conflicts of interest

Authors declare that they have no conflicts of interest.

Authors' contributions

Conceptualization: Natasha L. Gallant, Elizabeth Oddone Paolucci, Ray Jewett, Chantelle Recsky, Katherine Ford, and Deborah A. Marshall.

Data curation: Natasha L. Gallant, Chelsea L. Russill, and Ray Jewett.

Formal analysis: Natasha L. Gallant, Elizabeth Oddone Paolucci, Chelsea L. Russill, Ray Jewett, and Deborah A. Marshall.

Funding acquisition: Natasha L. Gallant, Elizabeth Oddone Paolucci, and Deborah A. Marshall.

Investigation: Natasha L. Gallant, Elizabeth Oddone Paolucci, Chelsea L. Russill, Ray Jewett, and Deborah A. Marshall.

Methodology: Natasha L. Gallant, Elizabeth Oddone Paolucci, Chelsea L.

Russill, Ray Jewett, and Deborah A. Marshall.

Project administration: Natasha L. Gallant, Elizabeth Oddone Paolucci, and Deborah A. Marshall.

Resources: Natasha L. Gallant, Elizabeth Oddone Paolucci, and Deborah A. Marshall.

Software: Natasha L. Gallant, Chelsea L. Russill, and Ray Jewett.

Supervision: Natasha L. Gallant, Elizabeth Oddone Paolucci, and Deborah A. Marshall.

Validation: Natasha L. Gallant and Ray Jewett.

Visualization: Natasha L. Gallant.

Writing—original draft: Natasha L. Gallant, Elizabeth Oddone Paolucci, Chelsea L. Russill, Ray Jewett, Chantelle Recsky, Katherine Ford, Dina Idriss-Wheeler, Hina Ansari, Zeenat Ladak, and Deborah A. Marshall.

Writing—review & editing: Natasha L. Gallant, Elizabeth Oddone Paolucci, Ray Jewett, Chantelle Recsky, Katherine Ford, Victorine Tseung, Hina Ansari, Zeenat Ladak, Aida Fernandes, and Deborah A. Marshall.

Funding statement

the research, authorship, and/or publication of this article. This study was supported by the Institute for Health Services and Policy Research (IHSPR) of the Canadian Institutes for Health Research (CIHR) and supplemental extensions [Funding Reference Numbers: Grant #177252 and Grant #184412]. Additional cash and in-kind support was provided by several organizations: From June 2021 to May 2022, the University of Calgary Cumming School of Medicine, the University of Calgary Vice President of Research Office, and the Alberta Health Services Strategic Clinical Network; from June 2022 to May 2023, the BC Academic Health Science Network and the Michael Smith Foundation for Health Research; from June 2021 to May 2023, the Alberta Children's Hospital Research Institute; from June 2021 to May 2022, Alberta Innovates; from June 2021 to May 2023, Alberta Strategy for Patient Oriented Research SUPPORT Unit (AbSPORU); and, from June 2023 to May 2024, the O'Brien Institute for Public Health.

Authors' affiliations

¹Department of Psychology, Faculty of Arts, University of Regina, Regina, SK, Canada. ²Departments of Community Health Sciences and Surgery, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ³Department of Geography and Planning, University of Toronto, Toronto, ON, Canada. ⁴Ontario Health, Toronto, ON, Canada. ⁵School of Nursing, Faculty of Applied Sciences, University of British Columbia, Vancouver, BC, Canada. ⁶Department of Kinesiology and Health Sciences, Faculty of Health, University of Waterloo, Waterloo, ON, Canada. ⁷Interdisciplinary School of Health Sciences, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada. ⁸School of Rehabilitation Science, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada. ⁹MAP Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada. ¹⁰Department of Applied Psychology & Human Development, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada. ¹¹IMAGINE Network SPOR, McMaster University, Hamilton, ON, Canada. ¹²Departments of Community Health Sciences and Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

Supplementary files

Supplementary file 1. Search Strategies.

