



Evaluating the validity and comparability of PIRLS in South Africa

Heather Leigh Kayton
heather.kayton@bsg.ox.ac.uk



The reading crisis in South Africa



South Africa has 12 official languages

All 12 are used as media of instruction in early primary school

Most South African children are unable to read with understanding in any language in primary school

A system built on racism and inequality

The current educational ecosystem in South Africa is built on a relatively recent legacy of segregatory, racist policies and structures.

“There is no place for [the black community] in the European community above the level of certain forms of labour ... What is the use of teaching the [black] child mathematics when it cannot use it in practice? That is quite absurd. Education must train people in accordance with their opportunities in life, according to the sphere in which they live.”

- Hendrik Verwoed, 1950
(former apartheid prime minister)

Although apartheid is now over, and there are current structures and policies that are attempting to address these atrocities, there is still a long way to go.

What evidence do we have about reading in primary schools?

Annual National
Assessments
(2012-2015)

SEACMEQ

PIRLS

Systemic
evaluations

What evidence do we have about reading in primary schools?

Annual National
Assessments
(2012-2015)

SEACMEQ

PIRLS

Systemic
evaluations

What is PIRLS?

- International large-scale reading assessment
- Conducted every 5 years
- Tests Grade 4s (~10 years)
- 12 reading passages
 - Each student answers items related to 2 of those passages
- In 2016:
 - 300,000 students
 - > 50 countries
 - 42 language versions

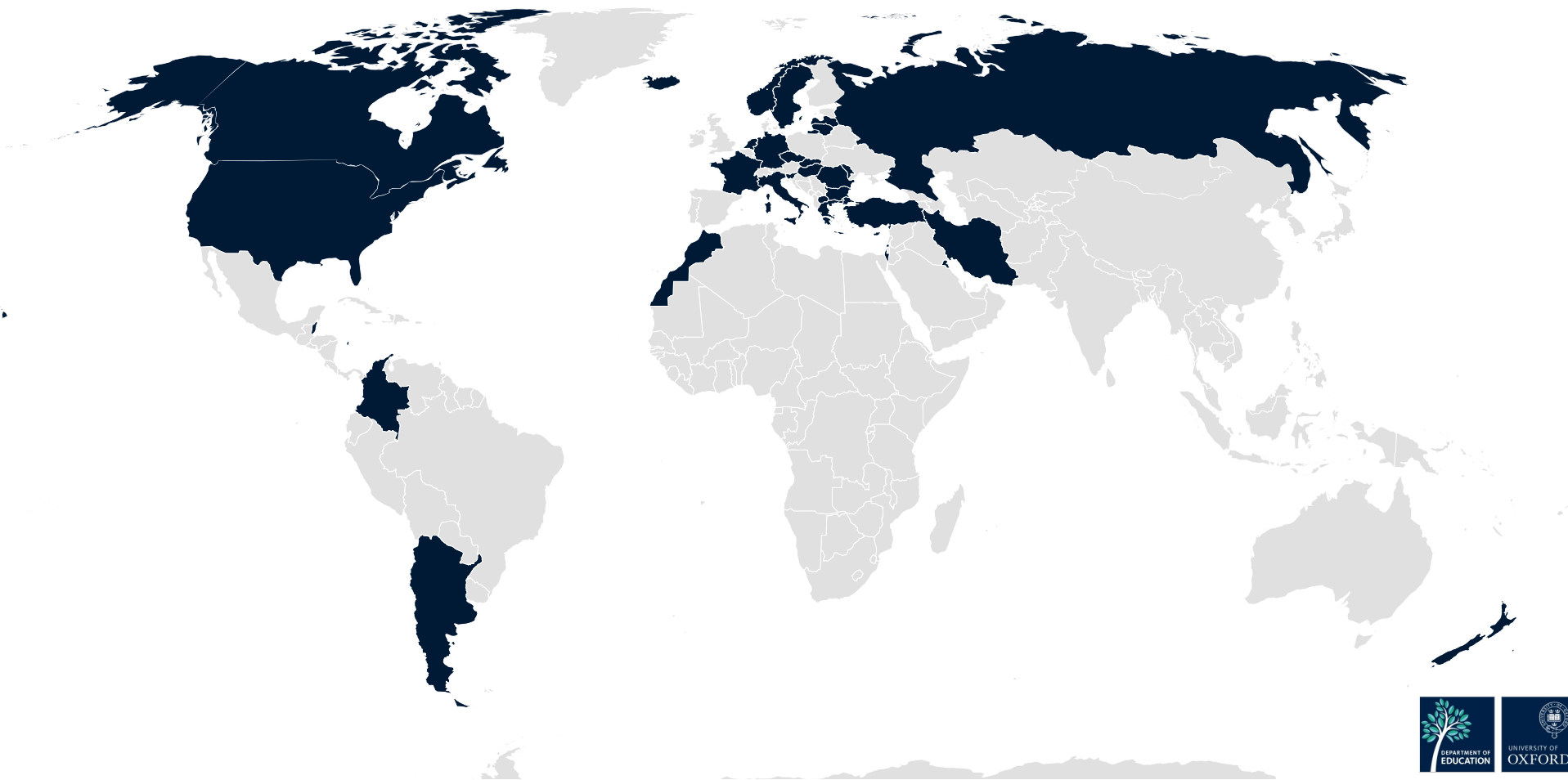
The image shows a sample PIRLS 2016 reading passage and questions. The passage is titled "Brave Charlotte" by Anu Stohner, illustrated by Henrike Wilson. The text describes a sheep named Charlotte who is different from the others because she is leaping around instead of standing shyly. The illustration shows a sheep leaping in the air above a group of sheep on a hillside. The questions are:

