International Journal of Evaluation and Research in Education (IJERE)
Vol. 14, No. 4, August 2025, pp. 2539~2554
ISSN: 2252-8822, DOI: 10.11591/ijere.v14i4.30399

GE-APROAch: an OBEdized survey instrument on GenEd outcomes achievement based on student personal reports

Rodrigo C. Rivera, Elymae Naldo Delos Santos, Juan Oliver S. Ofracio
School of Multidisciplinary Studies, Communication and Literature Area, De La Salle-College of Saint Benilde, Manila, Philippines

Article Info

Article history:

Received Feb 25, 2024
Revised Dec 12, 2024
Accepted Mar 2, 2025

Keywords:

General education
Higher education evaluation
Learning outcomes assessment
Outcomes-based education
Validation study

ABSTRACT

The Philippine commission on higher education (CHED) requires colleges and universities to align with its triad goals of individual, social, and global development, through the new general education curriculum (NGEC). Since its implementation, numerous studies have evaluated outcomes, often reporting positive results. However, these studies exhibit methodological inconsistencies due to varying assessment measures. While two studies from the same college employed robust self-developed instruments based on CHED standards to evaluate general education (GE) outcomes, questions about their reliability led to the creation of the GE-assessment of personal reports of outcomes achievement (GE-APROAch) instrument. Using a mixed-method descriptive-evaluation design, the research constructed and validated the instrument. Results indicated that the GE-APROAch effectively measures GE outcomes through: i) a solid conceptual framework of targeted competencies; ii) high interrater agreement from expert judges; iii) strong internal reliability to differentiate and consistently report scores; and iv) superior efficacy compared to existing tools. Establishing the validity of self-made instruments is crucial, particularly for evaluating curriculum learning outcomes. Statistical analysis confirms that the GE-APROAch is a reliable and valid tool, offering improvements over previously used instruments.

Corresponding Author:

Rodrigo C. Rivera
School of Multidisciplinary Studies, Communication and Literature Area
De La Salle-College of Saint Benilde
1004 Metro Manila, Manila, Philippines
Email: rodrigo.rivera@benilde.edu.ph

1. INTRODUCTION

As per the commission on higher education (CHED) memorandum order (CMO) 20, s. 2013, undergraduate students across all higher education institutions (HEIs) in the Philippines must complete twenty-four course units of general education (GE) courses and nine GE course electives. Formative evaluation of the extent to which the new general education curriculum (NGEC) has achieved its desired outcomes is paramount as HEIs adopt the outcomes-based education (OBE) framework. In OBE, educational systems are organized "around what is essential for all learners to know" because of their potential to "increase both the effectiveness of quality assurance (QA) and the effectiveness of higher education" [1]. Designing effective instructions requires schools to adopt diverse evaluation measures to decide how learners have reached the expected outcomes [2].

Through CMO 46, s. 2012, HEIs are mandated to bring out graduates who are intellectually competent, highly professional, with a focus on development, and technologically able, with a sense of social responsibility to improve the quality of human life [3]. Furthermore, the CHED recognizes the functional

differentiation of HEIs that it adopts a horizontal typology-based QA while expecting that colleges and universities commit to its mission through vertical typology-based QA, related to the level of program excellence and institutional quality [3]. This underscores the consistency of the learning environment with the institution's vision and goals, the demonstration of exceptional learning and service outcomes, and the development of a culture of quality.

OBE is student-centered which should lead to life-long learning, based on a purpose that is clearly stated; it expands opportunities, is ideal and designed for students, and is framed around the standards of acceptable performance that increases access to higher-level curriculum [4]. Despite its soundness and logic, critics raise issues against OBE. Its implementation challenges school leaders and researchers to examine its extent of effectiveness in meeting desirable outcomes, at program and school-wide levels. Moreover, OBE is criticized for lacking evidence that supports outcomes achievement as implemented, and often institutional practices do not adhere to the standards set by regulators [5]. The Organization for Economic Co-operation and Development (OECD) also reported a considerable gap in knowledge on outcomes achievement in HEIs, at an international scale [6].

In narrowing such gap, this study interrogates: how does the GE-assessment of personal reports of outcomes achievement (GE-APROAch) fair as an instrument to measure achievement of GE learning outcomes? Specifically, this validation study aims to determine whether the proposed instrument: i) Can discriminate between scores across diverse groups or classes; ii) Is reliable to measure shared perception of students from the same population or classes; and iii) Is organized to elucidate on the process of developing and validating a formative evaluation measure, leading to a reflexive evaluation on the proposed instrument's potential and limitations by testing the following null hypotheses:

- $H0_1$: group scores on reported outcomes achievement do not differ to the population mean ($\mu_d=0$), to determine the instruments' discriminating power in measuring scores across diverse groups.
- $H0_2$: using the GE-APROAch instrument, scores between the split-group taking the same course do not differ ($\mu_1=\mu_2$), to suggest its sensitivity in measuring shared perception of population in the same learning environment.
- $H0_3$: the GE-APROAch instrument is not significantly different in variance of scores ($T_1=T_2=T_3$), when compared with two other similar instruments that measure GE outcomes achievement following the CHED mandate.

The successful implementation and evaluation of the GE curriculum within the OBE framework in Philippine HEIs depend on reliable assessment tools like the GE-APROAch. HEIs must produce graduates who are academically skilled and socially responsible, necessitating that assessment instruments align with OBE goals and CHED standards. While OBE offers potential for improved learning outcomes and institutional quality, validating its effectiveness presents challenges due to criticisms and evidence gaps. This study rigorously evaluates the GE-APROAch to assess its reliability and validity to differentiate scores, measure shared perceptions, and compare with other instruments, providing insights to enhance educational outcomes and QA.

2. LITERATURE REVIEW

2.1. Instructional planning, outcomes, and quality education

Teachers must approach the instructional planning process deliberately and focus on the ideal results, with determined acceptable evidence, which becomes the basis of the planned learning experience [7]. The need for assessment tools, which are unbiased, valid, dependable, and adherent to standards, is vital in measuring learning outcomes [8]. Regionalization for HEIs is described as the foreground to internationalization, synonymous with competitiveness by adhering to global standards, requiring collaboration and programmatic to enhance the curriculum, to which Philippine HEIs invest in international accreditation such as with that of the ASEAN university network [9] which like any other accreditor and government regulator emphasize achievement of learning outcomes across the curriculum.

In the QA cycle, HEIs must evaluate the extent to which expected outcomes are met, striving for excellence and adherence to standards [10]. Accrediting and regulating bodies typically place this responsibility on academic institutions. Currently, there is no unified assessment tool for GE outcomes in the country, except for two recent studies [11], [12]. These studies were re-examined in this validation research and compared with the proposed GE-APROAch.

Nusche [13] highlights that many countries use direct assessments of learning, while HEIs in the USA and Australia sometimes use indirect methods, such as surveys of students' self-reported achievements. Both methods measure generic and domain-specific competencies but have limitations. Direct assessments provide concrete evaluations of learning outcomes, whereas indirect methods, though less precise, offer valuable insights into self-reported learning evidence. Previous studies [11], [12] concur on the need to

formatively evaluate GE programs, given that it has been implemented for several years already. Large-scale direct assessments of learning have relative distance from what is taught, while indirect measures do not examine the actualities of learning but offer some insights on evidence of learning as self-reported [14]. HEIs need to carefully choose the assessment approach that best fits their measurement goals and context, balancing the strengths and limitations of both direct and indirect methods. Hence, HEIs in the country must adopt what is fit for assessing GE outcomes achievement.

2.2. Divergent measures in evaluating general education outcomes achievement

In both practice and policies, HEIs in the Philippines continue to employ an inputs-based approach rather than fully embracing the OBE framework. This traditional approach emphasizes content coverage, segmented learning time, periodic testing, and quantifying performance instead of qualifying standards, focusing on teaching to test and evaluating students based on test results [15]. This divergence from OBE practices raises questions on the effectiveness and implications of educational methods.

A recent survey at a university in Southern Metro Manila, used a self-made instrument claimed to align with the CHED standards. The findings reported that the programs were successful in imparting relevant knowledge, critical thinking, communication skills, and ethical perspectives [16]. However, this survey did not fully reflect the prescribed CHED GE outcomes. Similarly, in 2016, the university of the Philippines utilized its student evaluation of teaching to gauge student motivation and satisfaction, reporting average to very good ratings on course satisfaction and outcomes achievement. Despite this, the evaluation did not specify which GE outcomes were met [17].

