



Educational evaluation research in Asia: a scoping review

Jason Loh, Wan Har Chong, Christina Lim-Ratnam, Michelle Tan & Phoebe Ow

To cite this article: Jason Loh, Wan Har Chong, Christina Lim-Ratnam, Michelle Tan & Phoebe Ow (2024) Educational evaluation research in Asia: a scoping review, Asia Pacific Journal of Education, 44:3, 552-568, DOI: [10.1080/02188791.2024.2390663](https://doi.org/10.1080/02188791.2024.2390663)

To link to this article: <https://doi.org/10.1080/02188791.2024.2390663>



Published online: 09 Oct 2024.



Submit your article to this journal [↗](#)



Article views: 215



View related articles [↗](#)



View Crossmark data [↗](#)



Educational evaluation research in Asia: a scoping review

Jason Loh , Wan Har Chong , Christina Lim-Ratnam , Michelle Tan 
and Phoebe Ow 

National Institute of Education, Nanyang Technological University, Singapore

ABSTRACT

This scoping review examines the landscape of educational evaluation research in Asia over the past decade, uncovering the methodologies, themes, and unique regional challenges within this domain. By delving into 62 articles from evaluation-centric journals, the study reveals a predominant focus on learning, teaching, and assessment, highlighting the importance of pedagogical methods and learning outcomes. The review also notes a significant emphasis on institutional-level evaluation, with China leading in contributions. Despite the rich insights offered, this review points to gaps in leadership and evaluation tools, suggesting areas for future exploration. Ultimately, this review aims to enrich global discussions on educational evaluation, emphasizing the need for a comprehensive understanding of Asia's diverse and evolving educational evaluation practices.

ARTICLE HISTORY

Received 15 March 2024
Accepted 26 June 2024

KEYWORDS

Scoping review; research on evaluation; research on educational evaluation; Asia; Singapore

Introduction

Context and significance

Asia, which hosts about 60% of the global population, offers a vibrant educational landscape, richly interlaced with diverse cultures, historical backgrounds, and developmental trajectories (United Nations Population Fund, 2023). In recent years, the region has witnessed a multitude of educational reforms, reflecting its diverse socio-economic and cultural landscapes. For instance, Japan's recent shift towards "independent and interactive in-depth learning" and "active learning" articulated in the New National Curriculum Standards reform, seeks to connect what is learnt in schools with life and society outside the classroom (Yamanaka & Suzuki, 2020, p. 93); South Korea's major national reform has focused on integrating technology and innovation in classrooms, building the capacities of in-service and pre-service teachers, merging early childhood education and childcare, and putting in place a devolution of government controls to the regions and local governments as a means to reinvigorate the Korean educational landscape (Ministry of Education, Korea, 2023). Similarly, India's National Education Policy 2020 marks a significant overhaul, aiming to increase accessibility and inclusivity in education (Ministry of Education, India, n.d.), resonating with movements across the region for more equitable education systems.

The role of educational evaluation research

Over these past few decades, as economies burgeoned and societies transformed, many Asian countries embarked on ambitious educational reforms aimed at aligning curricula with

CONTACT Jason Loh  jason.loh@nie.edu.sg  National Institute of Education, National Institute of Education, 1 Nanyang Walk, Singapore 637616, Singapore

This article was originally published with errors, which have now been corrected in the online version. Please see Correction (<http://dx.doi.org/10.1080/02188791.2024.2419251>)

© 2024 National Institute of Education, Singapore

global standards, fostering innovation, and promoting inclusive education (Goodwin & Low, 2020; Kim, 2022; Lin et al., 2014; Yakavets et al., 2023). To ascertain the outcomes and impacts of these reforms, the role of educational evaluation research has grown more salient. Kennedy and Lee (2008) highlighted the importance of educational evaluation in the pursuit of educational improvement in Asian societies. Educational evaluation has generally been understood as evaluating students' performance (particularly in high-stakes assessments), the quality of educational organizations (specifically accountability and quality assurance), the effectiveness of their academic programmes (especially curricular reforms), the use of certain pedagogical practices, the relevance of curricular/educational policies, the value of teacher education, among others (Peng & Lee, 2009). As educational evaluation covers a broad spectrum of areas, this paper will tap on Coryn et al. (2016) intentionally broad definition of research on evaluation (RoE). In this paper, educational evaluation research is defined as "[a]ny purposeful, systematic, empirical inquiry intended to test existing knowledge, contribute to existing knowledge, or generate new knowledge related to some aspect of evaluation processes or products, or evaluation theories, methods, or practices" within the field of education (p. 161).

Practical implications and growing interest

The information gleaned from an educational evaluation research endeavour allows the stakeholders to ascertain the "‘merit, value, or worth’, ‘efficacy’, ‘effectiveness or efficiency’, or extent to which the program was ‘meeting its stated goals and objectives’" (Wanzer, 2021, p. 36). Such a value judgement is considered the primary purpose for evaluation by evaluators, policymakers, school leaders and practitioners (Wanzer, 2021). With such knowledge, the quality of education can be further improved, and steps can be taken to ensure that students have better and greater access to high-quality learning opportunities. In recent years, the growing interest in educational evaluation research in Asia (Cheng & King, 2017; Loh et al., 2023; Mazur & Woodland, 2017; Pascua & Chang, 2015) is driven by several factors, including the increasing recognition of the importance of education for economic development and social progress, the expansion of educational systems, and the growing diversity of student populations. Educational evaluation research offers valuable insights into the efficacy of pedagogical interventions, shaping future strategies and policy directions (Phillips, 2018).

Expanding educational evaluation in asia

Addressing the literature gap and objectives

Despite the growing interest in educational evaluation research in Asia, there has been no comprehensive review of the field. Indeed, a comprehensive understanding of the methodologies, themes, and regional nuances of evaluation research in the Asian educational landscape remains fragmented. The western-centric bias in the existing literature inadvertently casts a shadow on the rich, diverse, and potentially contrasting perspectives from Asia. For instance, the unique challenges and innovations within Asian education systems, such as the balancing of traditional cultural practices with modern pedagogical approaches, often remain underrepresented. This scoping review aims to address this lacuna. By examining the educational evaluation research in the past decade, we seek to chart the evolution of educational evaluation research in Asia, highlighting its unique facets and contextual challenges. By doing so, it ultimately aims to enrich the global discourse on educational evaluation. The review will be of interest to a wide range of stakeholders, including researchers, policymakers, and practitioners. It will be particularly useful for those who are interested in learning more about the current state of educational evaluation research in Asia and in identifying areas for future research directions.

Beyond Western-centric Bias in evaluation research

The burgeoning importance of evaluation research in educational domains cannot be understated. It is a tool by which educational stakeholders gauge the success of interventions, identify areas of improvement, reconsider the cost-benefit and cost-effectiveness of these educational initiatives in which to base decisions to inform about further investment in these enterprises. Leaders in the field had previously highlighted the lack of empirical RoE in general (e.g., Alkin, 2003; Christie, 2011; Mark, 2008), and the value of placing “an additional focus of empirical inquiry and of meta-analyses of extant empirical work” (Christie, 2011, p. 13), so as to develop a sizable evidence base with which to offer insight and understanding of the discipline and aid in the refinement of its theories. Recent academic literature has seen a surge in systematic reviews, which aim to decode trends and synthesize findings in the field of evaluation research (see Brandon & Fukunaga, 2014; Christie & Fleischer, 2010; Coryn et al., 2011, 2017; Johnson et al., 2009; Lawaree et al., 2020; Vallin et al., 2015). However, a caveat emerges upon closer scrutiny. The vast majority of these reviews predominantly reflect the western Anglo-American context. This is understandable since the largest and most established evaluation association (i.e., American Evaluation Association) is situated in that context and logically its members would conduct the bulk of their evaluation research/endeavours in that same context (Coryn et al., 2016). This Western-centric bias unintentionally overlooks the wealth of diverse and potentially contrasting perspectives from other regions, especially Asia since 60% of the global population reside in this area. For instance, Henrich et al. (2010), in their “review of the comparative database from across the behavioural sciences” (p. 61), reported “substantial psychological and behavioural variation among human populations” (p. 78). Importantly, their findings highlighted the fact that samples drawn entirely from “Western, Educated, Industrialized, Rich, and Democratic societies” (Henrich et al., 2010, p. 61) may not apply to sample populations from other parts of the world. In fact, they underscored the need to have diverse populations in research programmes, because “[h]ow justified are researchers in assuming a species-level generality for their findings” when the findings are “largely based on findings from this subpopulation” (p. 61)?