1. Who is Jack?
2. What did Jack try to do with Charlotte?

The page number 142 is visible in the bottom left corner, and 143 is visible in the bottom right corner. The PIRLS 2016 logo is in the top left corner. The text "TIMSS & PIRLS International Study Group" and "PIRLS LITERACY 2016 SAMPLE PASSAGES, QUESTIONS, AND SCORING GUIDES" are visible at the bottom of the page.

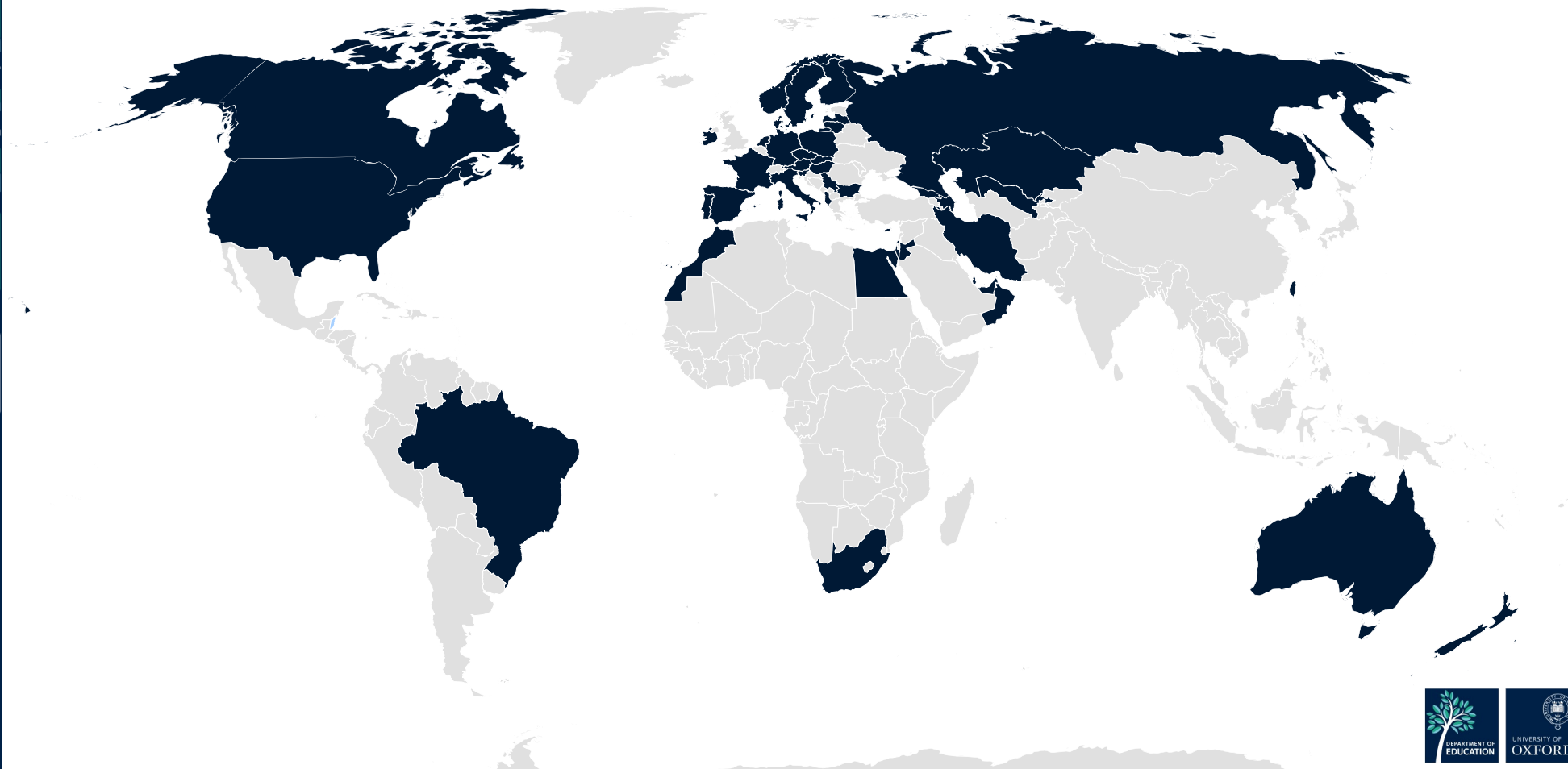
Widening participation in PIRLS

