

Comparing Writing Skills in Different Languages Using the Same Scale

Louise Courtney

EALTA June 2019

Presentation

Report on the field trial results of an assessment of **L1 writing** across a range of languages in South East Asia.

Overview

About ACER

Timeline for project

Framework design

Test development

Field Trial results

Questions / discussion

I. About ACER

- Independent, not-for-profit research organisation
- Established in 1930
- 400+ staff
 - Australia
 - India
 - Indonesia
 - United Arab Emirates
 - United Kingdom

Research...

- Early childhood education
- School education
- Higher education
- Vocational, adult and workplace education
- Indigenous education
- Education and development

➤ research.acer.edu.au

➤ rd.acer.org

and assessment

Assessment specialisation:

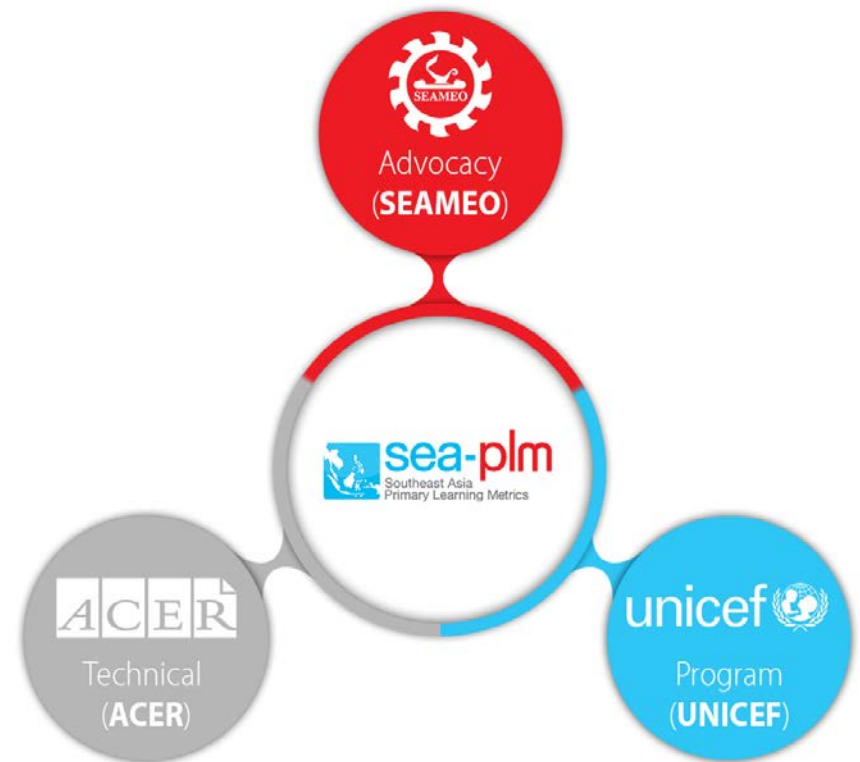
- Monitoring and benchmarking
- National and international
- Admission and selection
- Tailored assessment

2. The SEA-PLM project

What is SEA-PLM?

South East Asia

Primary Learning Metric



SEA-PLM in brief:

Which curriculum?

references the curricula of all countries of the region

Which values?

respectful of Southeast Asian values and context

What level?

measures learning outcomes for Grade 5 students

What subjects?

assesses Reading, Writing, Mathematics and Global Citizenship

What are the countries and languages?