Extensive cross-sectional studies to review OBE implications have been conducted, showing positive results after years of implementation [18], [19]. Local studies have focused on compliance with CHED standards, including institutional implementation [18], [20], syllabus alignment [14], [21], courseware relevance [22], and acceptance of the OBE framework [23], [24], reporting positive results. While these studies reported positive outcomes, many lacked substantial evidence of achievement and often used self-developed instruments with lower standards of measurement. The NGEC aims for holistic learner development, encompassing knowledge of self, Filipino society, and the world [25], setting higher standards of education.

Given these gaps, existing instruments need reworking, or a new instrument must be developed. The GE-APROAch, which aligns with the CHED prescribed outcomes also differs from previous instruments by using clear, student-tailored items that ensure comprehensibility and relevance, facilitating better engagement and response. This descriptive evaluation aims to compare GE-APROAch with its predecessors and assess its potential to effectively measure learning outcomes as prescribed by the NGEC. Validity remains a critical consideration in developing any effective outcome measurement tool.

2.3. Instrument development and validity

The design of an instrument and the process of its evaluation through expert judgment, must adhere to the guidelines and principles on content validity and educational assessment [26]. Expert evaluators are chosen based on their experience and knowledge, objectivity and common sense, communication competence, academic and scientific reputation, and their motivation to collaborate [27]. Most often instruments were face-content validated using binary codes which is questionable and unreliably resulting to type 1 errors [28]. Attempts to measure GE outcomes achievement have relied on self-made instruments, focused on compliance, that brings to question their validity and reliability, and even widens the gap in outcomes assessment.

Validity attributes to the consistency of an instrument's measure, ascertained through different techniques and analysis which should determine its internal consistency across all items, stability that consistently results the same way upon repeat testing, and equivalence that there is consistency in the responses of multiple users or among alternate forms of instrument [29]. Instrument validity and reliability are assessed in terms of relevance, content, internal structure, response process and consequences [30]. Face-content validation through expert judgment may address such concerns if it added layers of evaluation in the process [31]. It should be tested as to utility and appropriateness to meet purpose, supported by multiple layers of evidence demonstrating attributes of careful test construction and social consideration [32].

Assessment and learning are bound together, and program level assessment rely on valid and reliable instruments that institutions can widely adopt [33]. Self-reports are just one of them. There are many ways to measure learning outcomes achievement, and at the global level, regions and country adopt diverse means of assessing learning outcomes [34]. However, institutions in the country vary in their approaches in assessing GE-outcomes achievement [14], [16]–[19], [21], [22], [24], despite the fact that there is a set of outcomes that CHED mandates for GE providers to meet. Deviating from these prescribed outcomes makes any assessment attempts tangentially off.

Other studies [11], [12] on GE-outcomes adopted the CHED prescribed outcomes in earlier attempts to determine the extent to which they are met in specific GE courses. The department of education in the Philippines, through The Second Congressional Commission on Education (EDCOM) underscored limited opportunities for large-scale assessment that should provide basis for reforms [35]. Since 2018, no large-scale assessment on the implemented NGEC has gone underway. Without a standardized measure such intention will not be fruitful. GE-APROAch could narrow that gap, which results could inform policy formulations, curriculum development and innovations in instructional delivery.

2.4. CHED prescribed general education outcomes and key competencies

CMO 20, s. 2013 builds on CMO 46, s. 2012 by integrating a competency-based approach and adopting an OBE framework for QA in Philippine higher education. The NGEC is designed to expose students in "various domains of knowledge and ways of comprehending social and natural realities" that develops their intellectual competencies, civic capacities and practical skills [25]. The NGEC policy specifies nineteen learning outcomes expected from graduates but does not detail the competencies associated with these outcomes. Towards learner's holistic development, GE should develop students as an individual, as a Filipino citizen and as a member of the global community with, intellectual competence, socio-civic responsibilities and practical skills [25].

Moreover, the CHED expects GE courses to foster students' personal identity, pride in their collective identity, respect for diversity, and awareness of global issues [25] which overarch the three key GE outcomes of intellectual competencies, personal and civic responsibilities, and practical skills with nineteen learning outcomes. The commission defines two main competencies: intellectual competence, which involves critical and creative thinking, and civic capacities, which relate to participation in community and global activities [25]. Researchers designing instruments must understand these concepts thoroughly.

Competencies include observable knowledge, skills, and attitudes [36]. Intellectual competence (LO1-LO5) involves cognitive abilities and personality traits that can predict future success, differing from IQ and assessable through self-evaluation [37]. Civic capacity (LO6-LO14) includes the beliefs, knowledge and skills needed to influence social agendas and contribute to societal development [38]. Practical skills (LO15-LO19). Although CHED does not define practical skills (LO15-LO19), Programme for International Student Assessment (PISA)-OECD describes global competencies as including cultural awareness, digital fluency, and global awareness, respect and appreciation of diversity, digital fluency, and global awareness that support sustainable development goals [34].

The shift to OBE emphasizes a student-centered approach, focusing on what students need to achieve by the end of their courses. Debates exist about the effectiveness of outcomes-based versus competency-based education, with some arguing that competency-based education might limit intellectual development. However, others find these approaches complementary and effective [39]. Conversely, outcomes-oriented competency-based education were found to be complementary and effective in higher education [40]. "Competencies mean more than learning outcomes: they can be measured, verified by evidence of student achievement, and applied in different situations" [41].

Competencies should be defined in relation to learning outcomes and measured through evidence of student achievement [36]. The GE-APROAch instrument was developed with comprehensive input from the literatures, reflecting essential such as creativity, global awareness, and civic literacy. The former CHED chairperson identified vital competencies in developing human and intellectual capital, that resonates to "ingenuity, agility, skills lead, competitiveness, civic literacy, global awareness, cross-cultural skills, critical and inventive thinking, communication, collaboration and information skills" [42]. The introduction of the GE-APROAch instrument underscores the importance of clearly defined competencies and measurable outcomes, aligning with the broader NGEC educational goals. The literature is rich to inform on the competencies linked to GE learning outcomes, a prior review of which compared with the CHED stated outcomes supported the construction of the items in the GE-APROAch instrument [11], [12].

3. METHOD

This mixed-method descriptive-evaluation study [43] aims to report, review, and reflect on the process, impact, and implementation of the GE-APROAch survey instrument, designed to measure personal reports of student outcomes achievement. The instrument was developed and validated using an exploratory sequential mixed-method research design, integrating both qualitative and quantitative data and analysis tools [44]. It supports assessment procedures that employ a forward logic model to monitor and evaluate existing educational programs, detailing intended components [45]–[47]. Using a mixed-methods design, this study offers deeper insights into the quality of the instrument and strengthens the evidence for its validity [48].

The need for a standardized, valid, and reliable instrument to measure GE outcomes achievement drove this research. The development of the GE-APROAch instrument was guided by CHED standards and informed by conceptual definitions from relevant literature. Four expert judges, comprising two external and two internal evaluators, were selected to assess the instrument's face and content validity [31]. These experts were chosen based on several reasonable criteria for inclusion [27]: advanced degrees (including master's and doctoral qualifications), at least three years of college teaching experience, a balanced representation from both within and outside the college, proficiency in evaluating written language, and a confirmed commitment to rigorously review the instrument.

The validation process focused on several key aspects of the instrument: item construction, including validity, reliability, and clarity. This was achieved through the use of a 4-point Likert scale for evaluating how well the items measured the 19 GE outcomes and six open-ended questions. The overall quality of the instrument was assessed using a 5-point Likert scale across ten criteria to avoid potential type I errors commonly associated with binary measures [28]. Construct validity was examined using the scale content-validity index [30], which guided revisions of the items. Internal reliability was measured using Cronbach's alpha, while interrater agreement was assessed with Fleiss Kappa, addressing the limitations of Cronbach's alpha in measuring agreement between raters.

Each item in the GE-APROAch was scrutinized for reliability, validity, and clarity. Reliability refers to the consistency of the item in measuring the identified GE learning outcomes, validity pertains to how well the item reflects those outcomes as experienced by students, and clarity involves ensuring that the language used is understandable to the general student population. The instrument underwent further pilot testing with 84 students from the department of communication and literature at De La Salle-College of Saint Benilde, Manila during AY 2022-2023, with a 98% participation rate. Statistical hypotheses were tested at a 0.05 significance level using ANOVA, T-tests, and Tukey honestly significant difference (HSD).

Ethical considerations were strictly adhered to throughout the study, in compliance with the data privacy act (RA 10173). Informed consents were obtained both from the experts and the pilot-test participants. Anonymity and confidentiality were strictly observed, as well as the safety of students and evaluators were guaranteed. The GE-APROAch was compared with instruments used in previous studies [11], [12], evaluating similarities and differences in item construction and conceptual foundations, referred to as instruments 2 and 3 (T2 and T3). In editing and proofreading this report, generative pre-trained transformer (GPT) AI tools (i.e. Microsoft Word Editor and ChatGPT4) were used to make the report narrative concise and ensure that the document is free of grammatical, spelling, and punctuation errors.