PIRLS Participation in 2001



Widening participation in PIRLS

PIRLS Participation in 2021



Consequences of widening participation

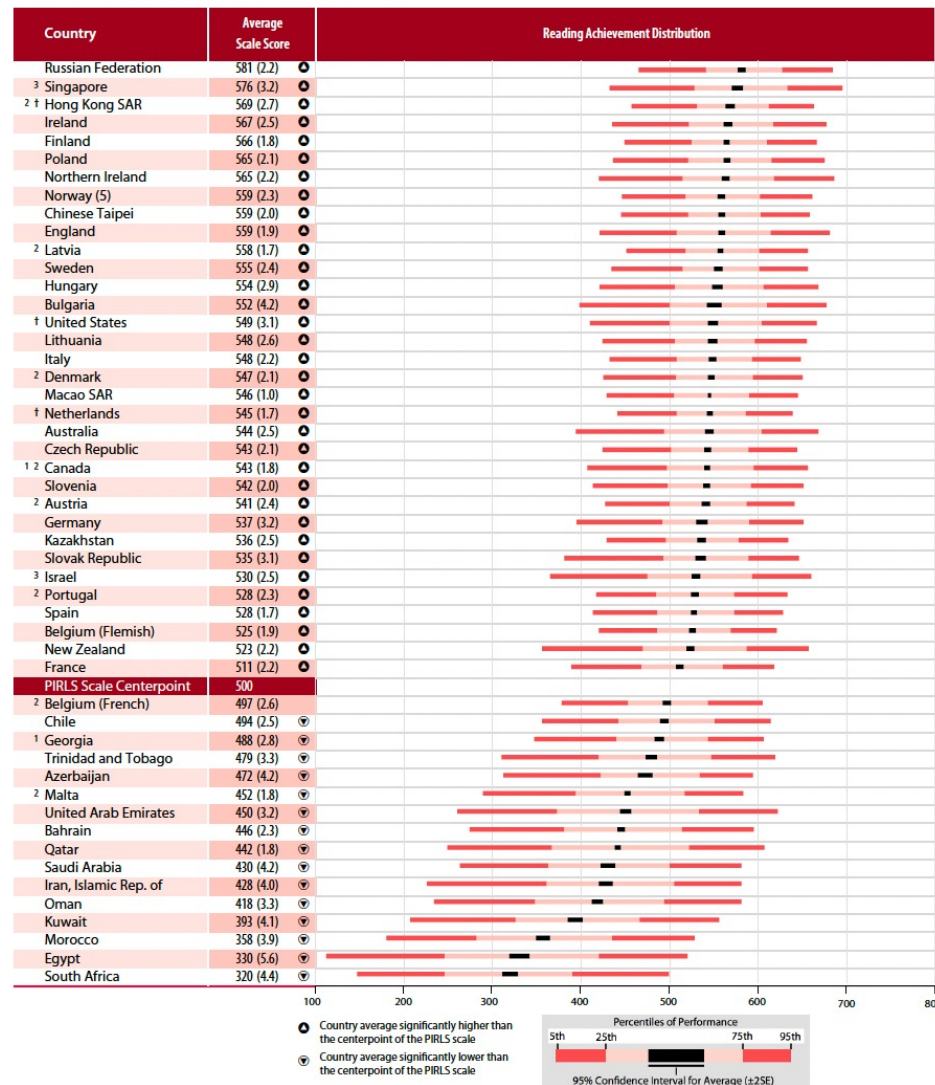
- Widening participation means wider ranges of student ability
- Testing agencies have extended and tailored their assessments to suit these needs
- Common changes:
 - ‘Easier’ versions of ILSAs
 - Adaptive testing
 - Adaptable contextual questionnaire items

