International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 2, April 2026, pp. 955~965
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i2.36369

Examining the ontological and epistemological differences between evaluation and assessment in the education context

Chew-Hung Chang, Christina Ratnam-Lim, Wong Hwei Ming, Phoebe Ow, Joshua Wong Jia Hao
National Institute of Education, Nanyang Technological University, Singapore, Singapore

Article Info

Article history:

Received Jul 27, 2025
Revised Jan 7, 2026
Accepted Feb 1, 2026

Keywords:

Assessment
Definition
Epistemology
Evaluation
Stakeholder

ABSTRACT

In the field of education, assessment and evaluation (AE) are defining strategies for improving school practices with the overall goal of enhancing student outcomes. While AE is understood and considered by educators as two distinctly different domains, they are often used interchangeably among practitioners in the field. The aim of this paper is three-fold: to identify and clarify the epistemological and ontological differences between AE; develop operational definitions of AE that are transferable from research to classroom through a comprehensive literature review; clarify any misconceptions between the two domains, if any. A comprehensive meta-synthesis of literature from articles published in 32 journals between 2014 and 2024 revealed four frames to distinguish assessment from evaluation: i) information gathered; ii) methodology; iii) purposes and outcomes; and iv) stakeholders. To corroborate these conclusions, we also conducted a social lab which is based on the principle of the Delphi method with 11 participants. Additionally, insights from two social lab sessions revealed that apart from the ontological distinctions between AE, there were also epistemological distinctions. Eventually, three frames for exploring the difference between AE emerged from the data: ontology, epistemology, and stakeholders. By refining and developing this area of AE, the research hopes to contribute to a more informed and integrated educational landscape, encouraging further work in areas of AE.

Corresponding Author:

Chew-Hung Chang
National Institute of Education, Nanyang Technological University
1 Nanyang Walk, Singapore 637616, Singapore
Email: chewhung.chang@nie.edu.sg

1. INTRODUCTION

Assessment and evaluation (AE) are two prominent concepts that are frequently used amongst various educational stakeholders. Traditionally, AE has focused on how to improve student learning [1], [2]. However, these concepts have expanded to become central key ideas in educational policy, seen as "indispensable tools for improvement, accountability, educational planning and policy development" [3] more recently. These concepts serve distinct yet interconnected roles–assessment focuses on measuring student learning [4], [5], while evaluation examines the effectiveness of programs, teaching strategies and policies besides overall student learning [6], [7]. Together, they provide critical insights that drive student learning outcomes–assessment at the individual or classroom level, and evaluation at the programmatic or systemic level. These insights, though derived at different scales, collectively inform curriculum development, instructional improvements, and education policy decisions.

However, both concepts have evolved over time that are shaped by the educational paradigms within which researchers, policy makers and practitioners operate. These paradigms are tied to perspectives

that influence how educators define the purpose, processes, and goals of education. These different paradigms offer different views on learning, the role of teachers and students, and the nature of knowledge. They are essential for understanding and shaping educational practices, policies, and research.

An educational paradigm that views knowledge as external, fixed, and transmissible from teacher to student, will emphasize quantitative measurement to ensure consistency and minimizes error. Assessment in this context, involves standardized tests and tools to capture observable learning outcomes and content mastery. Evaluation focuses on determining instruction effectiveness and program success through norm-referenced comparisons, reflecting scientific method principles of control, measurement, and prediction.

If the educational paradigm within which the researcher is operating in is more learner and learning centered [8], [9], the focus is more on the personal growth of students [10]. Assessments in this paradigm prioritize deep understanding, student agency, reflection, and development over time. Examples of assessments include exit tickets, portfolios, and learning journals which promote metacognition and self-awareness. Students evaluate their own learning and performance to build self-regulation and autonomy. Other domains of the learner's development–emotional and social–are also assessed, besides the cognitive aspects. Evaluation in this context shifts away from ranking and judgment toward guiding, supporting, and reflecting on the learning journey. The focus is on development, depth of understanding, and the process of learning, not just final outcomes. In a learner-centered, constructivist curriculum, assessment often refers to tools and methods used to gather evidence of learning (e.g., tests, rubrics, and peer reviews), while evaluation refers more broadly to how we interpret that evidence to make meaningful, holistic judgments about student growth, especially in relation to goals, values, and progress.

As Kuhn [11] suggests, ways of knowing and organizing knowledge continue to be contested and re-evaluated. These shifting epistemologies have blurred the lines between AE, as learners are no longer viewed atomistically but as embedded within complex networks. Perhaps due to their interconnectedness, where assessment usually acts as a tool and feedback to make informed decisions, conflation between the two terms is widespread and commonly used loosely with each other [12], [13].

One significant contributing factor to the conflation of the two terms includes the differing consensus and concepts found in the literature. For instance, Baehr [12] posits assessment as a term to look at how the level of quality of a performance, or outcome could be improved, while evaluation is used to look at the level of quality of a performance or outcome to enable decision-making based on the level of quality demonstrated. Conversely, Scanlan [4], suggests that educational assessment is the gathering and evaluating of data from learning activities or programs, which is often referred to as evaluation. Confusion is expounded when similar words such as 'judgements', 'decisions' and 'improvements' are used synonymously in describing the definitions of AE. Accompanied with the adoption of 'formative' and 'summative' terminologies as adjectives in both AE further contributes to the conflation between these two concepts.