This inadvertent oversight creates an epistemological gap. By marginalizing non-western paradigms, the universality, utility and translatability of findings and recommendations from these systematic reviews are, at best, limited. The nuanced intricacies of educational systems, pedagogical traditions, and societal expectations in Asia offer a fertile ground for evaluation research that could significantly contribute to, and transform, global educational perspectives. Given the identified gap in the literature and the paramount importance of understanding educational evaluation research in the Asian context, it is imperative to adopt a rigorous methodological approach. The challenge lies not just in cataloguing studies but in comprehensively understanding their context, scope, and relevance to both Asian and global educational landscapes. Thus, the following “Methodology” section delineates the systematic process we undertook to ensure a comprehensive, unbiased, and relevant scoping review. We detail our search strategy, criteria for inclusion and exclusion, data extraction, and analysis techniques, ensuring transparency and reproducibility in our endeavour to contribute meaningfully to the evaluation field.

Methodology

This study adopts a scoping review approach. A scoping review is “a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesizing existing knowledge” (Colquhoun et al., 2014, pp. 1292–1294). Unlike systematic reviews, which typically focus on specific research questions and often assess the quality of included studies, scoping reviews are particularly useful for examining emerging areas, where the literature might be heterogeneous and the research questions broad (Daudt et al., 2013; Pham et al., 2014).

There are primarily four reasons for embarking on a scoping review: to 1) evaluate the breadth, diversity, and essence of the research evidence; 2) ascertain the merit of conducting a comprehensive systematic review; 3) consolidate and share the results of the research; and 4) pinpoint areas of unexplored terrain within the current body of research (Arksey & O'Malley, 2005; Levac et al., 2010; Pham et al., 2014). These objectives make scoping reviews an ideal method for fields where knowledge is rapidly expanding or highly varied, as they allow for the inclusion of a wide range of study designs and methodologies, thus providing a comprehensive overview of the available evidence.

The choice of a scoping review for this study is justified given the fragmented nature of educational evaluation research in Asia. By mapping out the existing literature, this review aims to provide a broad understanding of the field, identifying key themes, methodologies, and gaps that have yet to be explored (Peterson et al., 2017; Pham et al., 2014). This is particularly important in the context of educational evaluation in Asia, where diverse educational systems, cultural practices, and developmental stages can result in varied research focuses and outcomes.

The steps for this scoping review follow the framework set out by Arksey and O'Malley (2005), ensuring a systematic and rigorous approach:

- Stage 1: Identifying the research question to guide the review and ensure relevance to the field.
- Stage 2: Identifying relevant studies through comprehensive and systematic search strategies across multiple databases and sources.
- Stage 3: Study selection based on predefined inclusion and exclusion criteria to ensure consistency and relevance.
- Stage 4: Charting the data involves extracting key information from each study, such as study design, context, outcomes, and limitations, to provide a detailed overview of the field.
- Stage 5: Collating, summarizing, and reporting the results to synthesize the findings in a way that highlights key insights, patterns, and areas needing further research. (p. 22)

By employing a scoping review, this study not only aims to map the current landscape of educational evaluation research in Asia but also to identify gaps and future research directions that can inform policy, practice, and further scholarly work. This approach is instrumental in providing a comprehensive understanding of the field, which is essential for advancing knowledge and improving educational practices across diverse contexts.

Research questions

The review will focus on the following questions:

- What are the principal topics and research methods employed in educational evaluation research across Asia?
- What key findings emerge from this scoping review of educational evaluation research in Asia, and what are their implications for researchers, policymakers, and practitioners?

Relevant studies: a systematic search

Firstly, our scoping review focused on evaluation-centric publications, drawing from the catalogue of journals highlighted by Coryn et al. (2016, 2017) and Lawaree et al. (2020) as pertinent to our investigation (see Table 1).

These 14 evaluation-oriented journals were selected because they cover a broad spectrum of evaluation-related topics and have published significant numbers of RoE articles. The criteria for selecting these journals included their focus on evaluation as a primary topic, their peer-reviewed status, and their contributions to the field of evaluation through the publication of articles that test

existing knowledge, contribute to existing knowledge, or generate new knowledge related to evaluation processes, products, theories, methods, or practices (Coryn et al., 2016, 2017; Lawaree et al., 2020). Since this scoping review intends to provide insights into the types of RoE investigations conducted, by focusing on these 14 journals, this paper aims to capture a comprehensive and representative sample of the RoE literature, thereby facilitating an analysis of trends, gaps, and directions for future research in the field of evaluation in Asia.

We manually searched for relevant articles, utilizing names of Asian states/cities as keywords. We categorized the search by geographical regions in Asia – East Asia, Southeast Asia, Central Asia, South Asia, and West Asia. Within each geographical region, we listed the countries/states (see Table 2).

Table 1. 14 evaluation-centric publications.

1	<i>African Evaluation Journal (AEJ)</i>	8	<i>Evaluation Journal of Australasia (EJA)</i>
2	<i>American Journal of Evaluation (AJE)</i>	9	<i>Evaluation Review</i>
3	<i>Canadian Journal of Program Evaluation (CJPE)</i>	10	<i>Journal of MultiDisciplinary Evaluation (JMDE)</i>
4	<i>Educational Evaluation and Policy Analysis</i>	11	<i>New Directions for Evaluation (NDE)</i>
5	<i>Evaluation: The International Journal of Theory, Research and Practice (EJTRP)</i>	12	<i>Practical Assessment, Research and Evaluation</i>
6	<i>Evaluation & the Health Professions</i>	13	<i>Research Evaluation</i>
7	<i>Evaluation and Program Planning (EPP)</i>	14	<i>Studies in Educational Evaluation</i>

Table 2. Search terms used within the evaluation-centric journals.

Geographical region	States/Cities
East Asia	China, ¹ Hong Kong, Japan, ¹ Macao, Mongolia, South Korea, Taiwan
Southeast Asia	Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Vietnam
Central Asia	Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan
South Asia	Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan, Sri Lanka
West Asia	Bahrain, Iraq, Jordan, Kuwait, Lebanon, Oman, Palestinian National Authority, Qatar, Saudi Arabia, Syria, United Arab Emirates, Yemen

As mentioned by Arksey and O'Malley (2005), from a pragmatic perspective, initial decisions must be made regarding the review's scope in terms of its time frame and language considerations. The language was not an issue, since the 14 journals only published in the English language. However, for the timeframe, we chose the past decade for the review, i.e., 2012–2022. The year 2011 marked the 25th anniversary of the American Evaluation Association (AEA); the *American Journal of Evaluation* published a series of anniversary issues for that year (volume 32) “to commemorate this significant milestone in the development of evaluation as a professional practice field and academic specialty” (Schwandt, 2011, p. 7). The final issue, 32(4), took stock of the past 25 years and exhorted the evaluation field to look forward for the next 25 years. Smith et al. (2011) anticipate that the field will be more culturally diverse and that evaluations will no longer assume cultural homogeneity but heterogeneity going forward. Smith et al. (2011) similarly anticipate a growth in international evaluation activities and in the profession. These views coupled with Christie's (2011) exhortation to advance the field by conducting more empirical inquiry helped us to decide on 2012 as the starting point of our scoping review. The intent was to ascertain the growth of the field since that key juncture in 2011. In addition, Gargani (2012), a Past President of the AEA, made 10 predictions “about what will happen over the next ten years” in the field of evaluation (para. 2). This timeframe of 2012 to 2022 is therefore used as a means of comparison with Gargani's 10 predictions –

- (1) Most evaluations will be internal.
- (2) Evaluation reports will become obsolete.

