SEA-PLM
Southeast Asia Primary Learning Metrics

Review of the SEA-PLM 2024 Writing Domain: Technical implications, comparability limitations, and options on the ways forward in SEA-PLM Surveys

This technical note presents the findings of the Regional Secretariat's ongoing comprehensive investigation into the SEA-PLM 2024 writing domain results. The review aims to ensure the appropriateness, fairness, and interpretability of student performance outcomes across participating countries and assessment cycles. It was prompted by the emergence of unprecedented patterns in country-level results, which warranted closer examination of item functioning, scoring consistency, and cross-country comparability. The process undertaken and the outcomes of this investigation are detailed below.

The SEA-PLM design and administration

i) SEA-PLM is grounded in rigorous theoretical and empirical foundations to ensure the reliability and validity of constructs and data across languages and survey iterations. Maintaining alignment with established methods, especially those developed during the first round in 2019 of SEA-PLM, is critical for preserving the robustness of test comparability and score interpretation. Any deviation from these international methods, procedures and standards could weaken the integrity of the data and lead to false or incorrect conclusions.

ii) By participating in SEA-PLM, countries and stakeholders commit to generating, applying, and transparently documenting the technical processes and methodological decisions at each stage of data collection, analysis, and reporting. This collective adherence is essential to support the comparability of student performance data and to uphold the credibility of SEA-PLM results across cycles.

iii) The SEA-PLM test design utilizes a rotated book design, with each student responding to only two out of the three domains (reading, mathematics, writing); each booklet contains only a subset of items from each of the two attributed domains. This incomplete test design structure (18 booklets composed of six clusters of items per domain - 18 in total - in rotation between the booklets) was also applied in 2019. The total time allocated for the test is one hour. Each booklet is attributed randomly to each participating student in the selected Grade 5 class while ensuring that the total number of students' responses to each item is equivalent in each school and country and, in each round of assessment.

iv) Scores for students are estimated as plausible values because each student completed only a subset of items contained in their randomly assigned booklet. Five plausible values were estimated for each student for each scoring scale per domain. These values represent the distribution of potential scores for all students in the population with similar characteristics and identical patterns of item response. During result analysis, the five plausible values are combined to produce estimates of group achievement. This approach (combining scores from

different test to estimate a given domain[1]) provides a methodology that allows for statistical tools to estimate population characteristics as used by many international large-scale assessments[2]. This methodology was approved by the 3rd Regional Steering Committee in February 2015 and has been adopted in the two rounds of assessments, SEA-PLM 2019 and SEA-PLM 2024. This resulted in having a test design of having booklets and scoring relying on 3 domains.

The writing domain in SEA-PLM

v) The writing domain is one of the mainstream cognitive domains in the SEA-PLM test design, along with reading and mathematics. While all of the domains provide evidence and insights to inform curriculum reform on literacy and more broadly on high-order skills in primary education, writing is not yet considered part of SDG 4.1 reporting; only reading and mathematics minimum proficiency levels are used for international benchmarking. The SEA-PLM 2019 writing domain received much attention during the development and implementation of the first study, especially after the release of the SEA-PLM 2019 writing scale, results and reports.

vi) As defined in the SEA-PLM 2024 Test and Questionnaire Framework, writing is assessed through constructed-response tasks that require students to produce extended written texts based on age-appropriate prompts and real-world contexts. The assessment evaluates multiple dimensions of writing proficiency, including content development, organization, language use, and grammatical accuracy. Student responses are scored using standardized rubrics such as coding guides to ensure consistency and comparability across languages and countries.

vii) The writing coding criteria are predetermined by the SEA-PLM writing domain framework and described in the writing coding guide. All writing questions pertain to complex, open-ended items, which are the most complex to code in comparative large-scale assessment. Students must produce written responses to all questions; in comparison, two-thirds of the mathematics and reading questions are open-choice questions, with only one third as open short responses. Marking the responses in the writing domain therefore requires the use of coding guides to direct human expert judgments and to assign appropriate code for each of the responses.

viii) The coding guide for the SEA-PLM writing domain specifies that the open-ended questions are coded with partial credits. Responses may receive zero (0), partial credits (1 or 2, depending on the number of criteria) or full credits (2 or 3 depending on several criteria). The reading and mathematics domains rely on zero (0) or full credit score of one (1) only.