4. RESULTS

4.1. Qualitative evaluation of GE-APROAch compared with two other instruments

A standard, valid, and reliable instrument to measure outcomes achievement in the GE curriculum will not only benefit program enhancement or instructional development, but it is vital for QA, vertically and horizontally that should inform teachers, education administrators, regulators, and accreditors. Table 1 shows that the GE-APROAch has strong content validity, meeting criteria described by Fitzpatrick [49] including sampling adequacy, relevance, and clarity. Unlike instrument 2, which used fewer qualitative questions and instrument 3 that lacked extensive pre-testing, GE-APROAch's comprehensive development and high reliability provide a deeper understanding of student learning experiences. All three instruments use a 4-point Likert scale but differ in their score interpretations, with GE-APROAch maintaining a strong alignment with CHED outcomes through appropriate and effective language and scoring methods.

Data shows that the GE-APROAch (instrument 1) demonstrates robust content validity among the three tools, adhering to the CHED CMO 20 s. 2013 guidelines with a comprehensive approach that includes 19 items and six open-ended questions. Its validation procedures are thorough, with an alpha coefficient of 0.959 and high internal consistency (α=0.949) based on expert judges' evaluation. The other two instruments while reporting higher interrater agreement, lacked documentation of its process of face-content validation. Such reflects the instrument's detailed customization, including qualitative questions, that enhances its ability to provide in-depth insights into student learning experiences. The GE-APROAch aligns well with the criteria for content, such as item adequacy, relevance, and clarity.

In contrast, instruments 2 and 3, while also based on CHED CMO 20 s. 2013 and using 4-point Likert scales, exhibit notable differences. Instrument 2, with a validation coefficient of 0.912, includes fewer qualitative questions and lacks detailed pre-test validation, focusing primarily on three LO categories. Instrument 3 has unclear validation procedures and no pre-test data, with some differences in item wording and interpretation. Both instruments use similar Likert scales and NA options, but their interpretations of scores differ significantly from GE-APROAch. The proposed instrument's comprehensive design and high reliability highlight its superior alignment with CHED outcomes and its effective discrimination of learning outcomes across different domains, compared with that of what [11], [12] employed.

Table 1. Similarities and differences between the GE-APROAch and other instruments

Instrument 1 (GE-APROAch)	Instrument 2 [11]	Instrument 2 [11]
Framework and design: based on CHED CMO 20 s. 2013, it includes 19 items plus 6 open-ended questions. It underwent extensive validation procedures with an alpha coefficient of 0.959, demonstrating high internal reliability.	Framework and design: also based on CHED CMO 20 s. 2013, this instrument includes 19 items and 2 open-ended questions. Validation procedure is described with an alpha coefficient of 0.912, but details are sparse.	Framework and design: based on CHED CMO 20 s. 2013, it includes 19 items and 1 open-ended question. Validation is unclear, with no statistical results provided, and 5 internal experts involved.
Pre-test and scale: pre-tested with 84 students and shows high internal consistency (α=0.949). Utilizes a 4-point Likert scale without a "not applicable" (NA) option.	Pre-test and scale: pre-tested with 114 students but lacks validation on pre-test results. Uses a 4-point Likert scale with a NA=0 option.	Pre-test and scale: no pre-test reported. Uses a 4-point Likert scale with a NA option.
Customization and revision: customized for clarity, cultural sensitivity, and efficiency, and revised based on expert feedback. Results from 5 courses under the arts and humanities domain show learning outcomes (LOs) were achieved at a great extent (X=3.69).	Customization and revision: rephrased CHED outcomes, maintaining fidelity but with some omissions. Personal pronouns are used in some items (n=7). Results show the MATWORLD course under the math, science, and technology domain achieved outcomes at a great extent (X=3.22; N=1,572).	Customization and revision: worded 12 LOs differently while maintaining fidelity to CHED, though some omissions are evident. Personal pronouns are used in some items (n=7). Results indicate the BIBCHUR course under the philosophy and social science domain achieved outcomes at a great extent (X=3.33; N=176).
Focus: 19 LOs and emphasizes detailed indicators and provides examples for each item, enhancing the specificity of assessment.	Focus: interpretation is narrowed to three LO categories, potentially limiting the breadth of the assessment. Few open-ended questions	Focus: also focused on three LO categories, similar to instrument 2. Included only 1 open ended question

Given that the GE-APROAch uses a 4-point Likert scale, interpreted in five ranges with a single freedom of choice of response and 0.6 interval, it sets a higher standard of expectation than its predecessor(s) with such scale of interpretation, as Spady [4] posits. On the assumption that all responses are relevant and no "not applicable" category is needed, the GE-APROAch proves to be a more effective tool for detailed and precise assessment. By omitting the "not applicable" category and starting its scale at 1.00, it ensures that every response is classified into relevant categories, streamlining the evaluation process. Its finely segmented categories—such as "at a moderate extent," "at a high extent," and "at a great extent"—allow for a nuanced interpretation of varying degrees of applicability. The higher threshold for the "at a great extent" category (starting at 3.44) ensures that only responses with a significant degree of relevance are included, which enhances the precision of high-applicability assessments.

On the other hand, the interpretation of scores shared in previous studies [11], [12] includes a "not applicable" category and has broader classifications like "to a very slight extent" and "to a slight extent". While this approach covers a wide range of scores, it may introduce unnecessary complexity and lack the finer granularity needed for detailed analysis. The threshold for high applicability in their measure starts at 3.30, encompassing a wider range of scores and potentially diluting the distinction between high levels of perceived outcomes achievement. Overall, this may be suitable for general assessments but lacks the precision of the GE-APROAch which offers a more precise and streamlined evaluation of relevance, making it more preferable for detailed analysis where all responses are relevant and nuanced insights are required.

4.2. Expert validation results

Data in Table 2 indicates a high level of perceived effectiveness of the GE-APROAch item constructs as assessed by expert judges, with most items categorized as "outstanding." Reliability scores are uniformly high at 4.00 across all learning outcomes, reflecting a robust consistency in the measures. Similarly, validity scores are predominantly 4.00, though a few outcomes such as LO5, LO8, and LO9 show slightly lower validity, indicating minor areas for improvement. Clarity scores, however, vary more significantly, with some outcomes scoring as low as 3.50, suggesting that while most items are clear, there is room for enhancing clarity in certain areas. Overall, the mean score of 3.80 reflects a generally high level of reliability, validity, and clarity across all learning outcomes. The Cronbach's alpha value of 0.959 supports excellent internal consistency, and the individual alpha values for each criterion further reinforce the reliability of the assessment items. The observed agreement is greater than the expected (P_o=0.807>0.250) which indicates that the results are reliable and not due to random chance, adding to the credibility of the assessment. Consistency in higher interrater agreement indicates reliability of the GE-APROAch as an instrument to measure learning outcomes achievement [28], [32], [49].

Table 2. Interrater agreement on GE-APROAch quantitative item constructs

Items	Reliability score	Validity score	Clarity score	$\bar{X}$	Least extent	Some extent	Moderate extent	Great extent	α	P_o	k	SCV-I	Interpretation
LO1	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO2	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO3	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO4	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO5	4.00	3.75	3.50	3.75	0	0	4	8	0.953	0.667	0.556	1.00	Outstanding
LO6	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO7	3.75	3.75	3.75	3.75	0	0	4	8	0.953	0.667	0.556	1.00	Outstanding
LO8	3.50	3.50	3.50	3.50	0	3	1	8	0.963	0.667	0.556	0.75	Very satisfactory
LO9	3.50	3.50	3.50	3.50	0	3	0	9	0.959	0.750	0.667	1.00	Very satisfactory
LO10	4.00	3.75	3.75	3.83	0	0	2	10	0.954	0.833	0.778	1.00	Outstanding
LO11	3.75	3.75	3.75	3.75	0	0	3	9	0.953	0.750	0.667	1.00	Outstanding
LO12	4.00	4.00	4.00	4.00	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO13	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO14	3.75	3.75	3.50	3.67	0	0	5	7	0.955	0.583	0.444	1.00	Outstanding
LO15	3.75	4.00	3.50	3.75	0	0	3	9	0.953	0.750	0.667	1.00	Outstanding
LO16	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO17	3.50	3.50	3.50	3.50	0	0	6	6	0.959	0.500	0.333	1.00	Very satisfactory
LO18	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
LO19	4.00	4.00	3.75	3.92	0	0	1	11	0.958	0.917	0.889	1.00	Outstanding
All items	3.87	3.86	3.68	3.80	0	6	38	184	0.959	0.807	0.743	0.98	Outstanding
A	0.777	0.791	0.926	0.890	0.00	0.03	0.18	0.81	0.250=(P_e) probability of chance agreement*				