However, the question of whether these adapted or easier versions are still able to suit the needs and motivations of participating countries has not yet been fully demonstrated.

Key issues for South Africa in PIRLS

- Lowest performer: Ranked lowest in every cycle of participation
- Average score 320 points is well below PIRLS centrepoint
- Very wide achievement gap, ranging from 5 to 710 points

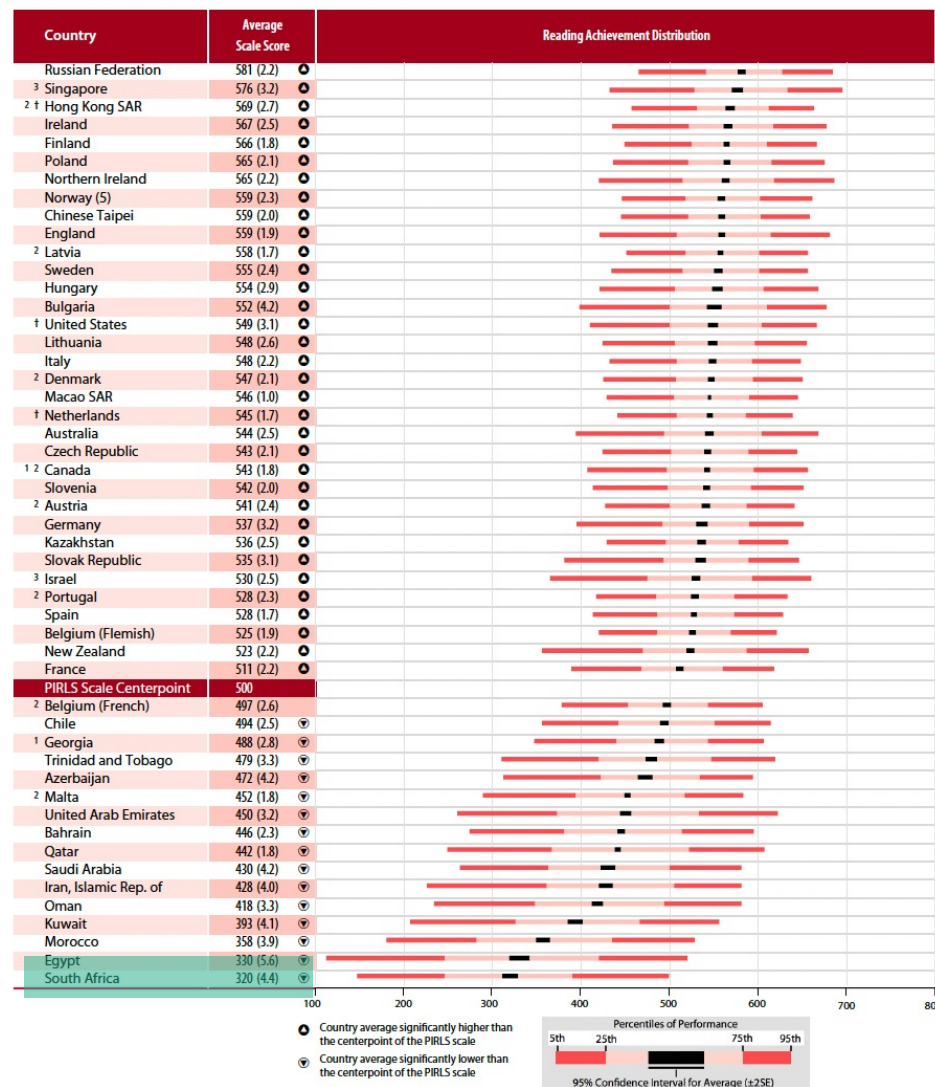
Exhibit 1.1: Distribution of Reading Achievement