[1] As reported in SEA-PLM 2019 Main Regional Report p35 - "A conditioning 3-dimensional model was built for each country. The school mean performance variable adjusted for the student's own performance (WLE), dummy variables for the sampling stratum variable, school type, school location and student gender were used as direct regressors in the model. Most of the other student background variables – POCC (parental highest occupation), PARED (parental highest education), HOMERES (resources in the home) – and responses to questions in the student questionnaire were re-coded into dummy variables that were transformed into components by a principal component analysis (PCA). The principal components were estimated for each country separately. Subsequently, the components that explained 99% of the variance in all the original variables were included as regressors in the conditioning model."

[2] Wu, M. (2005) The role of plausible values in large-scale surveys. Studies in Educational Evaluation. https://www.acer.org/files/plausiblevaluesinsee.pdf

ix) Consequently, in SEA-PLM (as in other international large-scale assessments) the writing domain is more resource-intensive and time-consuming; for example, coding the writing items requires 60-80% more resources and time than marking the reading and mathematics domains as it requires human expert judgement to assign zero, partial or full credit.

The writing domain in SEA-PLM 2024

x) A standalone writing scale with eight described proficiency levels was generated based on scores derived from the student ability in reading, mathematics, writing, and contextual questionnaire data from the 6 participating countries for the SEA-PLM 2019 cycle.

xi) While the SEA-PLM 2019 writing domain showed solid validity and reliability in the context of the first cycle of SEA-PLM implementation, the SEA-PLM 2019 writing results have been questioned by some stakeholders in comparison to the overall pattern of performance observed in the reading and mathematics domains for some countries, mainly due to the constraints imposed in coding the domain assessment.

xii) As with other ILSAs, new rounds of data collection offer the opportunity to update and confirm methodology and measurement validity. This can be done through the revision of some components of the design and methodology and by examining empirical linkages between assessments and scores. The SEA-PLM 2024 cycle collaborated with countries and experts to align those criteria and investigate methods and processes to deliver the most reliable measurements.

xiii) All of the 2019 writing test items have been used in 2024, with no new items created for the new round. All writing questions are therefore considered as eligible trend items. Comparatively, in 2024, two-thirds of the reading and mathematics items were revised and field-trialed in collaboration with the participating countries.

xiv) In consultation with participating countries, based on lessons learnt in the first cycle, a double coding process with consolidated quality assurance was agreed for the writing domain in the SEA-PLM 2024 cycle. This double-coding procedure and quality assurance process has been standardized across the countries to ensure greater accuracy in coding writing responses..

xv) In 2024, the external quality assurance process for training, supervising, and capacity building the national teams, coding leaders, and markers was enhanced, in comparison to the 2019 cycle. As part of the new quality assurance process, coding discrepancies between coders were identified daily and adjudicated by a third party when necessary. Coders were provided with additional guidance to improve their practices and discrepancies recorded. The SEA-PLM Regional Secretariat commissioned this to ACER in 2022 as this revised approach required additional technical assistance resources further to treating the reading and mathematics tests.

xvi) The SEA-PLM 2024 Main Survey administration took place from April (Viet Nam) to December 2024 with the final country (Myanmar). Each country team was responsible for entering and

coding their own datasets, with technical support from ACER, following the completion of their own survey.

xvii) The following changes in the coding process were documented by the Secretariat:

2019	2024
• The capacity-building process included a regional one-week workshop for all participating countries, along with an additional national workshop in Malaysia for the country only(due to time difference between data collection of other countries and the number of languages applied in Malaysia). • Code leaders were trained in pre-coding operations, covering test framework content, coding guide application, and Q&A sessions on student sample responses from the field trial in each language. • A simplified coding scheme was used during the 2018 field trial and the 2019 Main Survey, with countries operating independently and receiving virtual assistance upon request.	• The capacity building process included ACER directly training test markers and providing hands-on practice and continuous monitoring. Each country also received support from 1–2 individual experts. • A double-blind coding procedure was introduced to enhance coding reliability and adjudication, fostering a significant learning curve as booklets were coded. • Under this approach, each student script is independently marked by two coders—without seeing each other's scores to ensure the independence of each score. • In cases of discrepancies, the coding leader intervened to adjudicate and provide targeted training to coders to minimize inconsistencies in subsequent booklets.

Additionally, the changes and further details were identified and presented by ACER to countries on April 9-10 April 2025 during an online webinar on " *SEA-PLM 2024 Main Survey technical updates and psychometric outcomes*". More detailed information is available in the SEA-PLM writing comparison and quality assurance PowerPoint presentation delivered at that webinar.