- (3) Evaluations will abandon data collection in favour of data mining.
- (4) A national registry of evaluations will be created.
- (5) Evaluations will be conducted in more open ways.
- (6) The RFP will RIP.
- (7) Evaluation theories (plural) will disappear.
- (8) The demand for evaluators will continue to grow.
- (9) The number of training programs in evaluation will increase.
- (10) The term evaluation will go out of favour. (paras. 4-13)

Our scoping study's search across 14 evaluation-centric journals, with the names of Asian states/cities as keywords as well as in the article title, within the timeframe of 2012–2022, yielded a total of 118 references.

Study selection: screening using inclusion/exclusion criteria

We screened the obtained 118 references based on specific inclusion criteria. These criteria were designed to ensure that the data we identified and analysed aligned closely with our research objectives and questions. We adopted the subsequent criteria to maintain uniformity in our decision-making process (see [Table 3](#)).

The methodical screening process started with evaluating article titles and abstracts; if the titles or abstracts did not specify whether the article met the inclusion criteria, the full text was subsequently read through and screened. Initially, two authors individually sorted and reviewed all 118 articles. To address any sorting divergences, periodic consultations/arbitrations were conducted with the first author (i.e., third reviewer). The screening process resulted in the omission of 56 articles. This scoping review draws from a collection of 62 articles focused on research in the field of Education and featured in journals centred on Evaluation (see [Table 4](#)); the complete list of articles analysed in this scoping review can be found in [Appendix A](#).

Table 3. Inclusion/Exclusion criteria.

Articles were included when the articles	Articles were excluded when the articles
(i) report evaluation research in these three different orientations in Evaluation: (a) Theory – A contemplative discourse, viewpoint, or commentary related to a specific domain within the evaluation field; (b) Practice – Empirical investigations evaluating systems, establishments, programs, and the like; (c) Method – Pertaining to advancements in methodology, encompassing evaluation instruments and techniques; (ii) report evaluation research in the field of *Education.	(i) report research which used the Asian state as a subject of study, and which was done from outside of the Asian state using an outsider perspective; (ii) report evaluation research outside the field of *Education.

*The field of Education is defined broadly for the inclusion/exclusion criteria.

Table 4. Number of papers.

	Region	1st level screening: A systematic search	2nd level screening: Inclusion/exclusion screening
		All sectors	Education
1	East Asia	60	43
2	Southeast Asia	20	7
3	Central Asia	2	1
4	South Asia	28	9
5	Western Asia	8	2
Total		118	62

Data extraction

At this juncture, three authors independently reviewed and extracted the relevant data. A structured data extraction spreadsheet was used to extract and record the relevant information and various study characteristics. The spreadsheet provided clear guidance on which variables to extract, ensuring the research team could effectively address the research questions. The first level of information to be extracted was the general characteristics: *Context* (e.g., *country/city of origin*), *Data collection methods* (e.g., *interviews, lesson observation, written reflections*) and *Participants* (e.g., *University lecturers, teachers, students*). The next level of information was more evaluation-oriented: *Evaluation levels*, *Evaluation focus/topics* and *Evaluation perspective*. The Evaluation levels are meant to ascertain if the research studied at the policy/institutional, programmatic or classroom levels. This category of information was based on Mark's (2008) evaluation context and Deng's (2012) curriculum-making domains. The next category is that of the focus/topic of the evaluation research. We agree with Vallin et al. (2015) that there is a need to document "which topics have been most researched and the common designs and data collection methods employed in RoE" so that the findings from this review "can serve as a roadmap for where more research is needed" (p. 14). The final category is that of *Evaluation perspective*, which is essentially the extent stakeholders are involved in the evaluation. This is based on Johnson et al. (2009) category of "Stakeholder involvement", which was added to reflect "the increased research focused on participatory evaluation approaches, stakeholder or decision-maker participation, and/or stakeholder or decision-maker involvement since 1985" (p. 382).

This data extraction/charting process was an iterative one. The spreadsheet was reviewed by the research team and pretested by the first three authors before being implemented. The three reviewers extracted the data characteristics from one paper for initial calibration, with another eight papers for clarification. This step was recommended by Levac et al. (2010), where "two researchers independently extract data from the first five to ten studies using the data-charting form and meet to determine whether their approach to data extraction is consistent with the research question and purpose" (p. 6). This step provided our team an opportunity to discuss the data that was extracted so as "to reach consensus about our data collection process and to include additional categories relevant to our research inquiry" (Daudt et al., 2013, p. 6). The three reviewers addressed any discrepancies through discussion and achieved a consensus on the final interpretation, during the team meeting. Once consensus and mutual understanding were achieved, the three reviewers independently reviewed batches of 18, 18, and 26 articles. During the subsequent team meetings, the reviewers presented their data extraction and charting to facilitate discussion and clarify any potential disagreements. This collaborative approach ensured that all team members were aligned and that any discrepancies were promptly resolved. Finally, the extracted data was systematically collated and summarized to address the research questions and generate meaningful insights.

Findings & discussion

Having delineated the methodology employed in our research, we turn attention to the core findings and discussions that have emerged from our comprehensive scoping review. This next section presents an in-depth analysis of the general characteristics inherent in the field of educational evaluation, followed by an exploration of the specific, evaluation-oriented characteristics that define this domain. Through this examination, we aim to shed light on the multifaceted nature of educational evaluation in Asia, highlighting both its overarching trends and its nuanced, method-specific details. This synthesis not only reflects the breadth and depth of the existing literature but also provides critical insights into the evolving landscape of educational evaluation.

General characteristics

Contextual characteristics

Of the 62 articles, the bulk came from East Asia ($n = 43$; 69.4%): 26 (41.9%) were educational evaluation research on/from China, with eight from Hong Kong, six from Taiwan, two from Japan and one from South Korea. The region with the second highest came from South Asia ($n = 9$; 14.5%): five from Pakistan, two from India, one from Iran and one from Bangladesh. The region with the third highest came from Southeast Asia ($n = 7$; 11.3%): two from Vietnam, and one each from Brunei, Malaysia, the Philippines, Singapore and Thailand. West Asia had two articles: one from Syria and one from the United Arab Emirates, and Central Asia had one article: Kazakhstan.

A summary of the number of articles by region and by country can be found in [Table 5](#):

Table 5. Educational evaluation research papers by region/state/city of origin.

Region	State/city	n		Region	State/city	n	
East Asia	China	26	43	Southeast Asia	Vietnam	2	7
	Hong Kong	8			Brunei	1	
	Taiwan	6			Malaysia	1	
	Japan	2			Philippines	1	
	S. Korea	1			Singapore	1	
South Asia	Pakistan	5	9	West Asia	Thailand	1	2
	India	2			Syria	1	
	Bangladesh	1			United Arab Emirates	1	
	Iran	1			Kazakhstan	1	
				Central Asia			1

China, including Hong Kong, contributed slightly more than half of the educational evaluation research ($n = 34$; 54.8%) in East Asia. The main driver in the Educational Evaluation scene outside of East Asia is Pakistan, with five papers ranging from 2015 to 2022.

This dominance by China in evaluation journals could possibly be due to the fact that monetary and academic rewards (such as promotion and research grant awards) are tied to publications in high impact journals (Hyland, 2023; Shao & Shen, 2012). However, with the recent change in policy to discourage such an “inappropriate” practice (Mallapaty, 2020, p. 18), there is a possibility the numbers might change in the next 10 years. Despite this change in policy, Hyland (2023) predicts that “China’s emergence as a publishing superpower ... is unlikely to stall” (p. 401). It seems likely that educational evaluation research will continue to be driven by East Asia, with China as its leader. Interestingly, this is parallel to Vallin et al. (2015) analytical review of RoE articles in the *American Journal of Evaluation*, from 1998 to 2014. In their review, they found that 90% of the 177 RoE articles were published by North American authors, with the remaining 10% by non-North American authors.