Data in Table 3 demonstrates that open-ended question items in the GE-APROAch display a generally strong performance, with the majority categorized as "outstanding" or "very satisfactory", as rated by expert judges. Specifically, Q1, Q2, and Q6 are rated "outstanding," indicating high reliability, validity, and clarity. These items have high scores across all criteria, with reliability and validity scores reaching 4.00 and clarity scores also being strong. Notably, Q6 stands out with perfect scores in reliability, validity, and clarity, reflecting exceptional quality. Mean scores across all items are 3.67 for reliability, 3.88 for validity, and 3.58 for clarity, resulting in an overall average of 3.80. These figures suggest that, on average, the items are valid and reliable, though there is some variability in clarity. The Cronbach's alpha (α=0.956) for the overall dataset demonstrates excellent internal consistency. Individual α values for reliability (0.738), validity (0.733), and clarity (0.809) also reflect good consistency, though the clarity scores are slightly lower. As one rater observed:

> "*There are some items that are constructed in a less concise manner. Please improve the clarity of all the items. Make it more concise by trying to separate some items into two. Look at some words that have more understandable meaning for the students. Some items are too wordy, so please check it as well.*"

Table 3. Interrater agreement on GE-APROAch qualitative questions

Items	Reliability score	Validity score	Clarity score	$\bar{X}$	LE (1)	SE (2)	ME (3)	GE (4)	α	P_o	K	SCV-I	Interpretation
Q1	3.75	4.00	3.75	3.83	0	0	2	10	0.925	0.833	0.778	1.00	Outstanding
Q2	4.00	4.00	3.50	3.83	0	0	1	11	0.931	0.917	0.889	1.00	Outstanding
Q3	3.50	3.75	3.25	3.50	0	0	7	5	0.966	0.583	0.444	1.00	Very satisfactory
Q4	3.25	3.75	3.50	3.50	0	1	4	7	0.937	0.583	0.444	0.75	Very satisfactory
Q5	3.50	3.75	3.50	3.58	0	0	5	7	0.938	0.583	0.444	1.00	Very satisfactory
Q6	4.00	4.00	4.00	4.00	0	0	0	12	0.944	1.000	1.000	1.00	Outstanding
All items	3.67	3.88	3.58	3.80	0	1	19	52	0.956	0.750	0.609	0.95	Outstanding
α	0.738	0.733	0.809	0.825	0.000	0.014	0.264	0.722	= (P_e) Probability of chance agreement*				

Items Q3, Q4, and Q5 are categorized as "very satisfactory," showing some variability in their scores. These items have slightly lower reliability and validity scores compared to the "outstanding" items, and their clarity scores also reflect some room for improvement. The variability in these items' scores suggests that while they are generally effective, there is potential for enhancement in terms of reliability,

validity, and clarity. The interrater agreement is greater than the expected (P_o=0.750>0.250) which indicates that the results are reliable, adding to the credibility of the assessment indicating that the observed agreement in the data is significant and not due to random chance.

The whole instrument was further rated by experts guided by the criteria of a good questionnaire identified in previous studies [50], [51]. Based on the data presented in Table 4, the evaluation reveals that most criteria are rated as "outstanding," indicating high quality across the board. Specifically, criteria such as relevance, clarity, intelligibility, layout, efficiency, reliability, and ethics all achieved average scores of 4.0 or higher, with strong observed agreement scores of 1.00. The overall observed agreement for all criteria is 0.69, categorized as "strong," and the expected probability of chance agreement (P_o=0.690>0.200). This demonstrates a high level of consensus among evaluators and underscores the effectiveness of the items across criteria.

Table 4. Over-all quality assessment of GE-APROAch instrument

Criteria	P (1.0)	NI (2.0)	S (3.0)	VS (4.0)	O (5.0)	Σ	$\bar{X}$	Observed quality	P_o	K	Agreement
Relevance	0	0	0	0	4	4	4.0	O	1.00	1.00	Strong
Depth	0	0	0	2	2	4	3.6	O	0.50	0.38	Moderate
Structure	0	0	0	3	1	4	3.4	VS	0.75	0.69	Strong
Clarity	0	0	0	1	3	4	3.8	O	0.75	0.69	Strong
Consistency	0	0	0	2	2	4	3.6	O	0.50	0.38	Moderate
Intelligibility	0	0	0	1	3	4	3.8	O	0.75	0.69	Strong
Layout	0	0	0	1	3	4	3.8	O	0.75	0.69	Strong
Efficiency	0	0	1	0	3	4	3.6	O	0.75	0.69	Strong
Motivation	0	0	0	2	2	4	3.6	O	0.50	0.38	Moderate
Reliability	0	0	0	1	3	4	3.8	O	0.75	0.69	Strong
Adaptability	0	0	0	2	2	4	3.6	O	0.50	0.38	Moderate
Ethics	0	0	0	1	3	4	3.8	O	0.75	0.69	Strong
Observed agreement	0.00	0.00	0.02	0.33	0.65	1.0	3.7	Outstanding	0.69	0.61	Strong

P_e=0.200 (expected probability agreement). P=poor, NI=needs improvement, S=satisfactory, VS=very satisfactory, O=outstanding

Conversely, criteria such as depth, consistency, motivation, and adaptability received average scores of 3.6, falling into the "moderate" or "very satisfactory" categories. These criteria showed lower agreement levels, with scores ranging from k=0.38 to 0.50, suggesting some variability in evaluator opinions and indicating areas for potential improvement. One evaluator noted, "the instrument is appropriate to the level of comprehension of the target respondents." Another supports this, saying:

> "*It is a good thing that most of the items are valid enough*."
> "*Since the goal is to make this instrument standardized, pilot-test this instrument with a number of students who are not part of the study*."

Results showed that GE-APROAch has attributes of a good instrument being informed by a well-defined domain which provides "a working knowledge of the phenomenon under study, specify the boundaries of the domain, and ease the process of item generation and content validation" [52].

4.3. GE-APROAch pre-test results

Pre-testing is necessary to validate the effectiveness of an instrument, to assess it the extent of how questions reflect the domain being studied, and how those questions produce responses [52]. Table 5 presents the pilot test scores across five GE subjects (N=81) under the communication literature within the arts and humanities domain. The overall average reliability score for all LOs is α=0.949, indicating high internal consistency in student's ratings across the criteria. This suggests that the responses from pre-test respondents are consistently dependable across items, with several LOs demonstrating strong reliability, including LO6 (α=0.8095), LO13 (α=0.8214), and LO3 (α=0.7738).

Using ANOVA for independent measures, the F-ratio of 9.86 with a p-value <0.00001 indicates a significant result (p<0.05). This leads to rejecting $H0_1$, which posits no difference in student scores across GE classes (μ_d=0). Results show that the instrument effectively differentiates between group scores within the same knowledge domain. The standard deviations range from 0.053 to 0.217, reflecting high internal reliability. Additionally, pre-test results reveal excellent internal reliability (α=0.949), strong observed agreement (P_o=0.7386), and substantial inter-rater agreement (k=0.673), confirming the instrument's consistency in measurement. Across items, this interrater agreement shows higher reliability of the instrument, efficiently measuring variance in scores with consistency [53].

Table 5. Pre-test results using GE-APROAch across NGEC arts and humanities courses

LOs	GE1	GE2	GE3	GE4	GE5	$\bar{X}$	α	P_o	k
LO1	3.88	3.59	3.81	3.50	3.30	3.61	0.947	0.7143	0.643
LO2	3.88	3.59	3.83	3.50	3.30	3.62	0.947	0.7143	0.643
LO3	3.88	3.76	3.78	3.67	3.50	3.72	0.945	0.7738	0.717
LO4	3.75	3.82	3.86	3.42	3.60	3.69	0.945	0.7976	0.747
LO5	3.50	3.59	3.78	3.17	3.30	3.47	0.946	0.6667	0.583
LO6	3.75	3.59	3.89	3.75	3.70	3.74	0.947	0.8095	0.762
LO7	3.75	3.47	3.69	3.75	3.80	3.69	0.946	0.7143	0.643
LO8	3.63	3.71	3.81	3.25	3.50	3.58	0.947	0.7024	0.628
LO9	3.63	3.71	3.78	3.58	3.60	3.66	0.944	0.7381	0.673
LO10	3.63	3.59	3.83	3.25	3.78	3.61	0.945	0.7262	0.658
LO11	3.75	3.53	3.81	3.67	3.78	3.71	0.947	0.7262	0.658
LO12	3.75	3.59	3.78	3.92	3.50	3.71	0.947	0.7381	0.673
LO13	3.88	3.71	3.86	3.75	3.80	3.80	0.945	0.8214	0.777
LO14	3.88	3.53	3.69	3.42	3.60	3.62	0.946	0.6905	0.613
LO15	3.50	3.65	3.83	3.67	3.70	3.67	0.946	0.7738	0.717
LO16	3.75	3.76	3.80	3.75	3.70	3.75	0.947	0.7857	0.732
LO17	3.88	3.71	3.81	3.50	3.70	3.72	0.946	0.7619	0.702
LO18	3.50	3.47	3.72	3.17	3.70	3.51	0.944	0.6548	0.568
LO19	3.75	3.59	3.78	3.50	3.40	3.60	0.954	0.7143	0.643
Mean	3.73	3.63	3.80	3.54	3.59	3.66	0.949	0.7386	0.673