In evaluating the progress and benefits of an educational program or approach, teachers primarily rely on student assessment as a critical mechanism for gauging the effectiveness and impact of the implemented pedagogical strategies [14], [15]. As such, teachers tend to see assessment as the only means of evaluation, using the two terms interchangeably to mean the same thing. Furthermore, educators commonly perceive their role within an academic setting as instructional facilitators rather than evaluators. This perception stems from traditional conceptions of evaluation as a specialized domain requiring extensive statistical analysis and controlled experimentation to ascertain program effectiveness and outcomes [3] often associating evaluation with large-scale quantitative research and quasi-experimental methodologies [16].

The conflation of AE can consequently result in erroneous mental models formed. Mental models could impair learning if built with incorrect or incomplete materials, or when ideas are connected through anomalous linkages [17]. They become challenging to resolve, tending to persist and multiply once deeply ingrained [18]. This leads to differing epistemological ideas and ontologies between the two concepts, significantly impacting the determinations and outcomes derived from AE measurements in school systems used by various stakeholders in educational AE. Such impacts involve issues, such as poor policy design, lack of analysis of unintended consequences, little capacity for educational agents to put procedures into practice, lack of an evaluation culture or deficient use of evaluation results [3].

Similarly in Singapore, conflation between the two domains is widespread within the educational field. Firstly, the two terms "assessment" and "evaluation" are used interchangeably as they evoke a similar value of gathering evidence to produce a result. Secondly, there is a lack of epistemological clarity, in how these concepts are different when used in the vernacular. Within the fields of education, the specifics of the terms transform the words into educational jargon. In order to advance their work in educational AE, the Singapore National Institute of Education (NIE) Centre for Assessment and Evaluation (CAE) has established their operational definition and guiding principles to clarify the distinctions between AE in education, as seen in Table 1.

The NIE CAE defines assessment as "the process of obtaining and reflecting on evidence of student learning to determine whether and to what extent learners have achieved their intended learning outcomes. It provides constructive information and feedback to enhance student learning, teaching practices and assessment," while evaluation "involves the systematic process of collecting empirical data and contextual information to determine the merit of an educational initiative. It allows an informed decision to be made regarding its continuation, modification or termination in response to the needs of identified stakeholders" [19].

Although a definition has been initiated by NIE CAE, the derivations of these definitions was primarily based on epistemic discussions of faculty working in the area, often occurring at meetings or informal conversations; not one based on empirical evidence from a meta-synthesis of AE literature. This paper therefore seeks to develop research informed definitions of AE that are transferable to the classroom, and enhance the work of AE, reducing the gaps between knowledge and practice of AE in the educational context. It will aim to clarify the epistemological and ontological differences between educational AE through a comprehensive review of current literature, as well as through insights gathered from a social lab involving expert perspectives (the Delphi method).

Table 1. NIE CAE's principles of AE

Assessment	Principle	Evaluation
To provide information on student learning – Allows teachers to review the effectiveness of their teaching and students to reflect on their progress in learning To enhance teaching and learning – Provides constructive information and feedback to improve assessment development, teaching practices and student learning	Why: It is needed (purpose)	To examine the merit of an educational initiative – Captures descriptive, contextual and empirical data to review the merit and quality of the educational initiative. To make a judgement call – Provides insights, informs decision, assists in planning and further understanding of the phenomenon (e.g., continuation, improvement or termination)
Principles i) Central to effective teaching and learning ii) Valid and reliable iii) Explicit, transparent and accessible iv) Promote and improve learner's capacity for self-directedness v) Take into consideration impact on learner's well-being	What: It entails (principles)	Principles i) Rigor ii) Relevance iii) Utility
Individual – Looks at the student, classroom and institutional level	Which: Unit of analysis	System – Looks at the societal, organizational and/or institutional levels, or evaluation-specific issues. – Encompasses processes and products (e.g., professional development/training programs, educational context, curriculum quality).
Collects evidence of student learning – Various forms of assessments are conducted in the classroom, at the school and national level.	How: It is conducted (modes of inquiry and types)	Systematic process of inquiry – Mixed methods incorporating the perspectives of various stakeholders with interest(s) in the outcome of the educational initiative. Types – Needs assessment/analysis, evaluability assessment, formative evaluation, process evaluation, summative evaluation, impact evaluation.
Educators and learners – Includes assessment specialists, teachers and students	Who: The intended users are	Decision-makers – Includes policymakers, program developers, curriculum/school leaders and program funders

From "our beliefs", by NIE CAE [19]

2. METHOD

2.1. Research design

The paper adopted a qualitative research approach, structured into two stages. The first stage consists of a systematic literature review of the definitions of educational AE. The second stage incorporates discussion sessions through a social lab among different stakeholders working on AE. In the context of our research, the objectives are: i) to identify and describe the epistemological and ontological differences between educational AE and ii) to develop operational definitions of AE that are transferable from research to classroom through a comprehensive literature review. These objectives inform the research questions (RQ):

- What are the epistemological and ontological differences between AE as represented in the literature?
- How can we operationalize the difference in AE into working definitions that will guide the work of the strategic growth area and beyond?

2.2. Stage 1–systematic review: meta-synthesis

The first stage of this research employs a systematic review to address the first RQ, which aims to distinguish the epistemological and ontological differences between AE while identifying new patterns or potential gaps in the literature. A systematic review was chosen for this study to provide a more comprehensive and reliable synthesis of existing research through integration of findings from individual studies [20] to generate more robust conclusions and new interpretations. Specifically, this research utilizes a meta-synthesis approach that facilitates new theoretical insights and reinterpretation of meanings from other multiple high-quality qualitative studies [21].