Data collection characteristics

Of the 62 studies, the most common data collection method is the Questionnaire survey ($n = 33$; 54.8%), and the second most common is the Interview ($n = 15$; 24.2%). This is similar with two earlier reviews (Brandon & Fukunaga, 2014; Vallin et al., 2015). The evaluation literature is replete with studies using both surveys and interviews (Mertens & Wilson, 2019; Patton, 2015), since surveys enable evaluators to capture data from a huge population size, and interviews enable evaluators “to obtain descriptive and judgmental information from a wide range of persons who have important perspectives on a program, its setting, or its beneficiaries” (Stufflebeam & Coryn, 2014, p. 546). In this regard, the educational evaluation methodology in Asia is similar with the practice in the wider evaluation world.

Both secondary data analysis and the administration of a test constitute the third most common data collection method ($n = 11$; 17.7%), followed by Document/textual analysis ($n = 7$; 11.3%). The prevalent use of certain research methodologies in Asia (i.e., surveys and tests: $n = 44$; 71.0%)

contributes to the thematic skewness in educational research. Quantitative methods, often used to evaluate learning outcomes, may overshadow qualitative and/or mixed method approaches better suited for exploring complex themes like leadership or policy. This methodological bias might inadvertently limit the scope of evaluation topics and points to the need for a more diversified methodological approach.

Participant characteristics

Of the 62 studies, the bulk of the sample population came from K-12 ($n = 16$; 25.8%) and university students, both undergraduate and postgraduate ($n = 16$; 25.8%). Teachers make up the third largest sample population ($n = 20$; 16.1%), with Parents constituting just 6.5% ($n = 4$) of the sample population. There is another category of stakeholders, comprising a mix bag of industry experts, policy makers, government officials, among others ($n = 10$; 16.1%). Since the focus is on educational evaluation, it is logical that the bulk of the evaluation is on students and teachers ($n = 52$; 83.9%).

Evaluation-oriented characteristics

Evaluation levels

Mark (2008) categorized evaluation context into three sub-categories of *Societal*, *Organizational* and *Evaluation-specific* (local). In the curriculum field, Deng (2012) specified three curriculum-making domains of *policy/institutional*, *programmatic* and *classroom*; these three domains correspond closely with Mark's (2008) sub-categories. Within the institutional domain, the objective is to synchronize educational goals with societal aspirations, which inherently involves an ideological approach. An example of a study at the Institutional level is Zhang et al. (2020). Since higher education impacts the social development of the nation, this paper seeks to study the performance of Higher Education Institutes (HEIs) and the resource allocation to the different HEIs, so as ascertain the "development status of higher education more clearly and to narrow the regional gap in the field of higher education more effectively" (p. 2). In the programmatic domain, the task involves converting the principles outlined by the institutional domain into practical and applicable resources for schools, such as curriculum guides and teaching materials. An example of a study at the Programmatic level is Han (2018). This is a study of the validity of a Self-assessment (SA) and peer assessment (PA) undergraduate course, with two group of participants – student-participants coming from two different classes, and rater-participants from among lecturers and postgraduate students. The intent was to ascertain the value of programme for undergraduate students. In the classroom domain, the effort is centred on converting the curriculum defined at the programmatic level into actual learning experiences for students. An example of a study at the Classroom level is Hung (2018) which looked at how an elementary school teacher implemented peer assessment in her class, with a post-intervention survey of the students and a semi-structured instructor interview with the teacher. This study focuses on how the classroom teacher enacted an intervention for her own class. Since this Evaluation review focuses on Education, Deng's (2012) domains allow a better representation of what is evaluated and at which level.

About 73% of the 62 studies were undertaken at the Institutional level ($n = 45$), followed by the Programmatic level ($n = 11$; 17.7%), and the Classroom level ($n = 6$; 9.7%).

Evaluation focus/topic

This scoping review reveals a significant emphasis on learning and teaching, with 19 instances highlighting this area ($n = 19$) (see Table 6). This prevalence underscores the centrality of pedagogical methodologies and learning outcomes in the discourse on educational effectiveness. Such a focus suggests a trend towards evaluating educational environments not merely through outcomes but through a critical examination of the teaching and learning processes themselves. The Confucian heritage, a dominant cultural ethos in countries like China, Korea, and Japan, profoundly reveres education, emphasizing respect for teachers and the value of learning (Li, 2003). This cultural

dimension could be echoed in the substantial body of research focusing on pedagogical methods and learning outcomes, aligning with long-standing cultural values. This is further complemented by the notable emphasis on capacity, evident in 12 instances, which points to a growing recognition of the importance of developing and enhancing institutional and individual capabilities in education settings ($n = 12$). It is interesting to note that even though there are 19 instances on learning and teaching, which tends to be “classroom-related”, yet there is only 9.7% which are Classroom level studies. This is because a number of studies categorized as “learning and teaching” are similar to Nuncio et al. (2020). Nuncio et al. (2020) studied the implementation of an E-Learning Outreach Research Project across different schools, how teachers implemented it, what students gained from it, how attitudes towards the use of technology in learning changed as a result. Even though it is “learning and teaching” in focus, it is at the Programmatic level.

Table 6. Evaluation focus/topic.

Evaluation focus/topic	Number
Learning/teaching	19
Capacity	12
Assessment	12
Performance/quality	9
Policy	5
Evaluation tool	3
Evaluation practice	2
Leadership	1

Performance and quality ($n = 9$), as well as assessment ($n = 12$), continue to be prominent areas, emphasizing the ongoing importance of evaluating both organizational and individual performance in education. This prevalence on evaluating assessment underscores the pivotal role that assessment plays in both gauging and enhancing educational outcomes. The methodologies used in these studies predominantly revolve around direct measures of student performance, thereby providing concrete data to inform educational practices. Furthermore, the emphasis on assessment is also reflective of a broader trend in educational evaluation, where tangible outcomes and quantifiable measures are increasingly favoured to evaluate both organizational and individual performance in educational settings. Globalization and the influence of international benchmarks, such as PIRLS² and PISA, have a pronounced impact on educational evaluation research in Asia. The drive to meet international educational standards often propels research towards empirical assessments of teaching methodologies and learning outcomes (Liu et al., 2020; Zhao, 2009). This global-local interplay highlights the challenge of aligning with international educational trends while addressing unique local educational needs in Asian countries.

In contrast, the areas of leadership³ ($n = 1$), evaluation⁴ practice ($n = 2$) and specific evaluation⁵ tools ($n = 3$) are notably less represented in the literature. This disparity indicates a potential oversight in the broader narrative of educational evaluation, where the roles of leadership in shaping educational policies and practices and the nuances of evaluation methodologies might be under-explored. This gap highlights a potential area for further research, as leadership is crucial in shaping the educational environment (Cravens & Hallinger, 2012; Liu et al., 2020). Moreover, the studies evaluating evaluation practices and tools will enhance the rigour of the field.

The allocation of research funding and the prevailing academic discourse significantly steer the thematic focus in Asian educational research. With a traditional emphasis on teaching and learning, research in these areas is often favoured due to aligned funding opportunities and academic incentives (Huang, 2015; Shao & Shen, 2012). In contrast, less emphasis on educational leadership or novel evaluation tools may reflect limited funding or academic recognition in these domains.

Evaluation perspective

Fifty-three of the 62 studies are external evaluations. Most of the papers are not commissioned by stakeholders and tend to be driven by a research agenda. Research and evaluation seek different ends (Fitzpatrick et al., 2011). Research is meant to add knowledge to the field and contribute to the growth of theory within that field, whereas that is only a secondary consideration in evaluation studies (Fitzpatrick et al., 2011; Mark et al., 2000). Even though 50 of these papers were independently initiated and not requested by stakeholders, they still contribute to the advancement of knowledge because they are published in Evaluation journals overseen by field experts. Therefore, they are recognized as Evaluation research, meaning they aim to both assess and generate new insights, embodying the concept of research-oriented evaluation. This highlights the significant role of independent research in driving theoretical advancements and broadening the scope of educational evaluation. By not being limited to stakeholder-driven agendas, independent studies are free to explore innovative methodologies, diverse contexts, and emerging issues, thereby enriching the field with a wider array of perspectives and deepening our understanding of educational phenomena across different settings.

The remaining 9 papers are Internal research-oriented evaluations.