References

- Kolb AY, Kolb DA. Experiential learning theory as a guide for experiential educators in higher education. *Experiential Learning & Teaching in Higher Education*. 2022;1(1):38. doi:10.46787/elthe.v1i1.3362
- Government of Canada Canadian Institute of Health Research. Canadian health services and policy research alliance: training modernization in health services and policy research. <https://cihr-irsc.gc.ca/e/49883.html>. Modified April 5, 2016. Accessed June 3, 2024.
- Caley L, Williams SJ, Spernaes I, Thomas D, Behrens D, Willson A. Frameworks for evaluating education programmes and work related learning: a scoping review. *J Workplace Learn*. 2021;33(6):486-501. doi:10.1108/JWL-09-2020-0157
- Zackoff MW, Real FJ, Abramson EL, Li STT, Klein MD, Gusic ME. Enhancing educational scholarship through conceptual frameworks: a challenge and roadmap for medical educators. *Acad Pediatr*. 2019;19(2):135-141. doi:10.1016/j.acap.2018.08.003
- Pereira D, Flores MA, Niklasson L. Assessment revisited: a review of research in assessment and evaluation in higher education. *Assess Eval High Educ*. 2016;41(7):1008-1032. doi:10.1080/02602938.2015.1055233
- Fitzgerald JT, Burkhardt JC, Kasten SJ, et al. Assessment challenges in competency-based education: a case study in health professions education. *Med Teach*. 2016;38(5):482-490. doi:10.3109/0142159X.2015.1047754
- Bommu R. Advancements in healthcare information technology: a comprehensive review. *Innovative Computer Sciences Journal*. 2022;8(1):1-7.
- Zhang X, Lin D, Pforsich H, Lin VW. Physician workforce in the United States of America: forecasting nationwide shortages. *Hum Resour Health*. 2020;18(1):8. doi:10.1186/s12960-020-0448-3
- van Diggele C, Roberts C, Burgess A, Mellis C. Interprofessional education: tips for design and implementation. *BMC Med Educ*. 2020;20(S2):455. doi:10.1186/s12909-020-02286-z
- Cruz-Gomes S, Amorim-Lopes M, Almada-Lobo B. The demand for healthcare services and resources: patterns, trends and challenges in healthcare delivery. In: Alves MJ, Almeida JP, Oliveira JF, Pinto AA, eds. *Operational Research*. Cham: Springer International Publishing; 2019:91-106. doi:10.1007/978-3-030-10731-4_7
- Thibault GE. The future of health professions education: emerging trends in the United States. *FASEB Bioadv*. 2020;2(12):685-694. doi:10.1096/fba.2020-00061
- Aldriwesh MG, Alyousif SM, Alharbi NS. Undergraduate-level teaching and learning approaches for interprofessional education in the health professions: a systematic review. *BMC Med Educ*. 2022;22(1):13. doi:10.1186/s12909-021-03073-0