To ensure alignment between the literature review and the RQ, the team developed a set of inclusion and exclusion criteria. This included a selection of journals (n=32) related to AE that were grouped into four categories: i) evaluation journals; ii) assessment journals; iii) assessment and evaluation journals; and iv) journals in curricula study. The selection of evaluation journals (n=14) was based on a scoping review conducted by the NIE CAE. According to Loh *et al.* [6], these journals contain a broad coverage of evaluation-related topics and had significant contributions in the field, particularly in publishing a substantial number of research on evaluation (RoE) articles.

As literature suggests distinctions between AE are mostly seen by their measurement and testing processes [4], [7], [22], the team added in educational measurement journals (n=4) into the selection to increase further insight into the ontology of AE. To incorporate a better scope of educational AE, 1-2 journals from each discipline in the Singapore curriculum were also added into the list (n=18). These remaining 18 journals were selected based on their Scopus rankings to identify journals with higher readership and a clear robust peer-reviewed process. The list of journals is available through the digital repository that is published on the NIE digital repository at https://repository.nie.edu.sg/home.

Following the inclusion/exclusion criterion set by the research team, articles sourced from the listed journals were to focus on literature related to AE, specifically within the domain of educational AE. However, exceptions were to be made if AE studies in other domains were deemed relevant to revealing the epistemology and ontology of AE. A 10-year period (2014-2024) was chosen to filter the literature to ensure that findings remain relevant and reflective of the current educational landscape.

Next, a filtered search was conducted using the terms “evaluation”, “definitions”, “epistemologies” and “ontology”, yielding a total of 5,897 articles. To refine the selection, articles were further narrowed down according to the title and abstract level to ensure that they were relevant to the research topic. This process led to the final selection of 226 articles focused on AE. In addition to the selected journals, a list of references was shortlisted by the research team based on their expert knowledge and background of the domains.

After filtering down to 226 articles, including additional sources selected by the research team, a qualitative analysis of the collated data (n=226) was executed to generate themes relevant in distinguishing or finding similar patterns in the AE domain. Of all the articles reviewed, 176 articles were evaluative in nature, while 61 articles were assessment-based. Using the NVivo software, an inductive coding approach was applied, allowing themes to emerge from the broad review of the qualitative data collected. The analysis identified three overarching themes:

- Theme 1: parallels in data collection. Across the reviewed literature, data collection—whether through systematic or informal means—emerges as a foundational component of both AE processes.
- Theme 2: functional use of evidence. A key point of differentiation between AE lies in how the collected information is used, highlighting functional distinctions in intent and application.
- Theme 3: stakeholder orientation. The analysis reveals a recurring emphasis on the intended recipients or stakeholders of the information, underscoring the significance of “for whom” the data is collected in shaping the process.

These identified themes draw parallel to a distinction made by Scanlan [4], who argues that the difference between AE exists because of what is measured, why and how measurements are made for each of the domain. To capture these themes, the team summarized them into four main thematic codes of: i) type and nature of information gathered (what is being collected); ii) methodology (how data is collected); iii) purposes and outcomes (for what purpose and how will evidence be used); and iv) stakeholders (for whom and by whom).

The data was then further coded and deeper analysis was conducted to gain more clarity of the epistemological and ontological aspects of educational AE. It is important to note that the items categorized are not mutually exclusive. That is to say that they may have been coded more than once depending on the context of the article.

2.3. Stage 2–conducting the social lab

The second stage involves the conducting of a social lab which is based on the principle of the Delphi method. The Delphi method is regarded to be “a valuable technique for reaching consensus about

specific issues when empirical evidence is scarce or contentious" [23]. Furthermore, Förster and Gracht [24] express the Delphi method's usefulness in bringing together multiple diverse perspectives before coming to a consensus as it reduces bias. As the data gathered from the meta-synthesis indicates some theoretical distinctions between AE, what is needed to further this discourse is to expand upon an operational definition of the NIE CAE that is grounded in the experiences of people involved in AE. A social lab which is a highly collaborative and participatory method was conducted to allow expert opinions to develop in relation to other perspectives. The participants were posed with the question of what the difference between AE is.

To facilitate the discussion in the social lab, the research team used the findings from the meta-synthesis of the literature review to develop several diagrammatic representations of the relationship between AE. These diagrams were intended to serve as visual cues for discussion rather than as representations of the final results. Across both sessions, a total of 11 participants, ranging from experts in AE, specialists in the Ministry of Education to practitioners such as educational leaders, contributed their responses to the diagrams shown. Given that the focus was on discussing AE in the context of education, participants were brought in through purposive sampling with the intention of finding participants from different perspectives.

Once both sessions of the social lab were transcribed, they were first coded inductively to obtain a master code book based on the codes that emerged, such as: i) information gathered; ii) methodology; iii) purposes and outcomes; and iv) stakeholders. In addition, we found that there were also additional themes such as: v) response to the information gathered; vi) context that AE was carried out; and vii) paradigms in which AE occurs. While codes (i) to (iv) were also found in the meta-synthesis, codes (v) to (vii) were derived from social lab findings that revealed other codes previously unidentified. 'Response to the information gathered' identified responses towards AE that influences how AE is perceived and how the data is used, 'context that AE was carried out' highlighted how AE was perceived differently based on the context that it was conducted in, and 'shifts in the attitude towards AE' revealed how AE concepts have developed over time.

To identify the themes, the codes were reviewed through an inductive-deductive process. While the codes were being understood inductively, this was paired with a consistent returning to NIE CAE's definition to make sense of what was missing. By alternating between the induced codes and the concept maps derived from NIE CAE's definition of AE, this inductive-deductive approach allows for an exploration of themes, allowing for a thorough exploration of the themes available.