Conclusion

This review represents an important step towards creating a resource that evaluation researchers and scholars can use to traverse the wide and diverse educational landscape of Asia, and to provide a guide in exploring potential and empirical research areas to tackle prevalent issues in evaluation. Despite the modest number of studies reviewed, the scoping review highlights a diversity of evaluation topics of educational interest undertaken to date. We hope it contributes valuable insights into the efficacy of various pedagogical and innovative practices, guide the shaping of future educational strategies and offer directions in policymaking. This notwithstanding, the current landscape reveals significant gaps in educational evaluation research, particularly in areas of leadership and specific evaluation tools. Of note too is a lack of research in the use of evaluation theories or models to frame and appraise educational problems posed. We observed that much of the work to date leans towards research-oriented evaluation in knowledge generation. Addressing these gaps presents opportunities for a more comprehensive understanding of educational effectiveness. Research on educational leadership, for instance, could provide insights into the impact of leadership styles on educational outcomes in diverse Asian settings (Cravens & Hallinger, 2012); their respective roles in shaping formative evaluation processes; and the ways in each leadership type can offer different affordances to shape respective evaluation discourses. Similarly, focusing on the development and evaluation of innovative educational tools could enhance teaching practices and learning experiences, catering to the unique needs of different Asian educational systems (Nguyen et al., 2006; Wang, 2013; Wang & Lin, 2019). Testing the efficacy of prominent evaluation theories, expounded in the West, can reveal, and expand the rigour, relevance, and utility of these models in the culturally varied Asian educational landscape.

This scoping review provides an impetus by bringing together a body of education evaluation research undertaken in the Asian context. It aims to draw international attention to the state of educational evaluation work done in this region. In so doing, it seeks to offer direction to the development of an evaluation research agenda. Moving forward we think it is important to expand the review of educational evaluation in Asia to additional major education and educational psychology journals, to provide further empirical evidence to the thrust of observations made in this review and to examine if there are uncover areas of work that may not be evident in this current work. It would be remiss if such scholarly work were excluded in our efforts to develop and consolidate a comprehensive database of educational evaluation research in Asia.

Notes

1. Even though Hong Kong and Macao are part of China, due to their histories, they have different educational systems from mainland China. As such, they are reviewed as distinct cities.
2. PIRLS: Progress in International Reading Literacy Study; PISA: Programme for International Student Assessment.
3. An example is Liu et al. (2021), where the paper explores the relationship between distributed leadership and teacher job satisfaction in the Chinese educational context, highlighting the crucial role of distributed leadership in educational reforms aimed at improving teacher work conditions and overall school effectiveness.
4. An example is Al Hudib et al. (2016), where the paper details two large-scale evaluation initiatives, and discusses the importance of interaction between people from different cultures, learning about professional standards of evaluation, and adapting to different cultures.
5. An example is Sato (2012), where the paper expounded on “a method to measure the effect, impact and efficiency of treatment through using the standardized mean difference (SMD) of the outcomes between the treatment and the control groups” (p. 231).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Jason Loh is a Senior Lecturer and teacher educator at the National Institute of Education, Nanyang Technological University, Singapore (NIE/NTU). He is an instructional researcher who studies teachers and teaching in Primary and Secondary classrooms, largely due to his own experiences as a teacher in primary (English Language, Mathematics & Social Studies) and secondary (English Language & History) education contexts.

Wan Har Chong is an Associate Professor with the Psychology and Child & Human Development department, at the National Institute of Education, Nanyang Technological University, Singapore (NIE/NTU). She teaches educational psychology, counselling psychology and program evaluation at in-service and post-graduate levels. She has worked in various community-based evaluation and implementation projects as a principal and co-principal investigator, consultant and trainer. Dr Chong co-heads the Centre for Assessment and Evaluation at NIE.

Christina Ratnam-Lim is a Senior Lecturer with the Policy, Curriculum and Leadership department, at the National Institute of Education, Nanyang Technological University, Singapore (NIE/NTU). She specializes in curriculum design, implementation, and evaluation, as well as exploring perspectives on learning and the professional development of teachers.

Michelle Tan is a Senior Research Assistant at the Centre for Assessment & Evaluation at the National Institute of Education, Nanyang Technological University (NIE-NTU). She holds a Bachelor's degree in Sociology from NTU and a Master of Education in Developmental Psychology from NIE. Tan began her career as a contract teacher in a primary school. Her research interests focus on evaluating educational initiatives and exploring the intersection of sociological themes with developmental psychology.

Phoebe Ow is a Senior Research Assistant at the National Institute of Education, Nanyang Technological University, Singapore. Her research interests include evaluation and social-emotional learning in areas of climate change and sustainability education.

ORCID

Jason Loh  <http://orcid.org/0000-0002-6162-8550>
 Wan Har Chong  <http://orcid.org/0000-0002-8828-6634>
 Christina Lim-Ratnam  <http://orcid.org/0000-0001-6022-2703>
 Michelle Tan  <http://orcid.org/0009-0008-5397-4307>
 Phoebe Ow  <http://orcid.org/0009-0005-8551-117X>

References

- Alkin, M.C. (2003). Evaluation theory and practice: Insights and new directions. In C.A. Christie (Ed.), *The practice-theory relationship in evaluation* (pp. 81–89). Jossey-Bass.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>

- Brandon, P.R., & Fukunaga, L.L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, 35(1), 26–44. <https://doi.org/10.1177/1098214013503699>
- Cheng, S.-H., & King, J.A. (2017). Exploring organizational evaluation capacity and evaluation capacity building: A delphi study of Taiwanese elementary and junior high schools. *American Journal of Evaluation*, 38(4), 521–539. <https://doi.org/10.1177/1098214016672344>
- Christie, C.A. (2011). Advancing empirical scholarship to further develop evaluation theory and practice. *Canadian Journal of Program Evaluation*, 26(1), 1–18. <https://doi.org/10.3138/cjpe.026.001>
- Christie, C.A., & Fleischer, D.N. (2010). Insight into evaluation practice: A content analysis of designs and methods used in evaluation studies published in North American evaluation-focused journals. *American Journal of Evaluation*, 31(3), 326–346. <https://doi.org/10.1177/1098214010369170>
- Colquhoun, H.L., Levac, D., O'Brien, K.K., Straus, S., Tricco, A.C., Perrier, L., Kastner, M., & Moher, D. (2014). Scoping reviews: Time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology*, 67(12), 1291–1294. <https://doi.org/10.1016/j.jclinepi.2014.03.013>
- Coryn, C.L.S., Noakes, L.A., Westine, C.D., & Schroter, D.C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226. <https://doi.org/10.1177/1098214010389321>
- Coryn, C.L.S., Ozeki, S., Wilson, L.N., Greenman, G.D., Schröter, D.C., Hobson, K.A., Azzam, T., & Vo, A.T. (2016). Does research on evaluation matter? Findings from a survey of American evaluation association members and prominent evaluation theorists and scholars. *American Journal of Evaluation*, 37(2), 159–173. <https://doi.org/10.1177/1098214015611245>
- Coryn, C.L.S., Wilson, L.N., Westine, C.D., Hobson, K.A., Ozeki, S., Fiekowsky, E.L., Greenman, G.D., & Schröter, D.C. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation*, 38(3), 329–347. <https://doi.org/10.1177/1098214016688556>
- Cravens, X.C., & Hallinger, P. (2012). School leadership and change in East Asia: Building capacity for education reform. *Peabody Journal of Education*, 87(2), 157–161. <https://doi.org/10.1080/0161956X.2012.664461>
- Daudt, H.M., van Mossel, C., & Scott, S.J. (2013). Enhancing the scoping study methodology: A large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Medical Research Methodology*, 13(1), 48. <https://doi.org/10.1186/1471-2288-13-48>
- Deng, Z. (2012). Curriculum making in the new curriculum reform: Structure, process and meaning. In H.B. Ying & J.C. K. Lee (Eds.), *Curriculum reform in China: Current issues and challenges* (pp. 31–46). Nova Science.
- Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson.
- Gargani, J. (2012). *The future of evaluation: 10 predictions*. <https://evalblog.com/2012/01/30/the-future-of-evaluation-10-predictions/>
- Goodwin, A., & Low, E. (2020). *Students at the center of education reform in Singapore*. Oxford research encyclopedia of education. <https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-829>
- Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Huang, F. (2015). Building the world-class research universities: A case study of China. *Higher Education*, 70(2), 203–215. <https://doi.org/10.1007/s10734-015-9876-8>
- Hyland, K. (2023). Enter the dragon: China and global academic publishing. *Learned Publishing*, 36(3), 394–403. <https://doi.org/10.1002/leap.1545>
- Johnson, K., Greenesid, L.O., Toal, S.A., King, J.A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410. <https://doi.org/10.1177/1098214009341660>
- Kennedy, K.J., & Lee, J.C.K. (2008). *The changing roles of schools in Asian societies: Schools for the knowledge society*. Routledge.
- Kim, M.J. (2022). Happiness, politics and education reform in South Korea: Building 'happy human capital' for the future. *Comparative Education*, 59(4), 489–505. <https://doi.org/10.1080/03050068.2022.2147633>
- Lawaree, J., Jacob, S., & Ouimet, M. (2020). A scoping review of knowledge syntheses in the field of evaluation across four decades of practice. *Evaluation and Program Planning*, 79, 1–11. <https://doi.org/10.1016/j.evalprogplan.2019.101761>
- Levac, D., Colquhoun, H., & O'Brien, K.K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1), 69. <https://doi.org/10.1186/1748-5908-5-69>
- Li, J. (2003). U.S and Chinese cultural beliefs about learning. *Journal of Educational Psychology*, 95(2), 258–267. <https://doi.org/10.1037/0022-0663.95.2.258>
- Lin, T.-B., Wang, L.-Y., Li, J.-Y., & Chang, C. (2014). Pursuing quality education: The lessons from the education reform in Taiwan. *The Asia-Pacific Education Researcher*, 23(4), 813–822. <https://doi.org/10.1007/s40299-013-0135-4>
- Liu, Y., Yin, Y., & Wu, R. (2020). Measuring graduate students' global competence: Instrument development and an empirical study with a Chinese sample. *Studies in Educational Evaluation*, 67, 100915. <https://doi.org/10.1016/j.stueduc.2020.100915>