	2019	2024
Training	• 2.5 days. Country representatives (lead coders and/or national team leaders) to Bangkok training in English	• 5 days in country direct to markers. Interpreters used.
Code books and training scripts	• Translated into official language by national teams • Samples for all tasks were provided, in a range of languages	• Translated into official language by SEAMEO • Approx 5 samples for each task provided in the official language for training and practice for each country
QA techniques	• Practice sessions in Bangkok (and for 3 days in Malacca for Malaysia) • Lead markers train markers in their countries • Single marking of scripts	• Training in each country direct to markers from ACER experts • Double-blind marking of each script
Feedback	• Generally positive • Writing considered most difficult domain for training • Some concerns about logistics	• Excellent • Markers and lead markers "found it easy" after the complete training • Countries assisted with logistics where required
Monitoring	• Spot checking by lead markers • ACER helpline available	• Spot checking by lead markers and ACER staff • ACER staff available in person for 5 days and provided online help where needed

xviii) ACER shared further comments to review and comment on the variation between the 2019 and 2024 coding training and quality assurance processes commissioned to ACER in 2019 and 2024. The note is available in Annex 2.

xix) After the data collection, all responses were coded and collated in national datasets by countries. As a preliminary step before generating composite scores in writing, ACER conducted Intensive item-by-item analysis and construct validity checks to review the individual properties of each item, along with the time trend validity and reliability of constructs against psychometric modelling.

xx) The 2024 writing item analysis demonstrated strong psychometric performance overall, confirming the reliability and validity of the 2024 writing items. A list of items with limited national statistics parameters was identified and treated appropriately to control psychometric issues (exclusively score collapsing for 4 items only has been applied to all countries). Item-by-item statistics parameters for the writing domain are accessible for each country version.

xxi) Despite demonstrating acceptable item-level parameters in the 2024 assessment, initial psychometric comparisons between the 2019 and 2024 writing results revealed certain technical limitations that challenged the validity of robust time-trend comparisons on the writing scale. The 2024 data show unusually large increases in item facility (i.e. ease of items) across most participating countries, with some countries recording substantial, at times extreme, gains. When interpreted without consideration of these technical constraints, the results suggest unprecedented improvements, equivalent to an average proficiency gain of two to three years within a five-year period. Such rapid progress is highly atypical in the context of international large-scale assessments and raises questions about potential methodological inconsistencies. Notably, these significant increases are not observed uniformly across all countries. Further details on these findings are presented in the table below:

Table: Writing item facility difference 2024 vs 2019 (country are labelled randomly)

itmCode	Avg cnt	Country J	Country K	Country L	Country M	Country N	Country O
W15Y001A	7.3	22.8	-0.3	-7.0	8.2	0.7	19.0
W15Y001B	13.2	32.6	6.3	-2.2	23.4	0.9	18.2
W15Y001C	9.2	25.3	4.5	-8.1	16.7	-1.1	18.1
W15Y001D	13.4	36.1	3.2	-2.3	25.6	1.9	15.9
W15Y004A	10.7	20.5	5.8	1.6	13.2	5.1	17.9
W15Y004B	8.5	20.9	2.9	-3.0	11.1	7.3	12.0
W15Y004C	9.5	21.9	4.1	-4.6	12.0	5.1	18.6
W15Y004D		29.2	5.9	-4.7	19.8	6.1	23.7
W15Y004E	8.8	23.9	4.9	-9.7	13.5		11.5
W15Y010A	13.2	22.2	16.9	9.3	18.6	1.4	10.8
W15Y010B	13.1	19.9	12.6	5.0	24.4	3.8	12.7
W15Y010C	15.7	29.6	14.3	4.4	28.2	3.6	14.0
W15Y010D	15.3	32.6	15.9	4.9	18.4	1.2	19.0
W15Y014A	18.0	13.4	32.3	0.5	22.5	12.3	27.1
W15Y015A	13.9	46.3	-0.1	-1.2	46.0	-6.4	-1.3
W15Y015B	9.6	24.8	9.8	-2.5	20.2	-0.8	5.9
W15Y015C	14.0	28.4	11.9	2.8	32.3	4.8	4.0
W15Y015D	13.8	26.8	11.2	1.7	29.9	5.4	7.6
W15Y015E	10.4	28.1	5.2	-7.3	19.1		6.7
W15Y015F	8.4	27.5	6.2	-6.0	17.0	0.6	5.1
W15Y017A	15.8	32.9	14.5	-8.3	36.4	7.3	12.0
W15Y017C		33.0	15.1	-10.3	28.7	7.4	14.3
W15Y018A	11.5	30.4	13.7	-8.7	19.0	3.1	11.6
W15Y018B	15.1	37.5	16.8	-4.8	20.2	5.7	
W15Y018C	11.0	31.6	12.5	-14.6	20.1	2.6	13.6
W15Y020A	13.6	32.3	21.6	-2.7	20.7	0.1	9.5
W15Y020B	13.0	38.2	17.6	-11.6	20.0	2.0	11.9
W15Y020C	14.1	47.4	17.0	-15.1	25.3	3.6	6.0
W15Y020D	2.1	10.6		-17.4	22.7	5.3	-10.6
W15Y023A	9.9	17.7	13.5	1.7		10.1	6.2
W15Y023B	7.9	11.1	12.0	3.7	8.4	6.9	5.2
W15Y023C	9.5	22.1	2.9	2.5	8.7	13.9	7.1
Average	11.8	27.4	10.7	-3.6	21.0	4.0	11.4