3. RESULTS AND DISCUSSION

Thematic analysis of empirical data from social labs revealed three broad themes: ontological differences, epistemological differences, and stakeholders. The data on ontology expands upon the areas of purpose and what evidence is collected to show the concepts involved in AE and the relations between the concepts. The theme of 'epistemology' extends from the meta-synthesis that methodology can provide insights into the epistemology between AE. The theme 'stakeholders', as mentioned earlier in this paper, provides insights into the epistemology and ontology of AE. In fact, the social lab revealed how the understanding of AE shifts not just when different stakeholders are involved, but also from which stakeholder's perspective is AE viewed from. The data that support the findings of this study are openly available in the NIE digital repository at https://doi.org/10.25340/R4/FRDPEM.

The results from the metasynthesis and the social labs corroborate some of our initial discussions about the difference between educational AE. From the metasynthesis, what information is being gathered differs between AE. While evaluation generally focuses on broad, systemic data related to capacity building, strategic planning and program implementations, educational assessment tends to emphasize on individual and group data such as performance measures, abilities and design of assessment itself. Scheerens *et al*. [25] note that evaluation involves the usage of data to make an informed decision, or merit, while assessment collects data to understand an individual's learning at a specific point in time. Findings corroborate with Broom idea [26] that with evaluation, the type of information gathered tending to be macro, broad and encompassing while assessment tends to focus on the individual.

In terms of the methodology used to collect and analyze the data, both domains rely on systematic methods and theoretical frameworks, demonstrating a shared epistemological commitment to evidence-based decision making. Evaluation methodologies are predominantly deductive structured and quantitative, utilizing Delphi and quasi experiments, while assessment relies more inductive, qualitative approaches such as class observations and ethnographic interviews. These differences reveal the epistemological variation between AE, with evaluation grounded in more positivist epistemology and an ontological broader and systemic emphasis compared to assessment drawing on constructivist epistemologies, with ontology centering at the individual level. By utilizing structured mixed methods, evaluation can provide a more comprehensive and overall understanding of the subject being evaluated [27]. In contrast, assessment methods prioritize gathering data on student that usually entails knowledge, skills and performance, tying

heavily to the assessed individual [26]. These differences shed light on the intrinsic link between methodology and epistemology, where knowledge justification (epistemology) influences choice on method justification (methodology) [28]. Nonetheless, exceptions occur where evaluation studies can focus on formative evaluation, that is, increasing looking to qualitative methods, and assessment can be conducted through more quantitative large-scale testing methods. These will be discussed later when looking across the data collected from the social lab.

In addition to the obvious purpose and outcomes of improvement to learning outcomes across the individual, group, program, school, or policy scales, there are also other outcomes that emerged from the literature review. From the literature synthesis, evaluation frequently involved the discussion of new approaches and definitions, as well as the influence of evaluation practice and programs. The emphasis on new approaches and definitions indicates ongoing efforts to standardize or create an operationalized definition in evaluation, which is a common problem identified in the evaluation field [29], [30]. Conversely, assessment trends involve validation of assessment tools, reinforcing the validity of assessment that was also found in its methodologies. They also include factors pertaining to assessment, such as the design, assessment literacy and participation. A unique theme that emerged involved the beliefs and perspectives of students and teachers, highlighting the impact of the assessed individual in shaping and influencing assessment–something less commonly discussed in evaluation. Pertaining to education, these findings reflect how assessment underscores roles in student learning and instructional strategies based on outcomes based on curriculum objectives, whilst evaluation tends to provide actionable insights for policy and program improvements and effectiveness, adhering to outcomes at the programmatic level.

Evaluation's analysis at the systemic level compared to assessment's examination at the individual level is reinforced through the stakeholder dynamics in AE. Evaluation engages a broader range of stakeholders, spanning from those being evaluated to 'evaluation clients' at the organizational and directive level. In contrast, assessment stakeholders are mainly those administering the assessment and those that are being assessed. Assessment directs its value of knowledge and production on the assessed object, whilst evaluation's knowledge come from the evaluated clients rather than the evaluative object. In education, these directly translate to assessment planning constructed around the student's knowledge and ability, while evaluation planning focuses on the knowledge surrounding the school program or policy that is being evaluated.

The relationship between AE reveals its own ontology when focusing on how they interact with each other. One aspect is that its purpose which varies across stakeholders, contexts and scales. Recognizing that there could be a false dichotomy between AE, some of the data highlights that ontological presence of either assessment or evaluation shifts depending on the stakeholders and the scale in which the purpose that they are doing the measurement of. In particular cases, when it is conducted by more general stakeholders such as the institutions, evaluation would be more present. Likewise, when it is conducted by individuals such as students, assessment would be more present. This relationship echoes Broom statement [26] that evaluation is associated with the macro while assessment focuses on the individual. However, it would also depend on who are the evaluative or assessed objects and who are the evaluative or assessed clients which would determine whether it would be more appropriate to use assessment or evaluation.

This conflation between the two is further accentuated when contextualized in the type of information collected. Data from the social lab reveals that when talking about either assessment or evaluation, the other would always be referenced as an integral part of the process. As such, AE are presented as inseparable and intertwined. Assessment data is an integral part of evaluation, and evaluating is a central component of carrying out assessments.