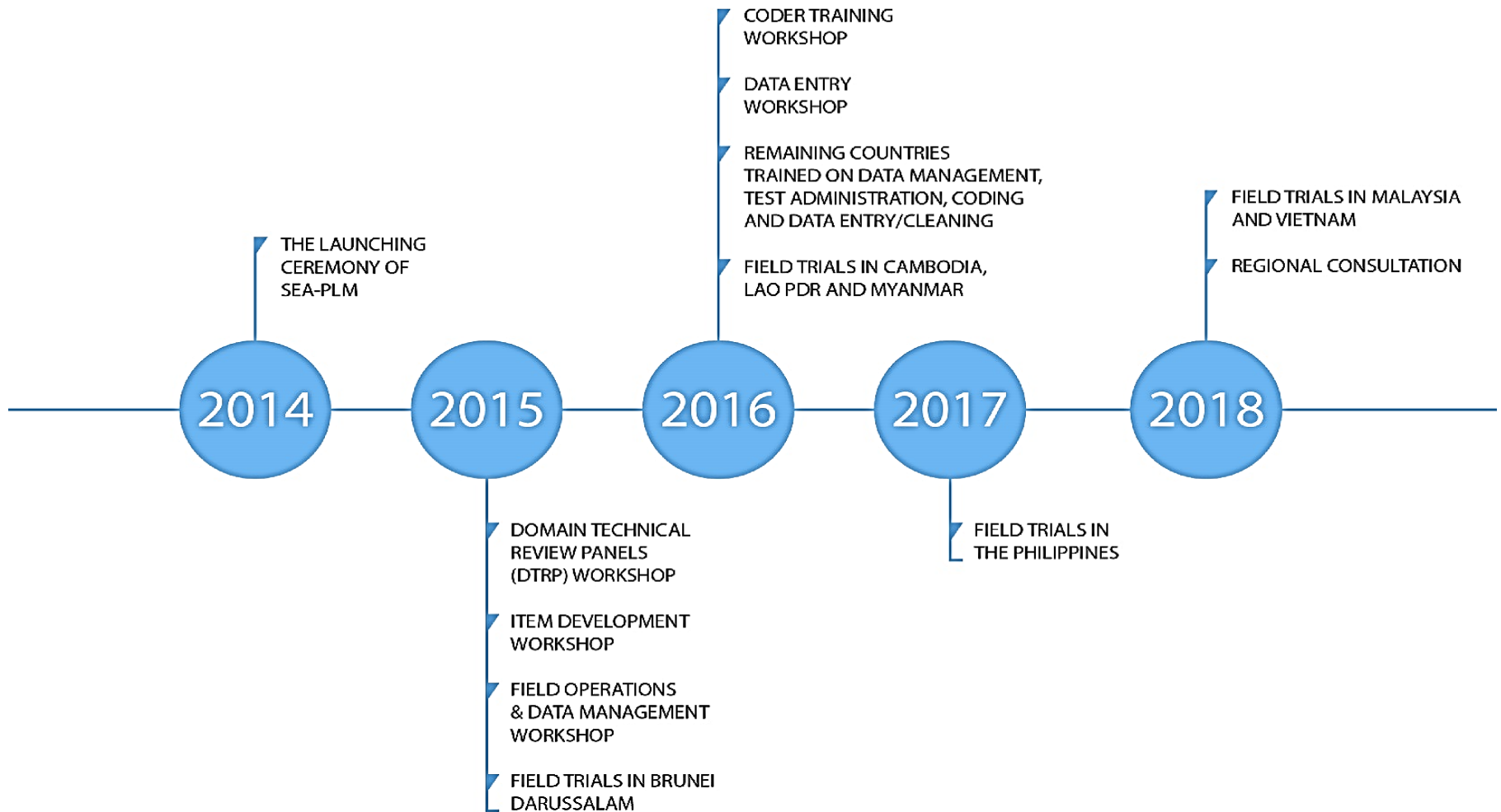
- Vietnamese - Vietnamese
- Laos - Lao
- Cambodia - Khmer
- Malaysia - Malay, Mandarin and Tamil
- Brunei – English*
- The Philippines - English
- Myanmar - Myanmar language/Burmese.

Where are these countries relative to one another?



Timeline so far

Timeline of SEA-PLM Field Trial



Developing the framework...

- What is writing?
- What are suitable **tasks** that would apply to writers of all languages?
- What are the main features of writing, **regardless** of language?
- What aspects of writing might **not** be able to be assessed in a variety of languages?

What is writing?

The working definition of 'writing' for SEA-PLM:

Writing literacy is constructing meaning by generating a range of written texts to express oneself and communicate with others, in order to meet personal, societal, economic and civic needs.