- Bridges D, Davidson RA, Soule Odegard P, Maki IV, Tomkowiak J. Interprofessional collaboration: three best practice models of interprofessional education. *Med Educ Online*. 2011;16(1):6035. doi:10.3402/meo.v16i0.6035
- Kaufman DM. ABC of learning and teaching in medicine: applying educational theory in practice. *BMJ*. 2003;326(7382):213-216. doi:10.1136/bmj.326.7382.213
- Allen LM, Hay M, Palermo C. Evaluation in health professions education—is measuring outcomes enough? *Med Educ*. 2022;56(1):127-136. doi:10.1111/medu.14654
- Brandt B, Lutfiyya MN, King JA, Chioreso C. A scoping review of interprofessional collaborative practice and education using the lens of the Triple Aim. *J Interprof Care*. 2014;28(5):393-399. doi:10.3109/13561820.2014.906391
- Sfantou D, Laliotis A, Patelarou A, Sifaki- Pistolla D, Matalliotakis M, Patelarou E. Importance of leadership style towards quality of care measures in healthcare settings: a systematic review. *Healthcare (Basel)*. 2017;5(4):73. doi:10.3390/healthcare5040073
- Gervais J. The operational definition of competency-based education. *Journal of Competency-Based Education*. 2016;1(2):98-106. doi:10.1002/cbe2.1011
- Thomas PA, Kern DE, Hughes MT, Chen BY, eds. *Curriculum Development for Medical Education: A Six-Step Approach*. 4th ed. Baltimore, MA: Johns Hopkins University Press; 2022.
- Kirkpatrick JD, Kirkpatrick WK. *Kirkpatrick's Four Levels of Training Evaluation*. Alexandria, VA: ATD Press; 2016.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9):63-67. doi:10.1097/00001888-199009000-00045
- Harden RM. Developments in outcome-based education. *Med Teach*. 2002;24(2):117-120. doi:10.1080/01421590220120669
- Kirkpatrick D. Techniques for evaluating training programs. *J ASTD*. 1959;13:3-9.
- Kirkpatrick D. Great ideas revisited: revisiting Kirkpatrick's four-level model. *Training and Development*. 1996;50(1):54-7.
- Kirkpatrick DL, Kirkpatrick JD. *Transferring Learning to Behavior: Using the Four Levels to Improve Performance*. San Francisco, CA: Berrett-Koehler Publishers; 2005
- Reio TG, Rocco TS, Smith DF, Chang E. A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development*. 2017;29(2):35-53. doi:10.1002/nha3.20178
- Cahapay M. Kirkpatrick model: its limitations as used in higher education evaluation. *International Journal of Assessment Tools in Education*. 2021;8(1):135-144. doi:10.21449/ijate.856143
- Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Eval Program Plann*. 2004;27(3):341-347. doi:10.1016/j.evalprogplan.2004.04.011