In the relationship between AE, its epistemology makes use of both AE regardless of whether it is solely for the purpose of assessment or evaluation. Some liken the approach between the two to be like a subset where the proportion between AE shifts depending on the context. Evaluation and assessment information are inherently value-laden, and an a priori approach to AE is shaped by the philosophical underpinnings of a particular culture or its culturally specific pedagogies. This differs slightly from the literature that raises contextual knowledge in relation to evaluation and not assessment [6], [31]. However, it is crucial to consider whether the statement that the approach towards AE is shaped by culturally specific pedagogies is made with the suggestion that AE are interchangeable. In addition, while the literature shows that evaluation methodologies are predominantly deductive structured and quantitative while assessment relies on qualitative approaches, data from the social lab reveals a different situation where both AE are increasingly less positivistic and adopt both deductive and inductive approaches.

The theme of 'stakeholders' is intertwined when considering the relationship between AE. The social lab data revealed a trend that when referencing AE from the individual-particular level to the general level, it would also develop from 'assessment' to 'evaluation'. Another area to note is the direction at which AE is conducted. At the individual-student level, teachers could assess students, and nowhere else is 'assessment' referenced. In addition, AE is always conducted 'downwards' from a stakeholder with greater

authority and power, or among stakeholders in the same level such as in self-assessment or self-evaluation. There are no cases found in which assessment or evaluation is conducted 'upwards'. This movement highlights the role relationships play in the ontological and epistemological dimensions of AE.

3.1. Development of the definitions

Scanlan [4] proposes that if there were to be a distinction between AE, it would probably involve 'what is being measured, why and how the measurements are made'. In line with NIE CAE's [19] definitions of AE and supported by the meta-synthesis and social lab, the 'what is being measured' would be (i) information gathered; the 'how the measurements are made' would be (ii) methodology; and the 'why' would be (iii) purposes and outcomes. Since NIE CAE situates its definition of AE in the context of an educational setting, (iv) stakeholders are considered in one of the three evaluation principles on relevance, a point that is raised in both the literature and the social lab. However, while NIE CAE's definitions include the four aspects of the who, what, how, and why, the research findings reveal that the (v) response to the information gathered, (vi) context that AE is carried out, and (vii) paradigms in which AE occurs are not explicitly mentioned. Before proposing how to include these aspects into NIE CAE's definition, their relation to (i) to (iv) must first be considered.

First, (v) to (vii) are not separate elements given that they share a connection with (i) to (iv). As such, they can be considered as overlapping elements that necessarily exist as part the process of conducting AE. Considering (vii) paradigms in which AE occurs, paradigms shape the epistemology in which one approaches (i) to (iv). If a person has been trained in the paradigms of positivism or behaviorism, it would shape the way they collect the data, what data would be collected, or how the data is going to be used, for instance. Chen [32] proposal to interface the three theoretical perspectives of reductionism, systems thinking, and pragmatic synthesis to further evaluation practice describes a configuration of stakeholder's assumptions on how to conduct evaluation and why it works. This then maps onto stakeholder's understandings on what is expected to happen to attain program goals and what should be done to attain them. As such, the need to consider an evaluation's paradigm has become one of the key factors in choosing an appropriate method [33].

While the focus on paradigms is more commonly found in discussions around evaluation, considering the paradigms in which assessment occurs is still necessary. The understanding of 'learning' is framed by one's epistemology and influenced by theories of learning and assessment theory [34]. Therefore, the purpose of AE, choice of what information is gathered in AE data, or its methodology is not just dependent on the individual or organization that is conducting AE, but it is ultimately framed by the paradigms they are influenced by, or the paradigm that is at the background of conducting AE at that moment.

In addition to the paradigm in which AE operates in, the (vi) context that AE is carried out would influence how AE operates. Referencing Eisner [35] levels of curriculum decision-making, decision-making in curriculum is based on the level at which it operates. Extrapolating from this, the level at which AE operates would affect how AE is conducted. In evaluation, the context such as the policy or decision setting shapes the way stakeholders are involved and how the findings are disseminated, which ultimately shapes the production of evaluation knowledge [36].

In assessment, Bennett [37] proposes that teaching and assessment should be structured according to the epistemologies of the students they are directed towards. Considering how standardized tests discriminate against students of a lower socioeconomic status given that it is measured to a particular socioeconomic competency, the consideration of the context in which assessment occurs at various levels of students would aid in reducing structural inequalities [37]. These findings in the literature align with participants of the social lab emphasizing that AE would appear and operate differently at different levels, both locally and internationally, which contributed to the difficulty of providing a single definition of AE that considered every context.

Finally, the (v) response to the information gathered, especially by the individuals conducting AE, reveals another dimension to the why and how of AE. Marsh *et al*. [38] highlight that in the context of schools' responses to teacher evaluation policy, schools may respond in reflective, distortive, or compliant ways depending on the context or stakeholders who are responding. In assessment, student responses to assessment could be used by teachers to refine and design assessment tasks so to include students' interests and experiences to better engage them [39]. Data from the social lab adds the emotional dimension to highlight how the use of AE data is shaped by how one responds to AE, which in turn could also be used as an additional form of data. Therefore, the response to the information gathered can be reframed as understanding what the data gathered is to be used for, as that affects the way the methodology and the purpose in conducting AE.

As a result, given that the distinctions between AE are defined by the who, what, why, and how, these findings from the meta-synthesis and the social lab support the principles of AE at NIE CAE [19]. What can be refined in the center's definition is 'understanding what the data is used for', which would be incorporated into the why and the how of AE. The incorporation of this element would support the purpose

of conducting AE which requires a consideration of the suitable methodology in achieve the purpose. However, the elements of who, what, why, and how of the definition are insufficient in coming to a cohesive definition of AE, even if it is supported with an understanding of what the data is used for. From the meta-synthesis and social lab, we can conclude that AE would be conceived differently depending on one's training and paradigmatic influence that they operate within. As such, we cannot include the other two elements of the definition such as the (vi) context that AE is carried out and (vii) paradigms in which AE occurs. Instead, these points can be inserted as a pre-amble to NIE CAE's definition as factors that influence the way in which AE operates.