What are suitable **tasks** that would apply to writers of all languages?

- Labelling
- Narrative
- Descriptive
- Persuasive
- Instructional
- Transactional

Contexts

- Personal
- Societal (local)
- Wider world (civic and economic)

What do we consider to be the main features of writing, regardless of language?

- Generate ideas
- Control structure and organisation
- Manage coherence
- Use vocabulary

What aspects of writing might NOT be able to be assessed across a variety of languages?

- Control of syntax and grammar
- Other language specific features

e.g. spelling, character formation, punctuation, register

How is writing assessed in SEA PLM?

What do students need to do?

write responses for all tasks...

basic labelling

sentence level

extended task

21 writing tasks in total
plus 14 pre-writing tasks

Shorter and longer tasks organized into 6 clusters

18 booklets provided in a rotating design

Test development process

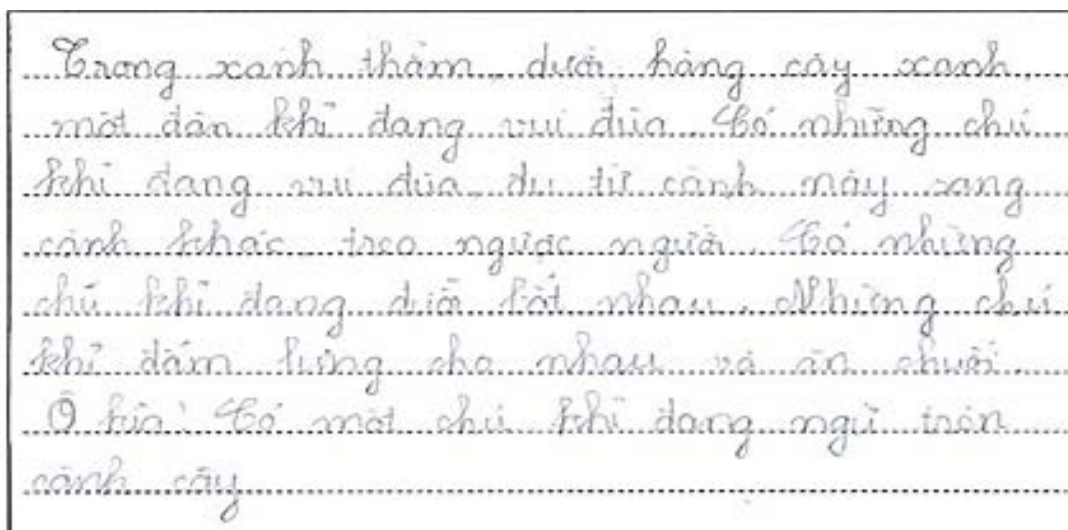
- 2015 all tasks developed in English by ACER literacy experts
- 2016 approved by each participating country
- 2016 translated into each target language
- 2016 coder training workshops for local teachers/assessors by ACER
- 2017 translations verified by advisory committee in each country

Scoring guide

criterion	score	description
Generate ideas: relevance W15Y016A	0	no relevant or comprehensible ideas
	1	one relevant idea OR most ideas are not relevant to the task of giving advice (ideas refer to spending money, but not presented as giving advice)
	2	relevant, plausible alternative ideas
language of advice W15Y016B	0	no attempt to use the language of advice
	1	some attempt at advice but weak (must, can)
	2	clearly offers advice (should, could, might, I suggest, I advise)
vocabulary W15Y016C	0	no content words relevant to the task
	1	(limited range): basic / simple vocabulary
	2	beyond basic

critrion	score	description
overall account of the picture W15Y011A	0	evidence of a response, but no relevant information is included
	1	very little content; describes no more than 2 elements
	2	focuses on isolated features or elements but has adequate ideas (3 or more?)
	3	gives a good sense of the whole picture and includes some detail without being a coherent text (disjointed or odd sentences).
	4	Effectively conveys a sense of the whole picture, even if the ideas are expressed concisely. The text is a cohesive and complete unit.
grammar W15Y011B	0	Sentences are incomplete or all contain significant errors
	1	at least one simple correct sentence;
	2	Generally there are at least 2 grammatically correct sentences, but repetitive or simple in structure.
	3	variety in sentence structure, but with errors
	4	Generally the sentences are accurate and varied in form
vocabulary W15Y011C	0	no relevant content words
	1	(limited range): basic vocabulary, repetitive, inadequate to describe the picture well
	2	adequate to describe the picture
	3	good range of vocabulary gives good sense of detail of the content the student presents: good range of verbs and nouns