29. Alsalamah A, Callinan C. The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959–2020). *Ind Commer Train*. 2022;54(1):36-63. doi:10.1108/ICT-12-2020-0115
30. Lee H, Song Y. Kirkpatrick model evaluation of accelerated second-degree nursing programs: a scoping review. *J Nurs Educ*. 2021;60(5):265-71. doi:10.3928/01484834-20210420-05
31. Nawaz F, Ahmed W, Khushnood M. Kirkpatrick model and training effectiveness: a meta-analysis 1982 to 2021. *JBER*. 2022;14(2):35-56. doi:10.22547/BER/14.2.2
32. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Implement*. 2021;19(1):3-10. doi:10.1097/XEB.000000000000027
33. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18(1):143. doi:10.1186/s12874-018-0611-x
34. Morrison A, Polisena J, Husereau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care*. 2012;28(2):138-144. doi:10.1017/S0266462312000086
35. UNESCO Institute for Statistics. International Standard Classification of Education ISCED 2011. <https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>. Published 2012. Accessed July 2023.
36. Covidence systematic review software. Veritas Health Innovation, Melbourne, Australia. <https://www.covidence.org/>.
37. Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall; 1984.
38. Blackmer AB, Thompson M, Brunner JM, Patel T, Saseen JJ. Implementation and assessment of a novel APPE intercession course to assess near-terminal student competence. *Am J Pharm Educ*. 2018; 82(10):6460. doi:10.5688/ajpe6460
39. Çalıřkan SA, Durmaz S, Akçiçek SF, et al. Medical students' opinions on career planning course: evaluations of the relationship between course and faculty attributes and student characteristics. *Turkish Journal of Biochemistry*. 2022;47(4):539-544. doi:10.1515/tjb-2021-0299
40. Carlford S, Roback K, Nilsen P. Five years' experience of an annual course on implementation science: an evaluation among course participants. *Implement Sci*. 2017;12(1):101. doi:10.1186/s13012-017-0618-4
41. Cuteanu A, Hellich A, Cardinal AL, et al. Evaluation of a microsurgery training curriculum. *J Reconstr Microsurg*. 2023;39(8):589-600. doi:10.1055/a-2003-7689
42. Hawley SR. Using adaptive leadership principles to support Public Health 3.0 in multidisciplinary undergraduate education. *Leadersh Health Serv*. 2021;34(3):248-262. doi:10.1108/LHS-07-2020-0051
43. Lasmarias C, Guanter L, Turrillas P, Peñafiel J, Gómez-Batiste X. Evaluating the impact of a multidisciplinary master programme in palliative care on professional learning outcomes: a cross-sectional study after 20 years of experience. *BMJ Support Palliat Care*. 2020;0:1-7. doi:10.1136/bmjspcare-2020-002528
44. McClintock AH, Starks H, Williams M. Women's health for a primary care workforce. *Clin Teach*. 2022;19(3):251-256. doi:10.1111/tct.13483
45. Naidoo N, Azar AJ, Khamis AH, et al. Design, implementation, and evaluation of a distance learning framework to adapt to the changing landscape of anatomy instruction in medical education during COVID-19 pandemic: a proof-of-concept study. *Front Public Health*. 2021;9:726814. doi:10.3389/fpubh.2021.726814
46. Pinilla S, Cantisani A, Klöppel S, Strik W, Christoph Nissen, Huwendiek S. Introducing a psychiatry clerkship curriculum based on entrustable professional activities: an explorative pilot study. *Acad Psychiatry*. 2021; 45(3):354-359. doi:10.1007/s40596-021-01417-y
47. Ramaswamy R, Mosnier J, Reed K, Powell BJ, Schenck AP. Building capacity for Public Health 3.0: introducing implementation science into an MPH curriculum. *Implement Sci*. 2019;14(1):18. doi:10.1186/s13012-019-0866-6
48. Sudario G, Wiechmann W, Youm J, Le-Bucklin KV. An effective COVID-19 medical student elective. *West J Emerg Med*. 2022;23(1):40-46. doi:10.5811/westjem.2021.11.53656
49. Swanson JR, Shook DC, Vacanti JC, Molloy LM, Fields KG, Palmer LJ. Implementation of a self-guided focused cardiac ultrasound curriculum for anesthesiology residents. *J Educ Perioper Med*. 2020;22(2). doi:10.46374/volxxii-issue2-swanson
50. van Dam PJ, Griffin P, Reeves NS, et al. Learning in practice: collaboration is the way to improve health system outcomes. *Healthcare*. 2019;7(3):90. doi:10.3390/healthcare7030090