xxii) A substantial part of the high differences observed could be reasonably explained by changes introduced in the more rigorous coding training and the increased quality assurance procedure (double coding).

xxiii) To ensure that the SEA-PLM 2019 and 2024 writing scaling accurately measures a similar range of writing ability, the item-by-item and domain psychometric characteristics were generated and compared between the two assessment rounds for each country and all countries. The 2024 writing domains show a high mismatch at all distribution points, and the overall ability

distribution does not overlay in acceptable range. It affects the reliability of time trend comparisons for this specific domain. These limitations do not impact the validity of the reading and mathematics domains or the comparability of reading and mathematics scores across the two cycles.

xxiv) The psychometric analysis of the SEA-PLM 2024 writing results confirms acceptable item-level parameters and overall consistency (high test reliability and high item validity) which supports the standalone reporting of 2024 outcomes. However, the analysis also revealed a significant increase of item facility in 2024, highlighting important technical limitations for directly comparing the constructs between the 2019 and 2024 cycles. These limitations affect the comparability of writing scales, proficiency bands, individual and aggregated scores and indicators - particularly in relation to the results published as part of the SEA-PLM 2019 study.

xxv) Some experiences of other comparative large-scale assessment in measuring writing and literacy skills are presented in Annex 1.

Implications and the way forward

xxvi) As the 2024 writing domain shows high item and test validity, individual student ability continues to be generated using a three-dimensional conditioning model (reading, mathematics and writing) to estimate the SEA-PLM 2024 reading, mathematics and writing individual scores.

xxvii) **However, the 2019 writing score and proficiency scales are no longer comparable over time although the 2019 writing score remains valid in the 2019 context only.**

xxviii) New scoring scale settings (such as a new distribution point range and standard deviation points range) will be established and will be applied for the entire region in the 2024 cycle and in future rounds of SEA-PLM. New proficiency band cut scores will be calculated, and a new writing proficiency scale with refreshed proficiency descriptors will be considered. This operation is underway and will mean that **scores and competencies between the 2019 and 2024 results cannot be compared.** This will impact the release and utilisation of the writing results at regional and national levels, while ensuring the reliability of the 2024 results in all three domains.

xxix) **Countries will receive the new 2024 final student dataset with the new writing score scheme and the new proficiency** at the time of the release of the 2024 report. Countries will receive more instructions at the time when the writing datasets and results will be considered ready for official communication in the public domain. Countries are encouraged to consider reporting 2024 writing separately from 2024 reading and math results.

xxx) **The public release of the writing domain in the national and regional database and preliminary and official reports of results prepared by the SEA-PLM Regional Secretariat (regional preliminary note, country snapshot, main regional report) this year will be delayed from the reading and mathematics domains.** The data and results reporting of the writing domain will be treated separately from the reading and mathematics datasets and results prepared and released by the SEA-PLM Regional Secretariat, very likely in a regional secondary

report of results and in a new version of the regional student dataset to be published next year in the public domain.

xxxi) **A standalone technical note to document** the reading, mathematics and writing scaling and equating process and outcomes will be released to countries and to the public prior to the official launch of the SEA-PLM results to maintain adherence and transparency. More information on technical limitations will be properly reported and communicated.