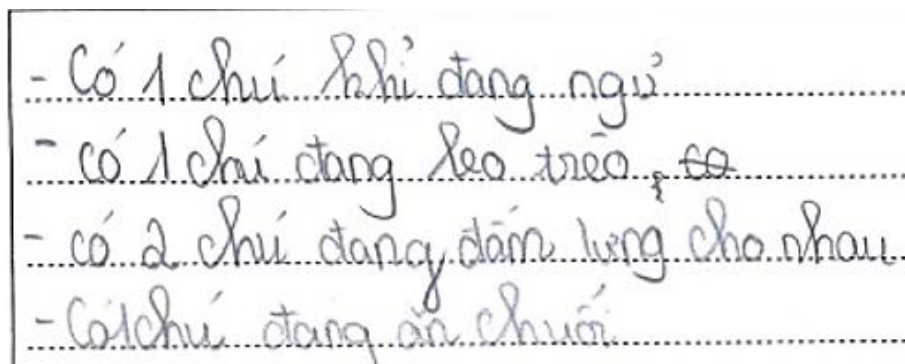
Sample script : Monkeys



under the green trees, there are monkeys playing. There are monkeys playing from branch to branch and climbing backwards. There are monkeys chasing each other. The monkeys backed each other. The monkeys bump their backs together and eat bananas. There is a sleeping monkey on the tree

Overall account	3	
Grammar	2	
Vocabulary	2	

Sample script : Monkeys



A monkey is sleeping.

A monkey is climbing.

Two monkeys are massaging their backs on each other.

A monkey is eating a banana.

Overall account	2	
Grammar	2	
Vocabulary	1	

Generate ideas

Criterion	Score	Description
Development of narrative (elaboration of ideas)	0	Evidence of a response, but no relevant information is included
	1	Fragments: few ideas or no complete ideas
	2	Limited writing related to the picture
	3	Simple writing related to the picture; limited detail
	4	Detailed writing with many relevant ideas

5. Field Trial

- Field trials in participating countries between 2014 and 2018
- 15 392 students in 277 primary schools completed assessments
- 14 479 parents and 2558 teachers completed surveys

Rater training

Every criterion for every task

Exemplified with sample student responses from each language

101 exemplars for each language.

Rater training meetings to standardise marking across the field trial, and across all participating countries.

Test *administrator* training on a large scale.



People's Democratic Republic of Laos



Analysis of field trial results

Analysis was conducted according to the Rasch model by ACER's psychometric division.

Differential Item Functioning (DIF) was examined to explore how the items functioned in different languages/contexts.