3.2. The disambiguation between AE

Based on the findings thus far, AE can be considered distinct fields, but some misconceptions remain. However, while recognizing that the two fields are interconnected, they are not interchangeable. The (vi) context that AE is carried out and (vii) paradigms in which AE occurs sheds light on how misconceptions have come about. In speaking about AE, the lack of a common context that frames the discussion results in multiple understandings of AE being discussed at the same time. Similarly, acknowledging that one's training under a particular paradigm shapes the way they understand AE, there are many impressions of AE that may be discussed under the shared terminology of 'assessment' and 'evaluation'.

These misconceptions cannot be eliminated, but they can be reduced. The findings from the meta-synthesis and the social lab reveals how the interplay between ontology and epistemology shifts and changes depending on the context and the paradigmatic reference. Furthermore, as previously mentioned, aspects such as 'stakeholders', or even AE itself are so intertwined that misconceptions will undoubtedly arise. While the ontology and epistemology may be the distinctive characteristics of AE, an awareness that there are varieties of AE, and that different individuals may have different epistemologies and ontologies of AE would aid in reducing the misconceptions between AE when they are discussed.

4. CONCLUSION

To conclude, we return to the issue that this paper began with: identifying and clarifying the epistemological and ontological differences between AE through: i) information gathered; ii) methodology; iii) purposes and outcomes; and iv) stakeholders. While NIE CAE's definitions reflect the 'what', 'why', 'who', and 'how' of the operation of AE, the findings from the meta-synthesis and social lab can provide an understanding of what the data is used for to support the aforementioned elements of AE. Furthermore, the addition of a pre-amble to the definition to consider the context and paradigm in which AE occurs would aid in sharpening how AE is understood in its particular situation, which would be effective given that AE is understood differently at different levels and across different stakeholders. Allowing for a pre-amble rather than a fixed contextual or paradigmatic basis to be part of the definition allows the definition to be appropriately contextualized for the specific area or group of study.

For the future, the refined definition of AE would aid in research by considering the various factors that shape the way in which AE is understood. Not only would the understanding of AE be defined to fit the purpose of the research, but an awareness of the other factors that shape the ontological and epistemological foundations of researchers and participants would help to guide the way in which research is conducted. Furthermore, a definition that is contextually fitted would provide a culturally relevant or localized approach that avoids misinterpretations using a preconceived cultural understanding of AE. At the same time, future research on the definitions of AE could build upon the findings of this paper and broaden the scope of inquiry. While our social lab involved educators, scholars, and experts in the field of AE, including other stakeholders such as students would deepen an understanding of how AE is understood at various levels of stakeholders and contexts.

In the research of sharpening NIE CAE's definitions of AE, several methodological and theoretical challenges emerged. There were difficulties in understanding the relationship between AE because of the overlap that exists between the two. This overlap was dependent on the scale, context, and the epistemological foundations of the individual or practitioner. As such, the definition of AE is not something that can be prescribed but instead needs to be descriptive because some ways of thinking are more a priori in nature and others are a posteriori. This proposal of the definition does not encourage people to subscribe to either epistemology, but this research acknowledges that these are the differences, and it is important to have awareness of these differences while working in the field of education, instead of being bothered about definitions. The frames of i) information gathered; ii) methodology; iii) purposes and outcomes; and iv) stakeholders will continue to help us in identifying these differences.

With the world facing numerous challenges, including anti-globalization and climate change, education is under immense pressure to prepare children for times of uncertainty. AE remain as core aspects

of what we do in education. While we continue to collect data for the purposes of refining programs, improving student learning experiences, and enhancing teacher capacity, many of us are not fettered by the need to have a clear disambiguation between the definitions of AE.

FUNDING INFORMATION

This work would not have been possible without the funding from the Strategic Research Funds for the project SRF 01/24 CCH.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Chew-Hung Chang	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Christina Ratnam-Lim		✓			✓	✓		✓		✓	✓			✓
Wong Hwei Ming		✓			✓	✓		✓		✓	✓			✓
Phoebe Ow		✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓
Joshua Wong Jia Hao			✓	✓	✓	✓		✓	✓	✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study in accordance with the protocol as approved by the NTU Institutional Review Board (Ref- IRB-2024-731).

ETHICAL APPROVAL

The study has undergone ethics approval by the Nanyang Technological University (NTU) Institutional Review Board (Ref- IRB-2024-731).

DATA AVAILABILITY

The data that support the findings of this study are openly available in the NIE digital repository at https://doi.org/10.25340/R4/FRDPEM.

REFERENCES

[1] J. A. King and S. Ayoo, "What do we know about evaluator education? A review of peer-reviewed publications (1978–2018)," *Evaluation and Program Planning*, vol. 79, p. 101785, Apr. 2020, doi: 10.1016/j.evalprogplan.2020.101785.
[2] J. Kwako, "A brief summary of traditional and alternative assessment in the college classroom," University of Wisconsin-Madison, 2008.
[3] P. Santiago, "The focus on evaluation and assessment," in *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD, Ed., Paris: OECD Publishing, 2013, pp. 29–34, doi: 10.1787/9789264190658-4-en.
[4] C. Scanlan, "Assessment, evaluation, testing and grading," 2012. Accessed Dec. 01, 2024. [Online]. Available: https://www.mariancollege.org/miitle/assets/downloads/mitle/resources/Assessment&Evaluation.pdf
[5] H. M. Wong, D. Kwek, and K. Tan, "Changing assessments and the examination culture in Singapore: a review and analysis of Singapore's assessment policies," *Asia Pacific Journal of Education*, vol. 40, no. 4, pp. 433–457, 2020, doi: 10.1080/02188791.2020.1838886.
[6] J. Loh, W. H. Chong, and C. Lim-Ratnam, "Educational evaluation research in Asia: a scoping review," *Asia Pacific Journal of Education*, vol. 44, no. 3, pp. 552–568, 2024, doi: 10.1080/02188791.2024.2390663.