Key issues for South Africa in PIRLS

- Lowest performer: Ranked lowest in every cycle of participation
- Average score 320 points is well below PIRLS centrepoint
- Very wide achievement gap, ranging from 5 to 710 points

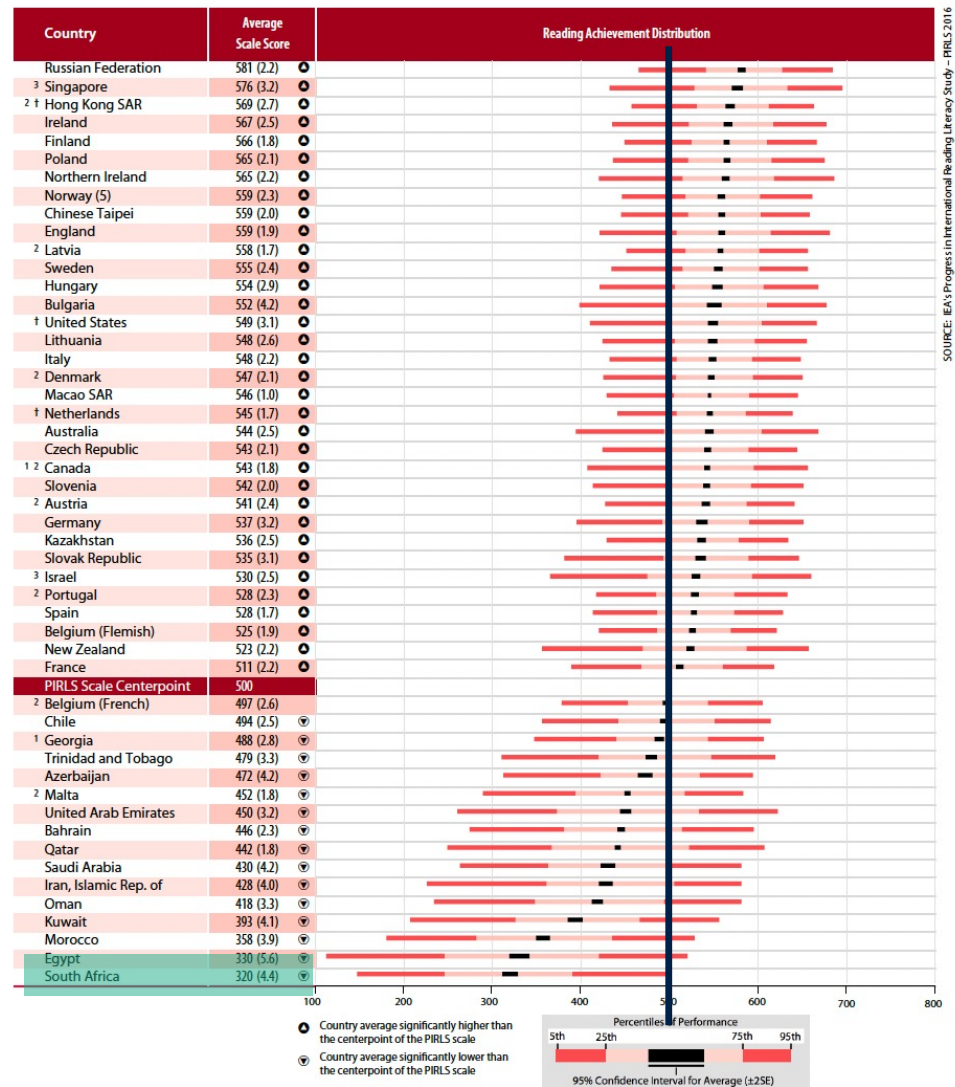
Exhibit 1.1: Distribution of Reading Achievement