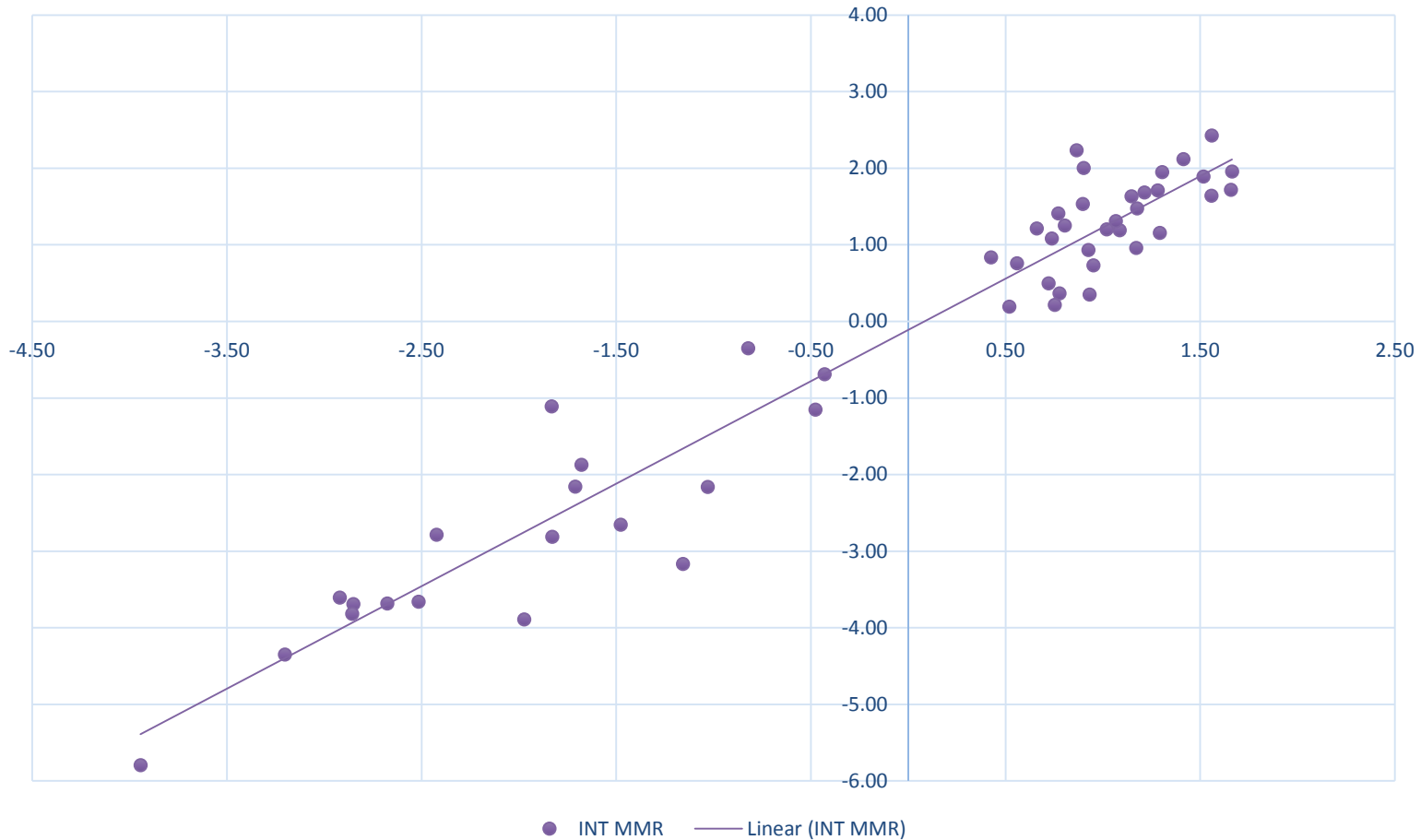
Hypothesis:

that the criteria anticipated to be language-specific will show more variation in item functioning between languages than other criteria.