The data in Table 6 highlights that most outcomes have high reliability and significant agreement, although some variations are present. Given the outstanding experts' evaluation of the GE-APROAch, the lower scores with high agreement among student respondents reflect there is not much evidence that they were met at greater degree. Low scores despite clear constructs can often be attributed to perceptions of relevance, alignment, practical utility, and expectations. If respondents agree that certain LOs are less impactful, challenging, or well-aligned with course goals, this consensus can lead to lower ratings. The reliability of an instrument to differentiate high and low ratings indicates its efficacy to inform future instructional adjustments.

Table 6. Paired split group comparison of pre-test scores using the GE-APROAch

LOs	Treatment group 1	Treatment group 2	Diff (T2-T1)	Dev (Diff-M)	Sq. Dev
LO1	3.66	3.67	0.01	-0.05	0
LO2	3.66	3.69	0.03	-0.03	0
LO3	3.74	3.71	-0.03	-0.09	0.01
LO4	3.68	3.83	0.15	0.09	0.01
LO5	3.61	3.57	-0.04	-0.1	0.01
LO6	3.71	3.81	0.1	0.04	0
LO7	3.68	3.69	0.01	-0.05	0
LO8	3.63	3.69	0.06	0	0
LO9	3.63	3.74	0.11	0.05	0
LO10	3.62	3.76	0.14	0.08	0.01
LO11	3.68	3.73	0.05	-0.01	0
LO12	3.76	3.69	-0.07	-0.13	0.02
LO13	3.74	3.86	0.12	0.06	0
LO14	3.58	3.67	0.09	0.03	0
LO15	3.66	3.76	0.1	0.04	0
LO16	3.76	3.80	0.04	-0.02	0
LO17	3.76	3.79	0.03	-0.03	0
LO18	3.47	3.64	0.17	0.11	0.01
LO19	3.58	3.71	0.13	0.07	0
Overall	3.66	3.73	M: 0.06		S: 0.08

A pair-wise comparison of scores (N=80) of split groups revealed that treatment group 2 (3.73) rated the course outcomes achievement higher than treatment group 1 (3.66), with a mean difference of 0.06 and a standard deviation of 0.08. A two-tailed t-test yielded a t-value of -4.29 and a p-value of 0.00044 to reject the null hypothesis that GE-APROAch is not sensitive to differentiate scores between the split-group taking the same course ($\mu_1=\mu_2$). Results confirm that these differences are statistically significant and not due to chance. This analysis demonstrates that the GE-APROAch instrument effectively discriminates between student ratings of outcomes achievement within the same class, indicating its reliability and effectiveness in measuring consistency in performance.

"Effective discriminating fineness is associated with both a high expected difference or steep slope, and a small error variance" [54]. Overall, the data supports that the instrument consistently measures

performance related to meeting learning outcomes, as evidenced by the low mean differences and standard deviations (M=0.06, S=0.08). The strong statistical significance and low variance suggest that while there are some variations in ratings, the instrument reliably captures the consistency of outcomes achievement within groups per course. Comparing these results with similar instruments could further validate the findings and reinforce the GE-APROAch's reliability and discriminative power.

4.4. Convergent and discriminant validity between existing instruments

In Table 7, scores from the three studies represent a picture of how the CHED mandated learning outcomes are met across the three GE knowledge domains, at the same college. Interpretations were based on the GE-APROAch's descriptive interpretation using a 4-point Likert scale. Bracketed interpretations at the last row reflect how the three studies interpreted their survey results. When compared to each researcher's standards, the earlier instruments [11], [12] seem to bring higher valuations in their interpretation, while the GE-APROAch is consistent with existing normative interpretation. Data for T_1 (N=84) reflects students' appreciation of the GE courses and electives as to the extent of outcomes achievement.

The instrument achieved high scores across three target competency categories, with arts and humanities outcomes reflecting superior performance compared to personal and civic responsibilities. Intellectual competence was achieved at a high extent in arts and humanities, contrasting with the moderate extent observed in previous studies (math and natural science and social science domains). Practical skills were rated highly in T_2 and arts and humanities but only moderately in T_1. These interpretations are based on the GE-APROAch standards, which better align with CHED QA standards compared to earlier scales. Distribution of scores in T_1 shares similar pattern with T_2 but not that of T_3, which either implies nuances in quality of instruction or limitations in the instruments used.

Table 7. GE-APROAch pre-test and actual evaluation results of two other instruments

Los	Target competencies	T_1	T_2	T_3	$\bar{X}$	Interpretation
LO1	Multiliteracies	3.67	3.23	3.32	3.39	High extent
LO2	Communication	3.68	3.14	3.12	3.29	Moderate extent
LO3	Integrative thinking	3.74	3.34	3.37	3.48	High extent
LO4	Innovative thinking	3.75	3.29	3.39	3.46	High extent
LO5	Analytical thinking	3.57	3.16	3.34	3.32	Moderate extent
CC1	Intellectual competence	3.62	3.23	3.31	3.39	High extent
LO6	Social awareness	3.77	3.64	3.15	3.51	High extent
LO7	Reflective thinking	3.68	3.51	3.18	3.46	High extent
LO8	Global perspective	3.64	3.32	3.03	3.31	Moderate extent
LO9	Sense of being Filipino	3.68	3.25	2.84	3.25	Moderate extent
LO10	Ethical reasoning	3.67	3.43	3.22	3.42	High extent
LO11	Moral responsibility	3.70	3.54	3.05	3.43	High extent
LO12	Artistic expression	3.70	3.39	3.20	3.43	High extent
LO13	Respecting others	3.81	3.56	2.98	3.45	High extent
LO14	Social responsiveness	3.62	3.38	3.07	3.36	High extent
CC2	Civic responsibilities	3.68	3.45	3.08	3.40	High extent
LO15	Collaboration	3.73	3.22	3.35	3.41	High extent
LO16	Information/Digital literacy	3.77	3.39	3.39	3.51	High extent
LO17	Digital fluency	3.74	3.28	3.31	3.44	High extent
LO18	Problem-solving	3.57	3.32	3.40	3.41	High extent
LO19	Professional readiness	3.65	3.33	3.39	3.44	High extent
CC3	Practical skills	3.60	3.33	3.39	3.44	Great extent
Mean and interpretations		3.66 [GE]	3.35 [GE]	3.22 [ME]	3.39	High extent

T_1: GE-APROAch, T_2: instrument 1, T_3: instrument 3

ANOVA analysis confirmed significant variability in scores across knowledge domains, with F-ratio value of 58.08549 and a p-value <0.00001, effectively rejecting the null hypothesis that GE-APROAch is as sensitive to yields similar results as its predecessors (T_1=T_2=T_3). The standard deviations demonstrate that GE-APROAch (S_1=0.065) offers greater reliability in distinguishing item scores than other instruments (S_2=0.13 and S_3=0.18). Significant differences were observed between GE-APROAch and earlier instruments (HSD T_1=0.3, Q=9.94, p=0.00000<0.05; T_1=0.44, Q=14.46, p=0.00000<0.05), highlighting its superior discriminative power. It can be inferred from the data that the GE-APROAch instrument exhibits strong internal consistency and effective discriminative power, distinguishing itself from previous instruments.

Overall, compared with two other instruments, the GE-APROAch which was pilot tested in five different courses has ability to measure learning outcomes across various knowledge domains and aligns with construct validity principles, where convergent validity is demonstrated through its alignment with other theoretically related measures [55], and discriminant validity is shown by its differentiation from unrelated

measures [56]. The instrument's high reliability and significant differences in results underscore its effectiveness and potential for broader application in educational evaluations. Future comparisons with similar instruments will further validate its construct validity and its impact compared to existing measures [57].