Key issues for South Africa in PIRLS

- Lowest performer: Ranked lowest in every cycle of participation
- Average score 320 points is well below PIRLS centrepoint
- Very wide achievement gap, ranging from 5 to 710 points

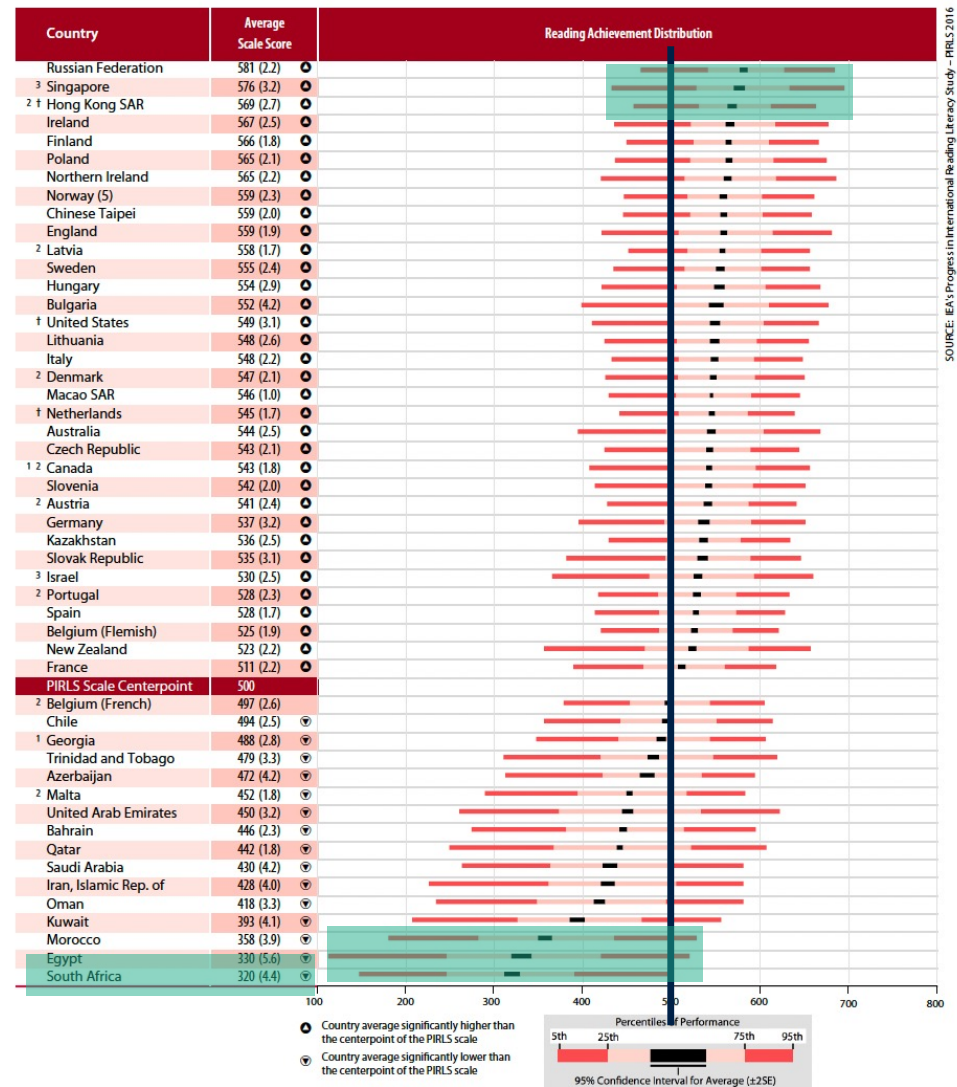
Exhibit 1.1: Distribution of Reading Achievement