[7] B. Kizlik, *Measurement, assessment, and evaluation in education*. Boca Raton, FL: Robert Kizlik & Associates, 2012.
[8] C. H. Chang and P. Ow, "Inquiry-based fieldwork assessment for and as learning in geography," in *Assessment in Geographical Education: An International Perspective*, T. Bourke, R. Mills, and R. Lane, Eds., Cham: Springer International Publishing, 2022, pp. 123–134, doi: 10.1007/978-3-030-95139-9_6.
[9] N. M. Lambert and B. L. McCombs, "Introduction: Learner-centered schools and classrooms as a direction for school reform," in *How students learn: Reforming schools through learner-centered education*, N. M. Lambert and L. McCombs, Eds., Washington, D.C.: American Psychological Association, 1998, pp. 1–22, doi: 10.1037/10258-017.
[10] A. C. Ornstein, D. U. Levine, G. L. Gutek, and D. E. Vocke, *Foundations of education*, 12th ed. Belmont, CA: Wadsworth Cengage Learning, 2014.
[11] T. S. Kuhn, *The nature of scientific revolutions*. Chicago, IL: University of Chicago, 1970.
[12] M. Baehr, "Distinctions between assessment and evaluation," in *Program Assessment Handbook*, Lisle, IL: Pacific Crest, 2005, pp. 441–444.
[13] G. Kidman and C. H. Chang, "Assessment and evaluation in geographical and environmental education," *International Research in Geographical and Environmental Education*, vol. 31, no. 3, pp. 169–171, 2022, doi: 10.1080/10382046.2022.2105499.
[14] B. K. Lynch, "Rethinking assessment from a critical perspective," *Language Testing*, vol. 18, no. 4, pp. 351–372, 2001, doi: 10.1177/026553220101800403.
[15] W. Popham, *Transformative assessment*. Alexandria, VA: ASCD, 2008.
[16] R. C. Richey, D. C. Fields, and M. Foxon, *Instructional design competencies: the standards, Syracuse*. New York: ERIC Clearinghouse on Information & Technology, 2001.
[17] M. T. Chi, *Three types of conceptual change: belief revision, mental model transformation, and categorical shift*. New York: Routledge, 2008.
[18] G. Liarakou, I. Athanasiadis, and C. Gavrilakis, "What Greek secondary school students believe about climate change?" *International Journal of Environmental and Science Education*, vol. 6, no. 1, pp. 79–98, 2011.
[19] National Institute of Education, Centre for Assessment & Evaluation, "Centre for assessment & evaluation - About Us," ntu.edu.sg. Accessed Dec. 01, 2024. [Online]. Available: https://www.ntu.edu.sg/nie/research-labs/centre-for-assessment-evaluation/about-us
[20] M. Newman and D. Gough, "Systematic reviews in educational research: methodology, perspectives and application," in *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*, O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, and K. Buntins, Eds., Wiesbaden: Springer Nature, 2020, pp. 3–22, doi: 10.1007/978-3-658-27602-7.
[21] C. Hoon, "Meta-synthesis of qualitative case studies," *Organizational Research Methods*, vol. 16, no. 4, pp. 522–556, 2013, doi: 10.1177/1094428113484969.
[22] M. McAlpine, *Principles of assessment*, 1st ed. Luton: CAA Centre, University of Luton, 2002.
[23] M. Barrios, G. Guilera, L. Nuño, and J. Gómez-Benito, "Consensus in the Delphi method: what makes a decision change?" *Technological Forecasting and Social Change*, vol. 163, p. 120484, 2021, doi: 10.1016/j.techfore.2020.120484.
[24] B. Förster and H. von der Gracht, "Assessing Delphi panel composition for strategic foresight - A comparison of panels based on company-internal and external participants," *Technological Forecasting and Social Change*, vol. 84, pp. 215–229, 2014, doi: 10.1016/j.techfore.2013.07.012.
[25] J. Scheerens, C. Glas, and S. M. Thomas, "Basics of educational evaluation," in *Educational Evaluation, Assessment and Monitoring: A Systematic Approach*, 1st ed., C. Glas, J. Scheerens, and S. M. Thomas, Eds., London: Taylor & Francis, 2006, pp. 33–48.
[26] C. Broom, "Assessment and evaluation: exploring their principles and purposes in relation to neoliberalism through a social studies case study," *Canadian Social Studies*, vol. 45, no. 2, pp. 17–36, Oct. 2012, doi: 10.29173/css238.
[27] S. SenGupta, R. Hopson, and M. Thompson-Robinson, "Cultural competence in evaluation: an overview," *New Directions for Evaluation*, vol. 2004, no. 102, pp. 5–19, Jun. 2004, doi: 10.1002/ev.112.
[28] S. M. Carter and M. Little, "Justifying knowledge, justifying method, taking action: epistemologies, methodologies, and methods in qualitative research," *Qualitative Health Research*, vol. 17, no. 10, pp. 1316–1328, 2007.
[29] C. Poth, M. K. Lamarche, A. Yapp, E. Sulla, and C. Chisamore, "Toward a definition of evaluation within the Canadian context: Who knew this would be so difficult?" *Canadian Journal of Program Evaluation*, vol. 29, no. 1, pp. 87–103, 2014, doi: 10.3138/cjpe.29.1.87.
[30] A. M. Gullickson, "The whole elephant: defining evaluation," *Evaluation and Program Planning*, vol. 79, p. 101787, 2020, doi: 10.1016/j.evalprogplan.2020.101787.
[31] L. Omosa, T. Archibald, K. Niewolny, M. Stephenson, and J. Anderson, "Towards defining and advancing 'made in Africa evaluation'," *African Evaluation Journal*, vol. 9, no. 1, p. a564, 2021, doi: 10.4102/AEJ.V9I1.564.
[32] H. T. Chen, "Interfacing theories of program with theories of evaluation for advancing evaluation practice: reductionism, systems thinking, and pragmatic synthesis," *Evaluation and Program Planning*, vol. 59, pp. 109–118, 2016, doi: 10.1016/j.evalprogplan.2016.05.012.
[33] S. Elahi, N. Kalantari, M. Hassanzadeh, and A. Azar, "A new perspective for mixed-methods evaluations," *Evaluation Journal of Australasia*, vol. 15, no. 3, pp. 15–29, 2015, doi: 10.1177/1035719X1501500303.
[34] J. A. Baird, D. Andrich, T. N. Hopfenbeck, and G. Stobart, "Assessment and learning: fields apart?" *Assessment in Education: Principles, Policy and Practice*, vol. 24, no. 3, pp. 317–350, 2017, doi: 10.1080/0969594X.2017.1319337.
[35] E. W. Eisner, *Cognition and curriculum reconsidered*, 2nd ed. New York: Teachers College Press, 1994.
[36] J. Maloney, "Evaluation: what's the use?" *Evaluation Journal of Australasia*, vol. 17, no. 4, pp. 25–38, 2017, doi: 10.1177/1035719X1701700404.
[37] R. E. Bennett, "Toward a theory of socioculturally responsive assessment," *Educational Assessment*, vol. 28, no. 2, pp. 83–104, 2023, doi: 10.1080/10627197.2023.2202312.
[38] J. A. Marsh, S. Bush-Mecenas, K. O. Strunk, J. A. Lincove, and A. Huguet, "Evaluating teachers in the Big Easy: how organizational context shapes policy responses in New Orleans," *Educational Evaluation and Policy Analysis*, vol. 39, no. 4, pp. 539–570, 2017, doi: 10.3102/0162373717698221.
[39] H. Kang and E. M. Furtak, "Learning theory, classroom assessment, and equity," *Educational Measurement: Issues and Practice*, vol. 40, no. 3, pp. 73–82, 2021, doi: 10.1111/emip.12423.