4.5. Consistency between quantitative ratings and qualitative responses

Three major themes emerged in the qualitative analysis of student responses, pertaining to individual development or intellectual competence (f=68), community and national development or personal and civic responsibilities (f=67) and global competence and practical skills (f=71). Quantitatively (Table 7), civic responsibilities (3.68) scored the highest among three outcomes clusters, followed by intellectual competence (3.62) then by practical skills (3.60). The sense of civic responsibility which "pertains to the responsibilities of a citizen in terms of connection to the community, civic awareness, and civic efficacy" [58], appears to extend to a global level. Responses show that GE courses under the arts and humanities domain have positive impact on students.

> "*The course played a significant role in shaping my perception and appreciation of my national or collective identity. By delving into topics related to societal development and cultural heritage, I gained a deeper understanding of the diverse perspectives and experiences within my society. This, in turn, motivated me to consider practical ways in which I could contribute to societal development, whether through community engagement, advocacy, or active participation in civic initiatives.*"

It also appears that this civic responsibility learned in the communication and literature classes is not only confined to the local context but also extends beyond national borders, as one student iterates:

> "*The course broadened my perspectives on global issues, emphasizing the importance of respect for diversity, empathy for others, and active participation in solving global challenges. It instilled in me a heightened awareness of interconnectedness and a commitment to fostering collaboration and understanding across cultural, geographical, and socioeconomic boundaries*."

Personal growth and intellectual competencies which also relate to individual development are descriptive of the students' sensed academic skills enhancement, attained communication proficiency, and also cultural competence. This broad concept relates to life long, reflexive, and holistic learning that leads to one's state of self-actualization [59]. The CMO's description of individual growth, "where a student is enabled to develop his/her identity as a person, conscious of his/her talents, rights and responsibilities towards others" [25] is well described in this student's reflection:

> "*The course provided a platform for me to explore and understand my individual identity, talents, rights, and responsibilities towards myself and others. Through various activities, discussions, and assignments, I was able to reflect on my strengths and weaknesses, clarify my values, and develop a deeper sense of self-awareness. This process enabled me to grow personally by identifying areas for improvement and setting goals for self-development.*"

Global competence and practical skills relate to goal of building a global community. Global competence [34] directly associates with practical skills outcomes. The CMO reads, "global community, where the Filipino student recognizes and respects the fundamental humanity of all, respects and appreciates diversity, and cares about the problems that affect the world" [25]. Such is reflected in this response:

> "*This course affected my perspectives as a member of a global community by changing the way I see others. Because before, I tend to judge them in my mind without thinking about their experience or thoughts first. My respect and appreciation of diversity gained a lot. My concern for others also changed because before I did not really care much about it. My desire to participate in solving problems also enhanced because I want our world to be a better place to live.*"

While it is evident that GE courses have strong positive impact on the student's development of intellectual competence, socio-civic responsibilities and practical skills, there are a few who do not find its value the way most students do. The impact of GE on students positively relates to instructional delivery and design as facilitated by the teacher, while the negative perspective on GE arises from student's academic load and managing their coursework. This contrast is evident in the following student views:

"*In my experience, my professor made the course's pacing just right when it came to the flow of topics to requirements. Additionally, the way our professor was able to disperse everyone into random groups after every activity was very helpful since you do not stick to one group for a long time which somehow trains you to adapt quickly with whoever you are working with*." To another: "*The pacing is okay; the only problem is that the other courses have conflicts with schedules especially with the due dates. It is hard to focus on a course especially when you know that there are other courses with the same due date*."

Under practical skills, information and digital literacy received the highest score but did not appear to be emphasized in the qualitative response. What students can recall and perceive may have longer lasting and transferable effect on them, which results from their active learning engagement [60]. This implies that although such outcomes are highly met in the courses, they may not be as memorable to them, as they are integral or common already in their life as digital natives. Challenges and areas of improvement have an impact on student learning and the achievement of outcomes. Perceived learning directly associates with students' satisfaction which relate to their appreciation of their level of engagement, interaction, motivation, and on the learning environment as to its structure, the instructor and the knowledge they acquired in the course [61]. Such information can be richly collected qualitatively. Hence, the consistency in quantitative and qualitative responses observed in the pre-test results supports the instrument's efficacy.

5. DISCUSSION

This study evaluated the validity of the GE-APROAch, an instrument designed to measure outcomes achievement in the Philippine GE curriculum, in alignment with the CHED's mandated outcomes. The statistical analysis demonstrated that the GE-APROAch: i) effectively discriminates between scores across different groups or courses; ii) reliably measures shared perceptions among students from the same course; and iii) quantitatively and qualitatively evaluates GE outcomes achievement with high efficacy and efficiency. While an instrument's reliability is important, it alone does not guarantee validity. A reliable instrument may still produce questionable findings if it is not valid for its intended purpose [62].

The absence of standardized measures for assessing learning outcomes can hinder decision-making and obstruct progress towards meeting sustainable development goals in education [63]. The reviewed literature highlights gaps in how different institutions measure GE outcomes [11], [12]. Validity and reliability extend beyond statistical analysis; they involve assessing the instrument's ability to accurately measure what it intends to measure and how well it adheres to theoretical and practical standards. In these aspects, the GE-APROAch surpasses earlier instruments, demonstrating its capacity to meet research purposes effectively. It is conceptually relevant, well-organized, and uses appropriate language register for efficient response strategies and communication approaches [51], [64]. Previous instruments based on CHED's outcomes [11], [12] were conceptually valid but lacked transparency in their validation processes. A well-designed questionnaire should be valid, reliable, clear, engaging, and succinct, grounded in sound concepts to provide meaningful information [50]. All results in this validation report points to that.

Interrater agreement is crucial for justifying data aggregation in quantitative studies [65]. Expert evaluations confirmed that the GE-APROAch items are relevant, clear, and well-structured, reflecting a strong alignment with the CHED outcomes. The pre-test results showed that the instrument was efficient, producing reliable and consistent responses, with mutual exclusivity, variability and unitary directions towards a dimensional response, indicating accurate measures of concepts and properties it is designed to for [50]. Statistical tests revealed consistent scores among split groups from the same course and notable differences between courses, indicating the instrument's effectiveness and discriminative ability [31], [66]. The consistency between qualitative and quantitative data further supports the instrument's logical and empirical validity [63].

The analysis reveals that the GE-APROAch can distinguish between group scores on reported outcomes achievement (H1: $\mu_d \neq 0$). It also reliably measures shared perceptions within the same learning environment (H2: $\mu_1 \neq \mu_2$) and demonstrates significant differences in score variance compared to two other instruments (H3: $T_1 \neq T_2$, $T_2 \neq T_3$, $T_1 \neq T_3$). These findings confirm the instrument's validity in measuring GE outcomes and its psychometric properties, reflecting mastery, performance, motivation, and approaches influencing CHED's learning outcomes.

Applying a multifactor model to assess goal achievement has implications for individual, course and program-level evaluations [64]. Despite criticisms of self-reporting, the GE-APROAch aligns with OBE, constructivism, and student-centered pedagogy, emphasizing the importance of students' perspectives and learning experiences. Self-reports, while relying on verbal responses, are essential for assessing learning outcomes achievement [67], and are vital in promoting constructivist student-centered education.

6. CONCLUSION

The GE-APROAch demonstrates superior validity and reliability compared to previous instruments used to measure outcomes achievement, adopting the CHED prescribed GE outcomes. It excels in discriminating between item scores across different classes and between student groups within the same class, outperforming its predecessors in both statistical and qualitative measures. The findings highlight the GE-APROAch's effectiveness in assessing learning outcomes, offering valuable insights for policymakers and education researchers. Its design, grounded in OBE principles and guided by the CHED mandates, ensures a commitment to quality and a clear understanding of what it measures—learning outcomes beyond mere compliance. Future research should continue to explore and test the efficacy of the instrument to other courses, as to its relevance and effectiveness, across knowledge-domains and at large scale across HEIs.

The GE-APROAch, validated rigorously, is well-suited for use across colleges and universities for evaluating existing GE programs. The instrument's high validity and reliability, supported by expert judgment, pilot test results, and cross-comparative analysis, make it a tool for formative evaluation. This validation report underscores the need for researchers to adopt a rigorous, critical, and self-reflective approach in instrument design. By aligning with OBE frameworks and regulators' mandates, the GE-APROAch can inform curriculum development and instructional innovations. Its effectiveness is not only due to positive statistical results but also because of its systematic design, empathy towards respondents, and clear focus on measuring learning outcomes, not just compliance.

FUNDING INFORMATION

This study was self-initiated. No fundings in forms of research grant or honoraria were obtained from any institution.

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
Rodrigo C. Rivera	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Elymae Naldo Delos Santos	✓	✓		✓	✓	✓	✓	✓	✓	✓			✓	✓
Juan Oliver S. Ofracio	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓

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

The authors whose name are listed in this article certify that they have no affiliations or entity with commercial or financial interest, and that study was undertaken as an academic research endeavor.