Key issues for South Africa in PIRLS

- Lowest performer: Ranked lowest in every cycle of participation
- Average score 320 points is well below PIRLS centrepoint
- Very wide achievement gap, ranging from 5 to 710 points

Exhibit 1.1: Distribution of Reading Achievement



Overview of PIRLS results in South Africa

Cycle	Number of students	Average reading score	% did not reach lowest benchmark	Gap between highest and lowest performing language groups
2006	16,073	253 ★	87%	-
2011	15,744	461 (original) 323 (IEA rescaled in 2016) 295 (SA disputed score) 278 (IEA unpublished reanalysis)	29%	137 points (Afrikaans vs Sepedi)
2016	12,810	320	78%	96 points (English vs Sepedi)
2021	12,426	288 ★	81%	176 points (Afrikaans vs Setswana)

★ “cross-language comparisons not available as reading achievement for African language groups too low for estimation”

★ “reservations about reliability because the percentage of students with achievement too low for estimation exceeds 25%”

The status of PIRLS in South Africa

- South Africa has participated since 2006, **lowest achievement every cycle**
- Currently **one of the only primary school reading assessments conducted in 11 official languages at a national scale**
- High media attention
- PIRLS is used to **guide policy reform and reading interventions**

South Africa's 10 year-olds are struggling to read – it can be fixed

Published: May 22, 2023 3:53pm BST

IN NUMBERS | SA produces one of worst global reading results among over 50 countries

Pirls results a blaring fire alarm for South African society



By Mmusi Maimane [Follow](#)

22 May 2023

SA's children have lost a decade of reading progress, study shows

International study confirms that eight out of ten grade 4 learners still can't read for meaning

Some impacts PIRLS has had in South Africa

The performance in PIRLS has led South African policymakers to prioritise reading in the early grades, and in particular prioritising the teaching of reading in African languages

- National development goal 2030
 - Top five goal: “Every **10-year-old** child will be able to **read for meaning**” – President Ramaphosa (SONA, 2019)
 - Previous goal (2010): “Increase the quality of education so that all children have at least two years of preschool education and all children in **Grade 3** can **read and write**”
- National reading policy
 - “the PIRLS Reading Literacy Framework would be used as the basis for a shared national framework for understanding reading comprehension” Motshekga 2021 (former Minister of Basic Education)



Reading crisis

Influential test

Complex
multilingualism

Extreme
inequality

The research study

What can be learned from PIRLS Literacy 2016 about the reading achievement of South African students across different languages?

CONTEXTUAL ANALYSIS

Paper 1: Language Matters
Exploring how contextual and linguistic factors influence reading achievement across language groups

PSYCHOMETRIC ANALYSIS

Paper 2: Fit for Purpose
Evaluating the psychometric properties and targeting of PIRLS across language versions

Paper 3: DIF Happens
Examining comparability and differential item functioning across language versions

The research study

What can be learned from PIRLS Literacy 2016 about the reading achievement of South African students across different languages?