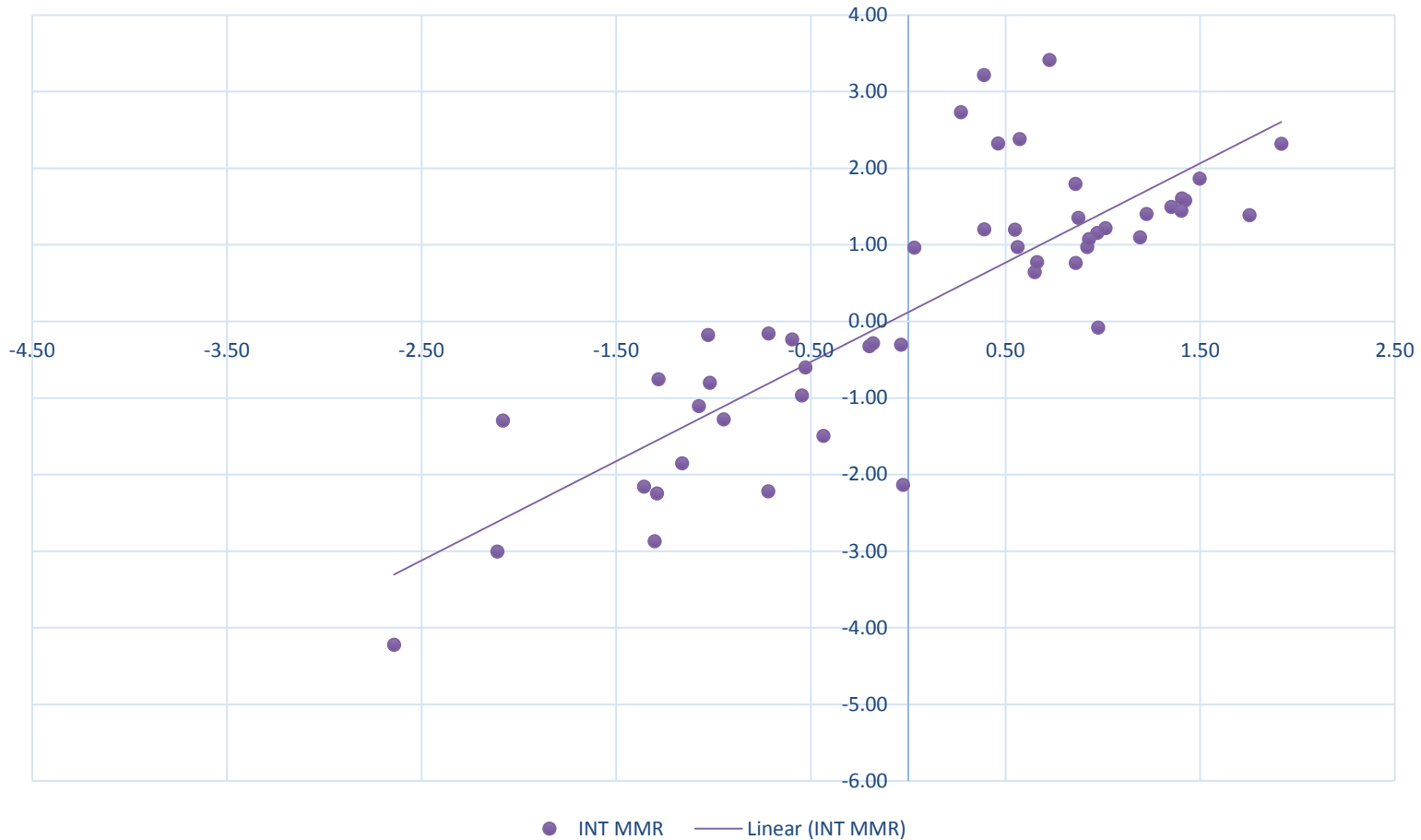
Definitions

- mean of the category thresholds as the ‘mean item location’ for each criterion
- ‘International’ data comprises all languages, including the language of interest, and represents the average difficulty of each criterion, for each writing task, across all the countries involved.
- preliminary indicators