INFORMED CONSENT

Informed consent for voluntary participation were obtained from students, who are of legal age to voluntary participate in this study. This consent was indicated in the online survey form.

ETHICAL APPROVAL

This study involved human respondents and was conducted in accordance with the law and ethical standards. Ethical approval to conduct the study was obtained from the appropriate school administrator. Data privacy was ensured under the Republic Act 10173. Anonymity, confidentiality, safety and the participants' right to withdraw from the study at any time were ensured, adherent to the National Ethical Guidelines for Research Involving Human participants by the Philippine Health Research Ethics Board.

DATA AVAILABILITY

The data of the study is available on the request of any interested researcher. The data can be accessed through direct email messages sent by the party concerned addressed to the primary author [RCR]. The data is intact in a drive kept by the authors.

REFERENCES

[1] Commission on Higher Education (CHED), *Handbook on typology, outcome-BSED education and institutional sustainability assessment*. Quezon City: Office of Institutional Quality Assurance and Governance, 2014.

[2] K.-T. Lindner and S. Schwab, "Differentiation and individualisation in inclusive education: a systematic review and narrative synthesis," *International Journal of Inclusive Education*, pp. 1–21, Sep. 2020, doi: 10.1080/13603116.2020.1813450.

[3] CHED Memorandum Order (CMO), "CMO 46, series of 2012: policy standard to enhance quality assurance QA in Philippine higher education through an outcomes based and typology based QA," Republic of the Philippines Office of the President, 2012. [Online]. Available: https://www.pacu.org.ph/wordpress/wp-content/uploads/2017/03/CMO-No.46-s2012.pdf

[4] W. G. Spady, *Outcome-based education: critical issues and answers*. Arlington, TX: American Association of School Administrators, 1994.

[5] M. A. A. Damit, M. K. Omar and M. H. M. Puad, "Issues and challenges of outcome-based education (OBE) implementation among Malaysian vocational college teachers," *Journal of Academic Research in Business and Social Sciences,* vol. 11, no. 3, pp. 197–211, 2021, doi: 10.6007/IJARBSS/v11-i3/8624.

[6] K. Tremblay, D. Lalancette, and D. Roseveare, *Assessment of higher education learning outcomes: Feasibility study report, Volume 1–Design and implementation.* Paris: OECD, 2012.

[7] G. Wiggins and J. McTighe, *Understanding by design*, 2nd ed. Alexandria, VA: Association for Supervision and Curriculum Development ASCD, 2005, doi: 10.14483/calj.v19n1.11490.

[8] O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach, and C. Kuhn, *Assessment of learning outcomes in higher education*. Cham: Springer International Publishing, 2018, doi: 10.1007/978-3-319-74338-7.

[9] J. E. Albia and S.-J. Chan, "Understanding regionalisation in Philippine higher education," *Higher Education Evaluation and Development*, vol. 11, no. 2, pp. 95–110, 2017, doi: 10.1108/heed-07-2017-0003.

[10] S. Mishra, *Quality assurance in higher education: an introduction*. Karnataka, India: National Assessment and Accreditation Council, 2007.

[11] B. Blay, "OBEdizing" the new general education course – mathematics in the modern world: an attempt to do it right - formative evaluation-students; perspectives," in *The Southeast Asian Conference on Education (SEACE)*, 2020.

[12] M. Pescasio, "Extent to which the competencies for the general education realized in Bib Chur," *European Journal of Teaching and Education*, vol. 4, no. 4, pp. 13–31, 2022, doi: 10.33422/ejte.v4i4.802.

[13] D. Nusche, "Assessment of learning outcomes in higher education: a comparative review of selected practices," *OECD Education Working Paper no. 15*, vol. 8, no. 45, pp. 36–77, 2008.

[14] K. H. Z. Salvaña and A. M. Costelo-Abrea, "General education science and mathematics courses of Philippine normal university's outcomes-based teacher education curriculum: intents and actualities," *Multicultural Education Journal (MEJ)*, pp. 41–54, 2021. [Online]. Available: https://www.researchgate.net/publication/331247509

[15] J. D. Dagdag, "Organizational structure and procedure barriers to OBEdizing Philippine higher education: implications to policies and practice," *Journal of Research, Policy and Practice of Teachers and Teacher Education*, vol. 10, no. 1, pp. 27–44, 2020, doi: 10.37134/jrpptte.vol10.1.3.2020.

[16] C. P. Custodio, G. N. Espita, and L. C. Siy, "The implementation of outcome-based education: the case of the university of perpetual help system DALTA Las Piñas Campus," in *SEMEAO RETRAC International Conference*, 2017, pp. 1–18.

[17] University of the Philippines-Diliman, "REPORT of the AD HOC committee: hybrid general education program review," Up Diliman, Quezon City, 2016.

[18] J. S. Fernando, "Compliance of higher education institutions to outcome-based education for the 21st century learners," *International Multidisciplinary Research Journal*, vol. 3, no. 2, pp. 226–236, 2021, doi: 10.54476/iimrj304.

[19] R. A. O.-D. Cruz, "Learners' attitude towards outcomes-based teaching and learning in higher education," *Tuning Journal for Higher Education*, vol. 9, no. 2, pp. 99–119, 2022, doi: 10.18543/tjhe.1965.

[20] R. T. Barrios, "Outcomes-based education practices in higher institutions in Aklan, Philippines," *Galaxy International Interdisciplinary Research Journal*, vol. 2, no. 8, pp. 29–37, 2021.

[21] E. Q. Tenedero and L. M. Pacadaljen, "Learning experiences in the emerging outcomes-based education (OBE) curriculum of higher education institutions (HEI's) on the scope of Hammond's evaluation cube," *Psychology and Education*, vol. 58, no. 2, pp. 10325–10332, 2021.

[22] J. T. Libo-on, D. Ph, E. O. Valencia, and D. Ed, "Purposive communication: CHED standards, teachers' syllabi, and curricular enhancement," *International Journal of Advance Study and Research Work*, vol. 3, no. 4, pp. 39–56, 2020.

[23] T. Pepito, "Perspectives on outcome-based education among faculty members teaching business courses at a Philippine university," *The Palawan Scientist*, vol. 11, pp. 49–64, 2019, doi: 10.69721/tps.j.2019.11.1.05.

[24] R. A. A. Ortega and R. Cruz, "Educators' attitude towards outcomes-based educational approach in English second language learning," *American Journal of Educational Research*, vol. 4, no. 8, pp. 597-601, 2016, doi: 10.12691/education-4-8-3.

[25] CHED Memorandum Order (CMO), "CMO 20, series of 2013: general education curriculum: holistic understandings, intellectual and civic competencies," Republic of the Philippines Office of the President, 2013. [Online]. Available: https://ched.gov.ph/wp-content/uploads/2017/10/CMO-No.20-s2013.pdf

[26] E. I. Obilor and G. U. Miwari, "Content validity in educational assessment," *International Journal of Innovative Education Research*, vol. 10, no. 2, pp. 57–69, 2022.

[27] P. G. Fernandez-Marcelo, B. L. Ho, J. F. Faustorilla Jr, A. L. Evangelista, M. Pedrena, and A. Marcelo, "Emerging ehealth directions in the Philippines," *Yearbook of Medical Informatics*, vol. 21, no. 1, pp. 144–152, 2012.

[28] B. Eagan, J. Brohinsky, J. Wang, and D. W. Shaffer, "Testing the reliability of inter-rater reliability," in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 2020, pp. 454–461, doi: 10.1145/3375462.3375508.

[29] R. Heale and A. Twycross, "Validity and reliability in quantitative studies," *Evidence-based Nursing*, vol. 18, no. 3, pp. 66–67, 2016.

[30] M. S. B. Yusoff, "ABC of content validation and content validity index calculation," *Education in Medicine Journal*, vol. 11, no. 2, pp. 49–54, 2019, doi: 10.21315/eimj2019.11.2.6.