51. Velusami D, Dongre AR, Kagne RN. Evaluation of one-month foundation course for the first year undergraduate students at a medical college in Puducherry, India. *J Adv Med Educ Prof*. 2020;8(4): 165-171. doi:10.30476/jamp.2020.86857.1272
52. Azar AJ, Khamis AH, Naidoo N, et al. Design, implementation and evaluation of a distance learning framework to expedite medical education during COVID-19 pandemic: a proof-of-concept study. *J Med Educ Curric Dev*. 2021;8:1-16. doi:10.1177/23821205211000349
53. Becker H, Timmerman GM, Delville C, Seo E. A multifaceted model to evaluate interprofessional education in clinical nurse specialist programs. *Clin Nurse Spec*. 2017;31(5):243-251. doi:10.1097/NUR.0000000000000320
54. Cangarella J, Elias K, Kalet A, Cohen E, Abramson S, Gillespie C. A preliminary evaluation of students' learning and performance outcomes in an accelerated 3-year MD pathway program. *J Grad Med Educ*. 2022; 14(1):99-107. doi:10.4300/JGME-D-21-00284.1
55. Cioffi I, Dale CM, Murphy L, Langlois S, Musa R, Stevens B. Ten years of interfaculty pain curriculum at the University of Toronto: impact on student learning. *Pain Rep*. 2021;6(4):e974. doi:10.1097/PR9.0000000000000974
56. Craig P, Hall S, Phillips C. Using the Freeth/Kirkpatrick model to evaluate interprofessional learning outcomes in a rural setting. *Focus Health Prof Educ*. 2016;17(1):84-99. doi:10.1157/fohpe.v17i1.124
57. Fiza B, Jain S. Evaluation of foundation course for medical students: student and faculty perspectives. *NeuroQuantology*. 2022;20(20):1600-1606. doi:10.48047/NQ.2022.20.20.NQ109162
58. Kennedy AG, Burnett M, Muthukrishnan P, Sobel H, van Eeghen C, Repp AB. "I think I was losing the forest for the trees": evaluation of an internal medicine residency quality improvement curriculum. *Med Sci Educ*. 2020;30(1):197-202. doi:10.1007/s40670-019-00854-7
59. Knox KE, Lehmann W, Vogelgesang J, Simpson D. Community health, advocacy, and managing populations (CHAMP) longitudinal residency education and evaluation. *J Patient Cent Res Rev*. 2018;5(1):45-54. doi:10.17294/2330-0698.1580
60. Mathura P, Lee DH, Thompson A, McMurtry N, Kassam N. Providing quality improvement training in an advanced pharmacy practice experience elective. *Curr Pharm Teach Learn*. 2021;13(4):397-402. doi:10.1016/j.cptl.2020.11.013
61. Pfeifle A, Glassburn S, Frank K, et al. Program evaluation of a new interprofessional geriatrics curriculum for advanced practice nursing and Master of Social Work learners. *J Interprof Care*. 2023;37(5):743-753. doi:10.1080/13561820.2022.2158182
62. Schussel KE, Forbes S, Taylor AM, Cooley JH. Implementation of an interprofessional medication therapy management experience. *Am J Pharm Educ*. 2019;83(3):6584. doi:10.5688/ajpe6584
63. Sieplinga K, Disbrow E, Triemstra J, Van De Ridder M. Off to a jump start: using immersive activities to integrate continuity clinic and advocacy. *J Med Educ Curric Dev*. 2021;8:238212052110596. doi:10.1177/23821205211059652
64. Stingl CS, Alexander KJ, Dittman JM, et al. Development and implementation of a longitudinal global acute care and systems strengthening program. *Ann Glob Health*. 87(1):125. doi:10.5334/aogh.3385
65. Wang ML, Chang AC. Student learning outcome assessment for an information organization curriculum based on the Kirkpatrick framework. *Qualitative and Quantitative Methods in Libraries*. 2017;6(4):527-544.
66. Dante A, La Cerra C, Caponnetto V, et al. Dose-response relationship between high-fidelity simulation and intensive care nursing students' learning outcomes: an Italian multimethod study. *Int J Environ Res Public Health*. 2022;19(2):617. doi:10.3390/ijerph19020617
67. Roca J, Gros Navés S, Canet-Velez O, et al. Service learning in the nursing bachelor thesis: a mixed-methods study. *Int J Environ Res Public Health*. 2022;19(19):12387. doi:10.3390/ijerph191912387
68. Venugopal V, Dongre AR. Effect of interactive lectures and formative assessment on learning of epidemiology by medical undergraduates—a mixed-methods evaluation. *Indian J Community Med*. 2020;45(4):526-530. doi:10.4103/ijcm.IJCM_46_20
69. Berry M, Rodrigues V, Evans JL, De Souza MIDC, Reher V. Design of a communication skills course in dentistry: Applying active-learning strategies to improve educational outcomes. *Eur J Dent Educ*. 2022;