Results for MMR Cross Language

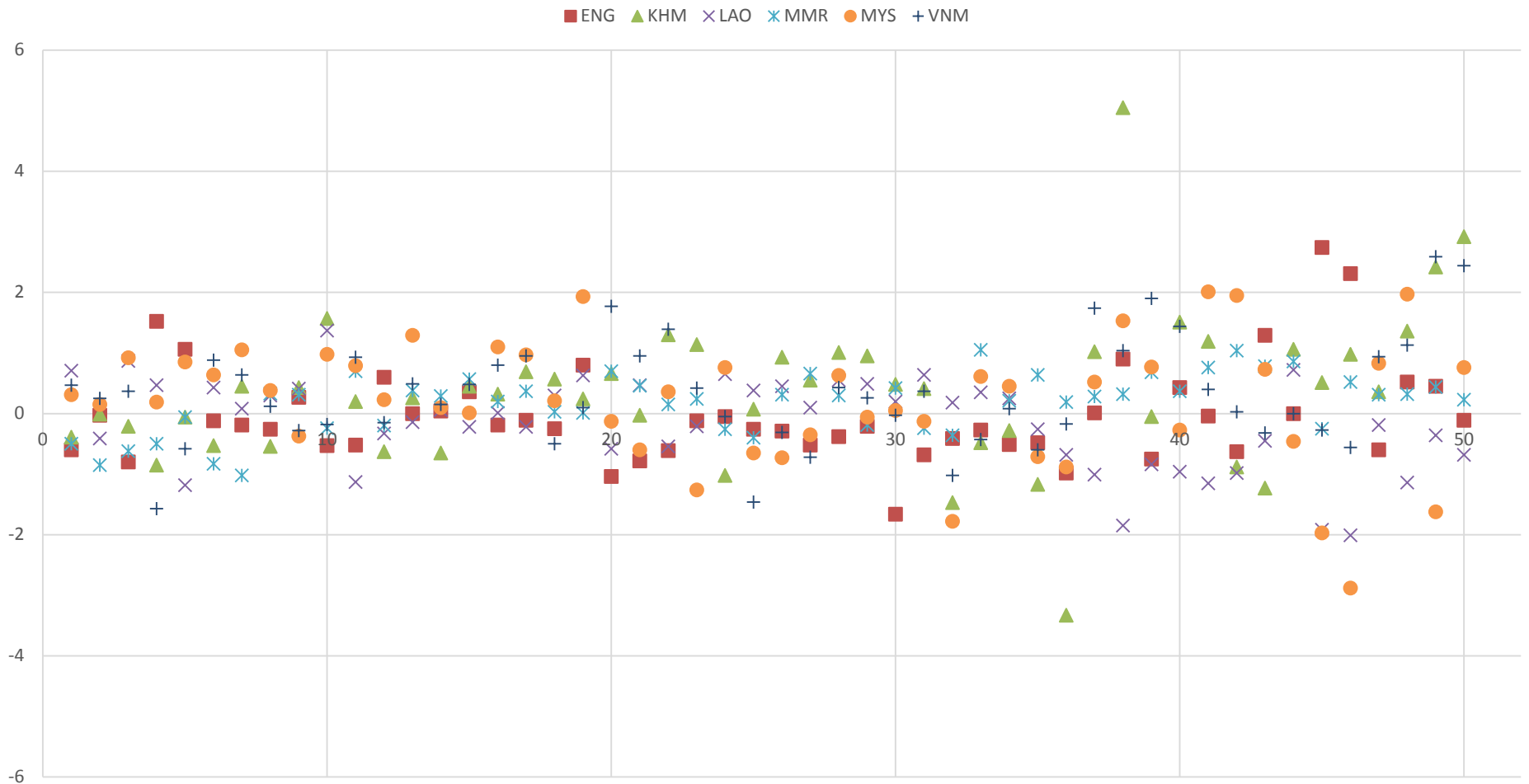


Results for MMR Language Specific



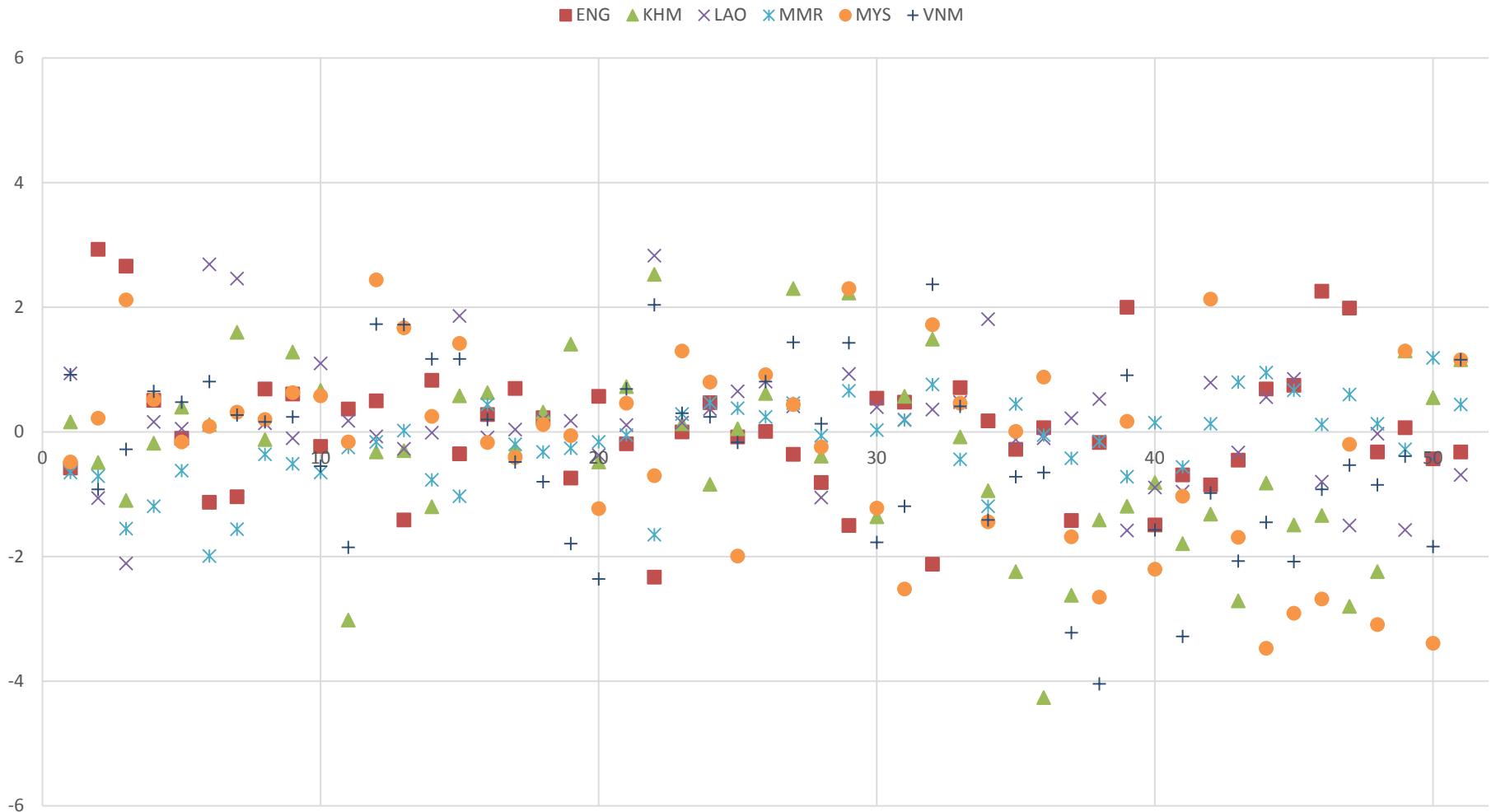
Results/data (15,392 cases)

CRITERIA HYPOTHESISED TO BE APPLICABLE ACROSS LANGUAGES



Language specific criteria

CRITERIA HYPOTHESISED TO BE LANGUAGE-SPECIFIC



The features of writing, across languages and text types

Process	Application by language	Application by text type
Generate ideas	Apply across languages	Vary by text type
Control structure	Apply across languages	Vary by text type
Manage coherence	Apply across languages	Apply across text types
Use vocabulary	Apply across languages	Apply across text types
Control syntax and grammar	May vary by language	Apply across text types
Other language-specific features	May vary by language	Apply across text types

Conclusion

A single instrument for use in multiple countries with diverse languages.

Field trial – a set of robust, valid and culturally sensitive instruments for use in the main survey.

Main survey being implemented now.

Sampling for each country is a minimum of 150 schools and 4000 students.

Thank you

Read the full report:

www.seapl.m.org.

www.acer.org