[31] H. Taherdoost, "Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research," *International Journal of Academic Research in Management (IJARM)*, vol. 5, no. 3, pp. 28–36, 2016, doi: 10.2139/ssrn.3205040.
[32] S. G. Sireci, "On validity theory and test validation," *Educational Researcher*, vol. 36, no. 8, pp. 477–481, 2007, doi: 10.3102/0013189x07311609.
[33] L. Goff *et al.*, "Learning Outcomes Assessment- A practitioner's Guide." [Online]. Available: https://heqco.ca/pub/learning-outcomes-assessment-a-practitioner-s-handbook/
[34] M. Barrett, *Preparing our youth for an inclusive and sustainable world: The OECD PISA Global Competence framework, assessments and findings*. Paris: OECD, 2018.
[35] The Second Congressional Commission on Education (EDCOM2, *Basic education: measurement of learning outcomes*. Pasay City, Philippines: EDCOM2, 2023.
[36] D. Kennedy and M. McCarthy, "Learning outcomes in the ECTS users' guide 2015: some areas of concern," *Journal of the European Higher Education Area*, no. 3, pp. 1–14, 2016.
[37] T. Chamorro-Premuzic, M. A. Quiroga, and R. Colom, "Intellectual competence and academic performance: A Spanish study," *Learning and Individual Differences*, vol. 19, no. 4, pp. 486–491, Dec. 2009, doi: 10.1016/j.lindif.2009.05.002.
[38] P. Auspos, P. Brown, and S. A. Sutton, "Living cities and civic capacity: leadership, leverage and legitimacy," The Aspen Institute Roundtable on Community Change, 2007. [Online]. Available: https://www.gsapp.org/labs/C2ARL/pdf_files/Living-Cities-and-Civic-Capacity.pdf
[39] M. Mulder, J. Gulikers, H. Biemans, and R. Wesselink, "The new competence concept in higher education: error or enrichment?" *Journal of European Industrial Training*, vol. 33, no. 8, pp. 755–770, 2009, doi: 10.1108/03090590910993616.
[40] Y. Matsuzuka, "Validity of outcome-oriented, competency-based education in the age of global student mobility: implications from an EU-Japan comparative study on competencies expected of university graduates," *Higher Education Forum*, vol. 17, pp. 1–20, 2020, doi: 10.15027/48951.
[41] T. Açıkgöz and C. Babadogan, "Competency-based education: theory and practice," *Psycho-Educational Research Reviews*, vol. 10, no. 3, pp. 67–95, 2021, doi: 10.52963/PERR_Biruni_V10.N3.06.
[42] P. Licaunan, "Human capital development and innovation: the role of higher education," pp. 1–10, 2017. [Online]. Available: https://industry.gov.ph/wp-content/uploads/2017/05/Keynote-Remarks-on-Human-Capital-Development-and-Innovation_CHED-Chair-Dr.-Patricia-Licuanan.pdf
[43] Y. Gu and J. Warren, "Methods for descriptive studies," in *Handbook of eHealth Evaluation: An Evidence-based Approach*, F. Lau and C. Kuziemsky, Eds., Victoria, British Columbia Canada: University of Victoria, 2016, pp. 199–211.
[44] J. Creswell and V. L. P. Clark, *Designing and conducting mixed method research*, 3rd ed. Thousand Oaks, CA: SAGE Publications, Inc., 2018.
[45] S. E. Eaton, *Logic models: what they are and how to prepare one*. Canada: Eaton International Consulting Inc., 2010.
[46] P. F. McCawley, *The logic model for program planning and evaluation*. Moscow, Idaho: University of Idaho Extention, 2001.
[47] The Strategy Unit, *Using logic models in evaluation*. Leyland: Midlands and Lancashire, 2016.
[48] J. A. Grand-Guillaume-Perrenoud *et al.*, "Mixed methods instrument validation: evaluation procedures for practitioners developed from the validation of the swiss instrument for evaluating interprofessional collaboration," *BMC Health Services Research*, vol. 23, no. 1, p. 83, 2023, doi: 10.1186/s12913-023-09040-3.
[49] A. R. Fitzpatrick, "The meaning of content validity," *Applied Psychological Measurement*, vol. 7, no. 1, pp. 3–13, 1983, doi: 10.1177/014662168300700102.
[50] S. Roopa and M. Rani, "Questionnaire designing for a survey," *Journal of Indian Orthodontic Society*, vol. 46, no. 4_suppl1, pp. 273–277, 2012, doi: 10.1177/0974909820120509s.
[51] M. Sha and T. J. Gabel, *The essential role of language in survey research*. Triangle Park, NC: RTI Press, 2020, doi: 10.3768/rtipress.bk.0023.2004.
[52] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quiñonez, and S. L. Young, "Best practices for developing and validating scales for health, social, and behavioral research: a primer," *Frontiers in Public Health*, vol. 6, p. 149, 2018, doi: 10.3389/fpubh.2018.00149.
[53] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International Journal of Medical Education*, vol. 2, pp. 53–55, Jun. 2011, doi: 10.5116/ijme.4dfb.8dfd.
[54] P. J. Ferrando, "Assessing the discriminating power of item and test scores in the linear factor-analysis model," *Psicologica*, vol. 33, no. 1, pp. 111–134, 2012.
[55] C.-L. Chin and G. Yao, "Convergent validity," in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed., Dordrecht: Springer Netherlands, 2014, pp. 1275–1276, doi: 10.1007/978-94-007-0753-5_573.
[56] A. M. Hubley, "Discriminant validity," in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos Ed., Dordrecht: Springer Netherlands, 2014, pp. 1664–1667, doi: 10.1007/978-94-007-0753-5_751.
[57] D. R. Lehmann, "An alternative procedure for assessing convergent and discriminant validity," *Applied Psychological Measurement*, vol. 12, no. 4, pp. 411–423, 1988, doi: 10.1177/014662168801200409.
[58] M. Vernon, E. S. Gracel, and A. S. Saban, "Sense of civic responsibility of national service training program completers and con-completers in a university setting in the Philippines," *International Forum*, vol. 23, no. 2, pp. 163–177, 2020.
[59] C. R. Jain, D. K. Apple, and W. Ellis, "What is self-growth?" *International Journal of Process Education*, vol. 7, no. 1, pp. 41–52, 2015.
[60] W. R. Klemm, "What good is learning if you don't remember it?" *Journal of Effective Teaching*, vol. 7, no. 1, pp. 61–73, 2007.
[61] H. Baber, "Determinants of students' perceived learning outcome and satisfaction in online learning during the pandemic of COVID19," *Journal of Education and e-Learning Research*, vol. 7, no. 3, pp. 285–292, Aug. 2020, doi: 10.20448/journal.509.2020.73.285.292.
[62] L. Sürücü and A. Maslakçi, "Validity and reliability in quantitative research," *Business and Management Studies: An International Journal*, vol. 8, no. 3, pp. 2694–2726, 2020, doi: 10.15295/bmij.v8i3.1540.
[63] Sudaryono, U. Rahardja, Q. Aini, Y. Isma Graha, and N. Lutfiani, "Validity of test instruments," *Journal of Physics: Conference Series*, vol. 1364, no. 1, 2019, doi: 10.1088/1742-6596/1364/1/012050.
[64] N. Zhao *et al.*, "Psychometric properties of achievement goal constructs for Chinese students," *Frontiers in Psychology*, vol. 11, p. 531568, Oct. 2020, doi: 10.3389/fpsyg.2020.531568.
[65] T. A. O'Neill, "An overview of interrater agreement on Likert scales for researchers and practitioners," *Frontiers in Psychology*, vol. 8, p. 777, May 2017, doi: 10.3389/fpsyg.2017.00777.
[66] K. A. Barchard and V. Brouwers, "Internal consistency and power when comparing total scores from two groups," *Multivariate Behavioral Research*, vol. 51, no. 4, pp. 482–494, 2016, doi: 10.1080/00273171.2016.1166422.
[67] R. Pekruna, "Commentary: self-report is indispensable to assess students' learning," *Frontline Learning Research*, vol. 8, no. 3, pp. 185–193, 2020.

BIOGRAPHIES OF AUTHORS

Rodrigo C. Rivera is an assistant professor and full-time faculty of De La Salle-College of Saint Benilde, Manila. He earned his master in communication management at the Pamantasan ng Lungsod ng Maynila and is currently completing his dissertation project for Ph.D. in Communication at the Polytechnic University of the Philippines, Manila. His research interests, include multicultural communication, technology and education, and curriculum development. He can be contacted at email: rodrigo.rivera@benilde.edu.ph.

Elymae Naldo Delos Santos is a full-time faculty member at the De La Salle-College of Saint Benilde, Manila. She majored in journalism at the University of the Philippines, Baguio and finished master of arts in curriculum and instruction with specialization in English at the Far Eastern University, Manila. She can be contacted at email: elymae.delossantos@benilde.edu.ph.

Juan Oliver S. Ofracio completed his master in English language and literature teaching at the Ateneo de Manila University and is a full-time faculty member at the De La Salle-College of Saint Benilde. He is the research coordinator for the communication and literature domain of the School of DLS-CSB's multidisciplinary studies. His research interests include English language teaching, novice teacher experience, facilitating learner engagement, and literature teaching, and art appreciation. He can be contacted at email: juanoliver.ofracio@benilde.edu.ph.