BIOGRAPHIES OF AUTHORS

Chew-Hung Chang works in sustainability, geography, and environmental education. His expertise is in climate change misconception diagnostics and in improvement to geography curriculum and technology-enabled refutation-oriented instruction. He is Professor of Education and the Chair of the Assessment and Evaluation Strategic Growth Area at National Institute of Education, Nanyang Technological University, Singapore. He is also currently Dean of Graduate Studies and Professional Learning and NIE Sustainability Officer. He is an elected Senator, who previously served as the Chairperson to the Senate Committee on Education (2020–2022) and Co-Chair to the Senate Steering Committee (2021–2022). He is also an Adjunct Professor of East China Normal University, Shanghai, China and he was previously a visiting scholar to the University College London, Institute of Education. He can be contacted at email: chewhung.chang@nie.edu.sg.

Christina Ratnam-Lim is a senior lecturer at the Policy, Curriculum and Leadership Academic Group, focusing on crafting the curriculum, curriculum implementation and evaluation, perspectives of learning, and teacher professional development and learning. She enjoys facilitating deep conversations on curriculum and pedagogical leadership, perspectives of learning and mindsets, and engaging in Q methodology in research. She can be contacted at email: christina.ratnam@nie.edu.sg.

Wong Hwei Ming is a senior education research scientist and the assistant center director at the Centre for Research in Pedagogy and Practice (CRPP), National Institute of Education (NIE). Prior to joining NIE, she was a primary school teacher and a specialist in MOE HQ for several years. She has more than 18 years of experience in education research, mainly in the area of assessment practices and assessment literacy and she teaches assessment courses and workshops. Upon requests, she has customized professional development workshops on student-involved assessment for primary and secondary schools for their staff development. Dr. Wong is also a registered educational psychologist with the Singapore Registry of Psychologists and a member of the Singapore Psychological Society since 2003. She can be contacted at email: hweiming.wong@nie.edu.sg.

Phoebe Ow is a senior research analyst at the Centre for Governance & Sustainability, National University Singapore and is currently pursuing a MSc in Sustainable Development at the SOAS University of London. Phoebe was previously working within the NTU, National Institute of Education's Centre for Assessment and Evaluation and has published in the domain of assessment in geographical and environmental education. She can be contacted at email: phoebe.ow13@nus.edu.sg.

Joshua Wong Jia Hao is a research analyst in the Studies in Inter-Religious Relations in Plural Societies (SRP) Programme at the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU). He was previously a research assistant at the NTU Wee Kim Wee School of Communication and Information and the NTU, National Institute of Education's Centre for Assessment and Evaluation. He can be contacted at email: joshua.wongjh@ntu.edu.sg.