xxxii) **The psychometric analysis of the reading and mathematics domains has confirmed the high validity of time trend comparisons between 2019 and 2024 scores and scales for the reading and for the mathematics domains.** This ensures the timely release of reliable comparative reading and mathematics indicators in the regional database and upcoming reports, in line with the committed timeline to release the regional results in November 2025..

xxxiii) In the longer term, the SEA-PLM Secretariat proposes to discuss making the writing domain an optional testing domain with participation and interest, particularly given the additional resourcing required.

xxxiv) In the scenario that the writing domain is withdrawn - either on a country-by-country basis or at the regional-level - from the list of compulsory testing domains, there will be no impact on the precision of estimates in reading and mathematics. To explore this, the SEA-PLM Regional Secretariat commissioned a study by ACER, which confirmed the feasibility of continuing with only two assessment domains, should participating countries collectively agree to this adjustment in the future cycles.

Annex 1 - Writing in other ILSA

i) Testing writing as a standalone domain and reporting writing proficiency on a unique scale is very singular in comparative large-scale assessment. For many experts, measuring writing as a single construct and in a comparative way is seen as a complex exercise largely dependent on the level of resources, capacity and mechanisms in place to ensure strong quality assurance within and between countries, language versions (9 in 2019, 11 in 2024) and over time.

ii) As per its nature and in comparison, to other domains assessed in the SEA-PLM surveys, the writing domain is highly vulnerable to validity and reliability measurement risks.

iii) Despite being one critical strand of literacy skills, assessing writing is seen by the internal community of experts as technically complex, notably in multilingual environments. In PISA, writing is measured through the reading comprehension scale and embedded in the literacy scale, notably the measurement of high-level cognitive tasks.

Box 2.7. The status of writing skills in the PISA 2018 reading literacy assessment

Readers are often required to write comments, explanations or essays in response to questions, and they might choose to make notes, outlines and summaries, or simply write down their thoughts and reflections about texts, while achieving their reading goals. They also routinely engage in written communication with others (e.g. teachers, fellow students or acquaintances) for educational reasons (e.g. to e-mail an assignment to a teacher) or for social reasons (e.g. to chat with peers about text or in other school literacy contexts). The PISA 2018 reading framework considers writing to be an important correlate of reading literacy. However, test design and administration constraints prohibit the inclusion of an assessment of writing skills, where writing is in part defined as the quality and organization of the production. However, a significant proportion of test items require readers to articulate their thinking into written answers. Thus, the assessment of reading skills also draws on readers' ability to communicate their understanding in writing, although such aspects as spelling, quality of writing and organization are not measured in PISA.

iv) Writing domain is also measured (in Spanish and Portuguese exclusively) as a standalone learning domain besides reading, mathematics and science in the Estudio Regional Comparativo y Explicativo, the regional Grade 3 and Grade 6 comparative assessments for Latin Americano countries. Based on official documentation, the writing results have been separated from the main results report and released in 2022, about 2 years after the main data collection. The initial results were made publicly available in November 2021. Writing results are reported country by country against scores rubrics rather than on a single comparative scale due to several technical limitation and risks.

'The results of this study are not comparable across countries or with the previous TERCE study, given that the ERCE incorporated new guidelines to include text genres not considered in the

previous cycle. This innovation responded to countries' interest in including new text genres to accommodate the curricular changes that have occurred since TERCE. Unlike the anchoring of items in other (multiple-choice) tests, the design of this test's administration does not allow for comparability of results across guidelines. Furthermore, the rubrics were adjusted and incorporated elements specific to the new text genres. Additionally, as noted in Chapter 2, when the percentage of consistency between coders is not the required one, even for the continuity rubrics between ERCE and TERCE, the comparison of the percentages of students in each category of the rubric with those observed in the previous study presents degrees of precision below what is recommended for making and reporting comparisons.'

To carry out the coding, each country trained a group of evaluators in both the procedure and the understanding of the rubrics, based on the technical guidelines provided by the LLECE (Spanish Institute of Educational, Technical, and Cultural Studies). After the training, each student's written output was coded, assigning performance to a rubric category for each indicator assessed. An essential input for coding responses is the coding manual, adapted from the pilot application, which details the rubrics, sample responses for each performance category, and considerations for coding. To monitor the consistent application of the rubrics, agreement between coders is verified by measuring the proportion of identical responses categorized in the same way by two different coders. Each country was required to double-code approximately 30% of the responses to obtain the percentage of consistency between coders.