- Loh, J., Renandya, W.A., Pang, E., & Aryadoust, V. (2023). Language programme evaluation. In Z. Tajeddin & C. Griffiths (Eds.), *Language education programs: Perspectives on policies and practices* (pp. 37–52). Springer.
- Mallapaty, S. (2020). China bans cash rewards for publishing papers. *Nature*, 579(7797), 18. <https://doi.org/10.1038/d41586-020-00574-8>
- Mark, M.M. (2008). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N.L. Smith & P.R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). The Guilford Press.
- Mark, M.M., Henry, G.T., & Julnes, G. (2000). Evaluation: An integrated framework for understanding, guiding, and improving policies and programs. *Jossey-Bass*.
- Mazur, R., & Woodland, R.H. (2017). Evaluation of a cross-cultural training program for Pakistani educators: Lessons learned and implications for program planning. *Evaluation and Program Planning*, 62, 25–34. <https://doi.org/10.1016/j.evalprogplan.2017.02.011>
- Mertens, D.M., & Wilson, A.T. (2019). *Program evaluation theory and practice: A comprehensive guide* (2nd ed.). The Guilford Press.
- Ministry of Education, India. (n.d.). *National education policy 2020*. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
- Ministry of Education, Korea. (2023). *Education reform, the first step to Korea's new leap forward*. <https://www.moe.go.kr/boardCnts/fileDown.do?m=0201&s=english&fileSeq=cf5bcbd2b87360af91f57ea47514a27a>
- Nguyen, P.-M., Terlouw, C., & Pilot, A. (2006). Culturally appropriate pedagogy: The case of group learning in a Confucian heritage culture context. *Intercultural Education*, 17(1), 1–19. <https://doi.org/10.1080/14675980500502172>
- Pascua, L., & Chang, C.H. (2015). Using intervention-oriented evaluation to diagnose and correct students' persistent climate change misconceptions: A Singapore case study. *Evaluation and Program Planning*, 52, 70–77. <https://doi.org/10.1016/j.evalprogplan.2015.04.001>
- Patton, M.Q. (2015). *Qualitative research & evaluation methods* (4th ed.). Sage.
- Peng, S.S., & Lee, J.C.K. (2009). *Educational evaluation in East Asia: Emerging issues and challenges*. Nova Science Publishers, Inc.
- Peterson, J., Pearce, P.F., Ferguson, L.A., & Langford, C.A. (2017). Understanding scoping reviews: Definition, purpose, and process. *Journal of the American Association of Nurse Practitioners*, 29(1), 12–16. <https://doi.org/10.1002/2327-6924.12380>
- Pham, M.T., Rajic, A., Greig, J.D., Sargeant, J.M., Papadopoulos, A., & McEwen, S.A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, 5(4), 371–385. <https://doi.org/10.1002/jrsm.1123>
- Phillips, D.C. (2018). The many functions of evaluation in education. *Education Policy Analysis Archives*, 26(46), 46. <https://doi.org/10.14507/epaa.26.3811>
- Schwandt, T.A. (2011). From the editor. *American Journal of Evaluation*, 32(1), 7. <https://doi.org/10.1177/1098214010390920>
- Shao, J.-F., & Shen, H.-Y. (2012). Research assessment and monetary rewards: The overemphasized impact factor in China. *Research Evaluation*, 21(3), 199–203. <https://doi.org/10.1093/reseval/rvs011>
- Smith, N.L., Brandon, P.R., Hwalek, M., Kistler, S.J., Labin, S.N., Rugh, J., Thomas, V., & Yarnall, L. (2011). Looking ahead: The future of evaluation. *American Journal of Evaluation*, 32(4), 565–599. <https://doi.org/10.1177/1098214011421412>
- Stufflebeam, D.L., & Coryn, C.L.S. (2014). *Evaluation theory, models, and applications* (2nd Ed). Jossey Bass.
- United Nations Population Fund. (2023). *Population trends*. <https://asiapacific.unfpa.org/en/populationtrends>
- Vallin, L.M., Philippoff, J., Pierce, S., & Brandon, P.R. (2015). Research-on-evaluation articles published in the American journal of evaluation, 1998–2014. In P.R. Brandon (Ed.), *Research on evaluation. New directions for evaluation* (Vol. 148, pp. 7–15). Jossey-Bass.
- Wang, J. (2013). Understanding the Chinese learners from a perspective of confucianism. In L. Jin & M. Cortazzi (Eds.), *Researching cultures of learning* (pp. 61–79). Palgrave Macmillan.
- Wang, J., & Lin, J. (2019). Traditional Chinese views on education as perceived by international students in China: International Student attitudes and understandings. *Journal of Studies in International Education*, 23(2), 195–216. <https://doi.org/10.1177/1028315318797356>
- Wanzer, D.L. (2021). What is evaluation? Perspectives of how evaluation differs (or not) from research. *American Journal of Evaluation*, 42(1), 28–46. <https://doi.org/10.1177/1098214020920710>
- Yakavets, N., Winter, L., Malone, K., Zhontayeva, Z., & Khamidulina, Z. (2023). Educational reform and teachers' agency in reconstructing pedagogical practices in Kazakhstan. *Journal of Educational Change*, 24(4), 727–757. <https://doi.org/10.1007/s10833-022-09463-5>
- Yamanaka, S., & Suzuki, K.H. (2020). Japanese education reform towards twenty-first century education. In F.M. Reimers (Ed.), *Audacious education purposes: How governments transform the goals of education systems* (pp. 81–103). Springer Open.
- Zhao, Y. (2009). *Catching up or leading the way: American education in the age of globalization*. ASCD.