- 26(3):577-585. doi:[10.1111/eje.12734](https://doi.org/10.1111/eje.12734)
70. van Tuijl AA, Calsbeek H, Wollersheim HC, Laan RF, Fluit CR, van Gurp PJ. Does a long-term quality and safety curriculum for health care professionals improve clinical practice? An evaluation of quality improvement projects. *J Contin Educ Health Prof.* 2020;40(1):3-10. doi:[10.1097/CEH.0000000000000277](https://doi.org/10.1097/CEH.0000000000000277)
 71. Alsuwaidi L, Kristensen J, HK A, Al Heialy S. Use of simulation in teaching haematological aspects to undergraduate medical students improves student's knowledge related to the taught theoretical underpinnings. *BMC Med Educ.* 2021;21(1):271. doi:[10.1186/s12909-021-02709-5](https://doi.org/10.1186/s12909-021-02709-5)
 72. Peterson C. Bringing ADDIE to life: instructional design at its best. *Journal of Educational Multimedia and Hypermedia.* 2003;12(3):227-241.
 73. Stufflebeam DL, Zhang G. *The CIPP Evaluation Model: How to Evaluate for Improvement and Accountability.* New York, NY: The Guilford Press; 2017.
 74. Freeth D, Hammick M, Koppel I, Reeves S, Barr H. *A critical review of evaluations of interprofessional education. Occasional Paper No. 2.* London, UK: Learning and Teaching Support Network for Health Sciences and Practice; 2002.
 75. Chermack TJ, Kasshanna BK. The use and misuse of SWOT analysis and implications for HRD professionals. *Hum Resour Dev Int.* 2007;10(4):383-399. doi:[10.1080/13678860701718760](https://doi.org/10.1080/13678860701718760)
 76. Kennedy PE, Chyung SY, Winiecki DJ, Brinkerhoff RO. Training professionals' usage and understanding of Kirkpatrick's Level 3 and Level 4 evaluations. *Int J Train Dev.* 2013;18:1. doi:[10.1111/ijtd.12023](https://doi.org/10.1111/ijtd.12023)
 77. Kirkpatrick Partners. The Kirkpatrick Model. Newnan, GA: Kirkpatrick Partners; 2024. <https://www.kirkpatrickpartners.com/the-kirkpatrick-model/>.
 78. Kusmiati M. A comprehensive evaluation in medical curriculum using the Kirkpatrick hierarchical approach: A review and update. *Med Res Arch.* 2025;13(5). doi:[10.18103/mra.v13i5.6557](https://doi.org/10.18103/mra.v13i5.6557)
 79. Caci L, Nyantakyi E, Blum K, et al. Organizational readiness for change: A systematic review of the healthcare literature. *Implement Res Pract.* 2025;6:1-33. doi:[10.1177/26334895251334536](https://doi.org/10.1177/26334895251334536)
 80. Lewis CC, Klasnja P, Lyon AR, et al. The mechanics of implementation strategies and measures: Advancing the study of implementation mechanisms. *Implement Sci Commun.* 2022;3:114. doi:[10.1186/s43058-022-00358-3](https://doi.org/10.1186/s43058-022-00358-3)
 81. Rucks L, Wingate L, Lopez M, Becho LW, FitzGerald M, Dean KL. Leveraging the Kirkpatrick four-level model to evaluate evaluation capacity building work. *New Dir Eval.* 2024;2024(83):81-94. doi:[10.1002/ev.20619](https://doi.org/10.1002/ev.20619)
 82. van de Mortel TF. Faking it: social desirability response bias in self-report research. *Aust J Adv Nurs.* 2008;25(4):40-48.
 83. Allen LM, Hay M, Palermo C. Evaluation in health professions education: Is measuring outcomes enough? *Med Educ.* 2021;56:127-136. doi:[10.1111/medu.14654](https://doi.org/10.1111/medu.14654)
 84. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implement Sci.* 2009;4:50. doi:[10.1186/1748-5908-4-50](https://doi.org/10.1186/1748-5908-4-50)
 85. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE Guide No. 67. *Med Teach.* 2012;34(5):e288-e299. doi:[10.3109/0142159X.2012.668637](https://doi.org/10.3109/0142159X.2012.668637)
 86. Moreau KA. Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? *Med Teach.* 2017;39(9):999-1001. doi:[10.1080/0142159X.2017.1337874](https://doi.org/10.1080/0142159X.2017.1337874)
 87. Moir T. Why is implementation science important for intervention design and evaluation within educational settings? *Front Educ.* 2018;3:61. doi:[10.3389/educ.2018.0006](https://doi.org/10.3389/educ.2018.0006)