Each country obtained consistency percentages for each indicator evaluated. Annex 1 details the percentages obtained in the coding carried out. If the consistency percentage obtained for an indicator is less than 70%, this will be noted in the results presented. In these cases, the data should be interpreted with great caution, given that the agreement between coders is lower than expected. 1 To carry out the coding process, the disciplinary teams from the different countries participating in the study attended a training session in Panama City in June 2019, led by the LLECE and the implementing partner."

Annex 2

This note reflects the views and opinions of ACER, April 2025.s

Comparison Report: SEA-PLM Writing Assessment Quality Assurance (2019 vs 2024)

1. Introduction

This report presents a comparison between the 2019 and 2024 SEA-PLM writing assessments, with a focus on the implementation and evolution of quality assurance procedures. It outlines the assessment frameworks, training processes, and strategies implemented to ensure consistency and reliability in assessing student writing across participating countries.

2. Quality assurance in 2019

In 2019, the SEA-PLM faced the challenge of implementing a valid and reliable writing assessment across multiple countries with varied languages and educational contexts. The main goals were to make tasks, marking standards, and coder training consistent. Key steps included:

- Translation and back-translation of writing tasks and coding guides.
- Centralized marker training conducted by ACER in Bangkok (2018) for all countries.
- Individualised training session for Malaysia in Malacca, working with English, Malay, Tamil and Chinese.
- Use of student writing samples from all countries during training to ensure consistency.
- Implementation of a 4-day regional coder workshop (for all domains) covering coding guide familiarization, exemplar review, and simulated training sessions.

3. Participant feedback from 2019 workshop

Participants found sessions on coding guides, exemplar discussions, and simulation exercises particularly useful. They emphasized the need for continued technical support, more international examples, and regular communication for sustained capacity building.

4. Enhancements in 2024

In 2024, the SEA-PLM quality assurance system was significantly strengthened with more direct and localized training and monitoring. Improvements included:

- Direct 5-day workshops in each country led by ACER trainers, bypassing 'train the trainer' models.
- Marker training conducted in local languages with increased and detailed task-based practice.
- Increased monitoring through spot-checking, daily 'refresher discussions', and double-blind coding.
- Enhanced security measures for script handling and data protection.

5. Double-blind coding and reconciliation

Double-blind coding was introduced to improve consistency. It is the international 'gold standard' for assessing student writing. Each script was independently marked by two coders, with discrepancies resolved through clear procedures. Differences of three or more points triggered a third, final marking by the lead coder to ensure fairness and accuracy.

6. Summary of comparison

The transition from the 2019 to 2024 SEA-PLM writing assessments shows a notable shift from centralized training to localized, task-specific support, enhancing both the reliability and fairness of the marking process. These upgrades represent a maturing quality assurance system that builds on regional collaboration and rigorous technical guidance.

7. Summary table: 2019 vs 2024 QA procedures

The table below summarizes the key differences between the 2019 and 2024 quality assurance procedures implemented in the SEA-PLM writing assessment.

Aspect	2019 QA procedures	2024 QA procedures
Marker training	Train-the-trainer approach with one regional workshop in Bangkok.	Direct 5-day country-level training by ACER in local languages, in each country.
Training content	Focus on sample review, coding guide, and simulated training.	Detailed task-based training and guided practice with feedback.
Language of training	Primarily English with translated materials.	Conducted in national languages where needed.
QA methods during coding	Standard practice with limited cross-checking.	Spot-checking, refresher discussions, double-blind coding.
Double coding	Not systematically implemented.	Each script is marked independently by two coders.
Discrepancy resolution	Handled case-by-case if markers requested assistance.	Clear reconciliation procedures with third marking by lead coder.
Security	Basic procedures in place.	Strict protocols including restricted room access and confidentiality agreements.

7. Recommendation

The SEA-PLM writing assessment has evolved significantly between 2019 and 2024, especially in the area of quality assurance. Through a shift from centralized to localized training, and by incorporating systematic monitoring and reconciliation practices, SEA-PLM has enhanced the fairness, accuracy, and reliability of its writing assessment process.

ACER recommends that this rigorous quality assurance process be maintained for future iterations of SEA PLM writing assessments.