Appendix A

Al-Qatawneh, S., & Al Rawashdeh, A. (2019). Gender representation in the Arabic language textbook for the ninth grade approved by the Ministry of Education for use in schools in the United Arab Emirates (UAE). *Studies in Educational Evaluation*, 60, 90-98. <https://doi.org/10.1016/j.stueduc.2018.12.001>

Bai, B., Shen, B., & Mei, H. (2020). Hong Kong primary students' self-regulated writing strategy use: Influences of gender, writing proficiency, and grade level. *Studies in Educational Evaluation*, 65, 100839. <https://doi.org/10.1016/j.stueduc.2020.100839>

Chan, C. K., & Luk, L. Y. (2022). Eight years after the 3-3-4 curriculum reform: The current state of undergraduates' holistic competency development in Hong Kong. *Studies in Educational Evaluation*, 74, 101168. <https://doi.org/10.1016/j.stueduc.2022.101168>

Chan, C. K., & Yeung, N. C. J. (2021). To assess or not to assess holistic competencies—Student perspectives in Hong Kong. *Studies in Educational Evaluation*, 68, 100984. <https://doi.org/10.1016/j.stueduc.2021.100984>

Chen, H., Su, Y., Zheng, Q., & Li, L. (2019). Towards an operationalized and effective school self-evaluation system: An ongoing action research study in Chinese secondary schools since 2011. *Studies in Educational Evaluation*, 60, 117-129. <https://doi.org/10.1016/j.stueduc.2018.12.005>

Chen, H., & Tseng, H. (2012). Factors That Influence Acceptance of Web-Based e-Learning Systems for the In-Service Education of Junior High School Teachers in Taiwan. *Evaluation and Program Planning*, 35, 398-406. <https://doi.org/10.1016/j.evalprogplan.2011.11.007>

Chen, L. M. (2018). School-based anti-bullying strategies in Taiwan and their effects. *Studies in Educational Evaluation*, 59, 218-223. <https://doi.org/10.1016/j.stueduc.2018.09.004>

Chen, X. (2022). The effects of individual-and class-level achievement on attitudes towards mathematics: An analysis of Hong Kong students using TIMSS 2019. *Studies in Educational Evaluation*, 72, 101113. <https://doi.org/10.1016/j.stueduc.2021.101113>

Cheng, S.-H., & King, J. A. (2017). Exploring Organizational Evaluation Capacity and Evaluation Capacity Building: A Delphi Study of Taiwanese Elementary and Junior High Schools. *American Journal of Evaluation*, 38(4), 521-539. <https://doi.org/10.1177/1098214016672344>

Courtney, M. G. R., Rakhymbayeva, Z., Shilibekova, A., Ziyedenova, D., Soltangazina, S., Muratkyzy, A., Goodman, B., Olzhayeva, A. (2022). Kazakh, Russian, and Uyghur child language literacy: The role of the updated curriculum on longitudinal growth trajectories in Kazakhstan. *Studies in Educational Evaluation*, 75, 101189. <https://doi.org/10.1016/j.stueduc.2022.101189>

Ellahi, A., & Zaka, B. (2015). Analysis of higher education policy frameworks for open and distance education in Pakistan. *Evaluation Review*, 39(2), 255-277. <https://doi.org/10.1177/0193841X15570046>

Fan, J., Frost, K., & Liu, B. (2020). Teachers' involvement in high-stakes language assessment reforms: The case of Test for English Majors (TEM) in China. *Studies in Educational Evaluation*, 66, 100898. <https://doi.org/10.1016/j.stueduc.2020.100898>

Fan, T., & Yan, X. (2020). Diagnosing English reading ability in Chinese senior high schools. *Studies in Educational Evaluation*, 67, 100931. <https://doi.org/10.1016/j.stueduc.2020.100931>

Guo, J. (2018). Building bridges to student learning: Perceptions of the learning environment, engagement, and learning outcomes among Chinese undergraduates. *Studies in Educational Evaluation*, 59, 195-208. <https://doi.org/10.1016/j.stueduc.2018.08.002>

Guo, J., Yang, L., & Shi, Q. (2017). Effects of perceptions of the learning environment and approaches to learning on Chinese undergraduates' learning. *Studies in Educational Evaluation*, 55, 125-134. <https://doi.org/10.1016/j.stueduc.2017.09.002>

Han, C. (2018). A longitudinal quantitative investigation into the concurrent validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course. *Studies in Educational Evaluation*, 58, 187-196. <https://doi.org/10.1016/j.stueduc.2018.01.001>

Hu, M. C., Hung, S. C., Lo, H. C., & Tseng, Y. C. (2016). Determinants of university–industry research collaborations in Taiwan: The case of the National Tsing Hua University. *Research Evaluation*, 25(2), 121–135. <https://doi.org/10.1093/reseval/rvw005>

Huang, L., Zhang, T., & Huang, Y. (2020). Effects of school organizational conditions on teacher professional learning in China: The mediating role of teacher self-efficacy. *Studies in Educational Evaluation*. 66. 100893. <https://doi.org/10.1016/j.stueduc.2020.100893>

Hudib, H. A., Cousins, J. B., Oza, J., Lakshminarayana, U., & Bhat, V. D. (2016). A cross-cultural evaluation conversation in India: Benefits, challenges, and lessons learned. *Canadian Journal of Program Evaluation*, 30(3), 329-343. <https://doi.org/10.3138/cjpe.30.3.06>

Hung, Y. J. (2018). Group peer assessment of oral English performance in a Taiwanese elementary school. *Studies in Educational Evaluation*, 59, 19-28. <https://doi.org/10.1016/j.stueduc.2018.02.001>

Ifitkhar, S., Fu, Y., Naureen, S., Cao, Y., & Zhou, C. (2022). Cascading of teachers training at higher education in Pakistan: An evaluation of a faculty professional development program. *Evaluation and Program Planning*, 94, 102130. <https://doi.org/10.1016/j.evalprogplan.2022.102130>

- Jamil, M. G. (2018). Technology-enhanced teacher development in rural Bangladesh: A critical realist evaluation of the context. *Evaluation and program planning*, 69, 1-9. <https://doi.org/10.1016/j.evalprogplan.2018.04.002>
- Leung, K. C., Leung, F. K., & Zuo, H. (2014). A study of the alignment of learning targets and assessment to generic skills in the new senior secondary mathematics curriculum in Hong Kong. *Studies in Educational Evaluation*, 43, 115-132. <https://doi.org/10.1016/j.stueduc.2014.09.002>
- Li, F., Miao, Y., & Ding, J. (2015) Tracking the development of disciplinary structure in China's top research universities (1998-2013). *Research Evaluation*, 24(3), 312-324, <https://doi.org/10.1093/reseval/rvv013>
- Lin, L., Lam, J. W. I., & Tse, S. K. (2019). Test takers' strategy use and L2 Chinese reading test performance in mainland China: A structural equation approach. *Studies in Educational Evaluation*, 60, 189-198. <https://doi.org/10.1016/j.stueduc.2019.01.002>
- Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation*, 48, 57-66. <https://doi.org/10.1016/j.stueduc.2016.01.002>
- Liu, S., Keeley, J.W., Sui, Y., & Sang, L. (2021). Impact of distributed leadership on teacher job satisfaction in China: The mediating roles of teacher autonomy and teacher collaboration. *Studies in Educational Evaluation*, 71, 101099. <https://doi.org/10.1016/j.stueduc.2021.101099>
- Liu, Y., Yin, Y.M., & Wu, R. (2020). Measuring graduate students' global competence: Instrument development and an empirical study with a Chinese sample. *Studies in Educational Evaluation*, 67, 100915. <https://doi.org/10.1016/j.stueduc.2020.100915>
- Ma, X., Fan, M., & Luo, X. (2018). Consistency of science achievement across science subjects among Chinese students and schools. *Studies in Educational Evaluation*, 56, 124-132. <https://doi.org/10.1016/j.stueduc.2017.12.006>
- Mariel, P., Sanko, N., & Vega-Bayo, A. (2021). The effect of the Free High School Tuition law on upper-secondary school choice in Japan. *Studies in Educational Evaluation*, 70, 101065. <https://doi.org/10.1016/j.stueduc.2021.101065>
- Mascareño, M. (2019). Constellations of risk factors as moderators of the impact of a randomized intervention on students' reading skills in rural India. *Studies in Educational Evaluation*, 60, 66-77. <https://doi.org/10.1016/j.stueduc.2018.11.007>
- Mazur, R., & Woodland, R. H. (2017). Evaluation of a cross-cultural training program for Pakistani educators: Lessons learned and implications for program planning. *Evaluation and Program Planning*, 62, 25-34. <https://doi.org/10.1016/j.evalprogplan.2017.02.011>
- Min, S., & He, L. (2020). Test fairness: Examining differential functioning of the reading comprehension section of the GSEEE in China. *Studies in Educational Evaluation*, 64, 100811. <https://doi.org/10.1016/j.stueduc.2019.100811>
- Mundia, L., & Metussin, H. (2019). Exploring factors that improve mathematics achievement in Brunei. *Studies in Educational Evaluation*, 60, 214-222. <https://doi.org/10.1016/j.stueduc.2018.10.003>
- Nuncio, R. V., Arcinas, M. M., Lucas, R. I. G., Alontaga, J. V. Q., Neri, S. G. T., & Carpena, J. M. (2020). An E-learning outreach program for public schools: Findings and lessons learned based on a pilot program in Makati City and Cabuyao City, Laguna, Philippines. *Evaluation and Program Planning*, 82, 101846. <https://doi.org/10.1016/j.evalprogplan.2020.101846>
- Pascua, L., & Chang, C. H. (2015). Using intervention-oriented evaluation to diagnose and correct students' persistent climate change misconceptions: A Singapore case study. *Evaluation and Program Planning*, 52, 70-77. <https://doi.org/10.1016/j.evalprogplan.2015.04.001>
- Peeraer, J., & Van Petegem, P. (2015). Integration or transformation? Looking in the future of Information and Communication Technology in education in Vietnam. *Evaluation and program planning*, 48, 47-56. <https://doi.org/10.1016/j.evalprogplan.2014.09.005>
- Razak, N. A., Jalil, H. A., Krauss, S. E., & Ahmad, N. A. (2018). Successful implementation of information and communication technology integration in Malaysian public schools: An activity systems analysis approach. *Studies in Educational Evaluation*, 58, 17-29. <https://doi.org/10.1016/j.stueduc.2018.05.003>
- Sato, Y. (2012). A trial of impact evaluation utilizing effect size statistics: Its application to the evaluation of Japan's foreign student policy. *Evaluation*, 18(2), 230-245. <https://doi.org/10.1177/1356389012443163>
- Shayeste-Far, P., & Kiany, G. R. (2018). Iran's university admission programme at change: Policies, prospects and pitfalls from planners' perspectives. *Studies in Educational Evaluation*, 56, 112-123. <https://doi.org/10.1016/j.stueduc.2017.11.006>
- Shen, B., Bai, B., & Xue, W. (2020). The effects of peer assessment on learner autonomy: An empirical study in a Chinese college English writing class. *Studies in Educational Evaluation*, 64, 100821. <https://doi.org/10.1016/j.stueduc.2019.100821>
- Sun, H., & Zhang, J. (2022). Assessment Literacy of College EFL Teachers in China: Status Quo and Mediating Factors. *Studies in Educational Evaluation*, 74, 101157. <https://doi.org/10.1016/j.stueduc.2022.101157>
- Simsar, A., Doğan, Y., & Sezer, G. (2021). The ecocentric and anthropocentric attitudes towards different environmental phenomena: A sample of Syrian refugee children. *Studies in Educational Evaluation*, 70, 101005. <https://doi.org/10.1016/j.stueduc.2021.101005>
- Siripitakchai, N., & Miyazaki, K. (2015). Assessment of research strengths using co-citation analysis: The case of Thailand national research universities. *Research Evaluation*, 24(4), 420-439. <https://doi.org/10.1093/reseval/rvv018>

- Sun, Y., & Cheng, L. (2014). Teachers' grading practices: meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21, 326 - 343. <https://doi.org/10.1080/0969594X.2013.768207>
- Ullah, R., Richardson, J. T., Malik, R. A., & Farooq, S. (2016). Perceptions of the learning environment, learning preferences, and approaches to studying among medical students in Pakistan. *Studies in Educational Evaluation*, 50, 62-70. <https://doi.org/10.1016/j.stueduc.2016.07.001>
- Vuong, Q. H., Bui, A. T., La, V. P., Nguyen, M. H., Pham, H. H., Pham, T. H., . . . & Ho, M. T. (2021). Mirror, mirror on the wall: Is economics the fairest of them all? An investigation into the social sciences and humanities in Vietnam. *Research Evaluation*, 30(1), 57-72. <https://doi.org/10.1093/reseval/rvaa036>
- Wang, L., Lee, I., & Park, M. (2020). Chinese university EFL teachers' beliefs and practices of classroom writing assessment. *Studies in Educational Evaluation*, 66, 100890. <https://doi.org/10.1016/j.stueduc.2020.100890>
- Wang, T. Y., & Hsieh, F. J. (2017). Taiwanese high school students' perspectives on effective mathematics teaching behaviors. *Studies in Educational Evaluation*, 55, 35-45. <https://doi.org/10.1016/j.stueduc.2017.06.001>
- Wu, X., Ma, X., & Luo, X. (2022). Can less be more? Effects of study load on science achievement among Chinese eighth-grade students. *Studies in Educational Evaluation*, 75, 101219. <https://doi.org/10.1016/j.stueduc.2022.101219>
- Xie, Q. (2022). The role of parents in the school accountability system: Insights from a Hong Kong case. *Studies in Educational Evaluation*, 72, 101114. <https://doi.org/10.1016/j.stueduc.2021.101114>
- Xu, S. (2021). Processes and effects of test preparation for writing tasks in a high-stakes admission test in China: Implications for test takers. *Studies in Educational Evaluation*, 70, 101015. <https://doi.org/10.1016/j.stueduc.2021.101015>
- Yan, Z., & Brown, G. T. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, 68, 100985. <https://doi.org/10.1016/j.stueduc.2021.100985>
- Yasmin, M., Naseem, F., & Masso, I. C. (2019). Teacher-directed learning to self-directed learning transition barriers in Pakistan. *Studies in Educational Evaluation*, 61, 34-40. <https://doi.org/10.1016/j.stueduc.2019.02.003>
- Yau, A. H. Y., Yeung, M. W. L., & Lee, C. Y. P. (2022). A co-orientation analysis of teachers' and students' perceptions of online teaching and learning in Hong Kong higher education during the COVID-19 pandemic. *Studies in Educational Evaluation*, 72, 101128. <https://doi.org/10.1016/j.stueduc.2022.101128>
- Yoo, J. (2019). The impact of conflict among political actors on implementing South Korea's new teacher evaluation policy: A case study with implications for education policymaking. *Studies in Educational Evaluation*, 61, 94-104. <https://doi.org/10.1016/j.stueduc.2019.03.006>
- Yu, S., Zhou, N., Zheng, Y., Zhang, L., Cao, H., & Li, X. (2019). Evaluating student motivation and engagement in the Chinese EFL writing context. *Studies in Educational Evaluation*, 62, 129-141. <https://doi.org/10.1016/j.stueduc.2019.06.002>
- Yue, J. J., & Chen, G. (2019). An exploration of the structure of pharmaceutical mentors' competence at universities in Western China. *Studies in Educational Evaluation*, 63, 1-8. <https://doi.org/10.1016/j.stueduc.2019.06.004>
- Zhang, H., Bournot-Trites, M. (2021). The long-term washback effects of the National Matriculation English Test on college English learning in China: Tertiary student perspectives. *Studies in Educational Evaluation*, 68, 1-21. <https://www.sciencedirect.com/science/article/pii/S0191491X21000031>
- Zhang, G., Wu, J., & Zhu, Q. (2020). Performance evaluation and enrollment quota allocation for higher education institutions in China. *Evaluation Program Planning*, 81, 101821. <https://doi.org/10.1016/j.evalprogplan.2020.101821>
- Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2017). Classroom assessment in the eyes of Chinese primary mathematics teachers: A review of teacher-written papers. *Studies in Educational Evaluation*, 52, 42-54. <https://doi.org/10.1016/j.stueduc.2016.12.002>