CONTEXTUAL ANALYSIS

Paper 1: Language Matters
Exploring how contextual and linguistic factors influence reading achievement across language groups

PSYCHOMETRIC ANALYSIS

Paper 2: Fit for Purpose
Evaluating the psychometric properties and targeting of PIRLS across language versions

Paper 3: DIF Happens
Examining comparability and differential item functioning across language versions

Language matters: PIRLS reading achievement in South Africa (Paper 1)

South African students demonstrate very low and unequal performance in PIRLS 2016 across language groups

Research questions

To what extent do:

- 1) contextual factors
- 2) use of the test language at home
- 3) test language

impact students' overall PIRLS 2016 reading achievement in South Africa?

Method

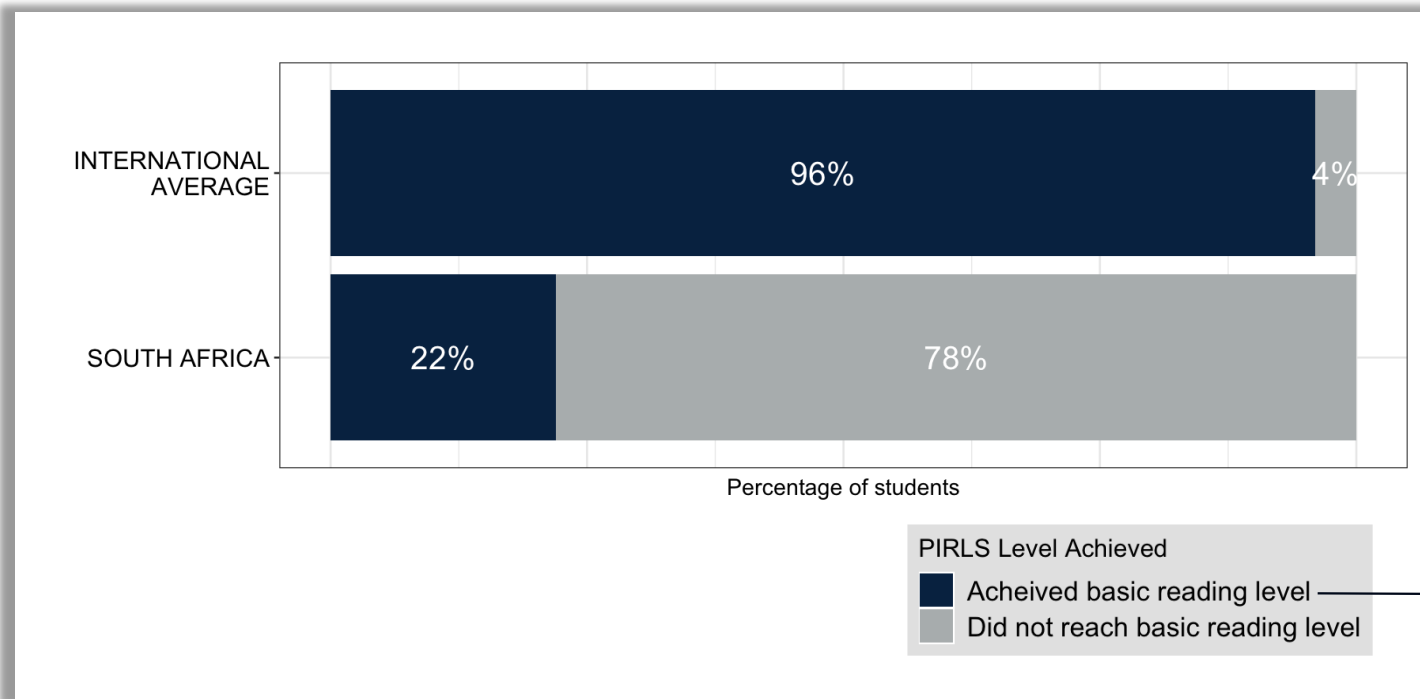
Multiple linear regression analysis

Dependent variable: PIRLS overall reading score

Independent variables:

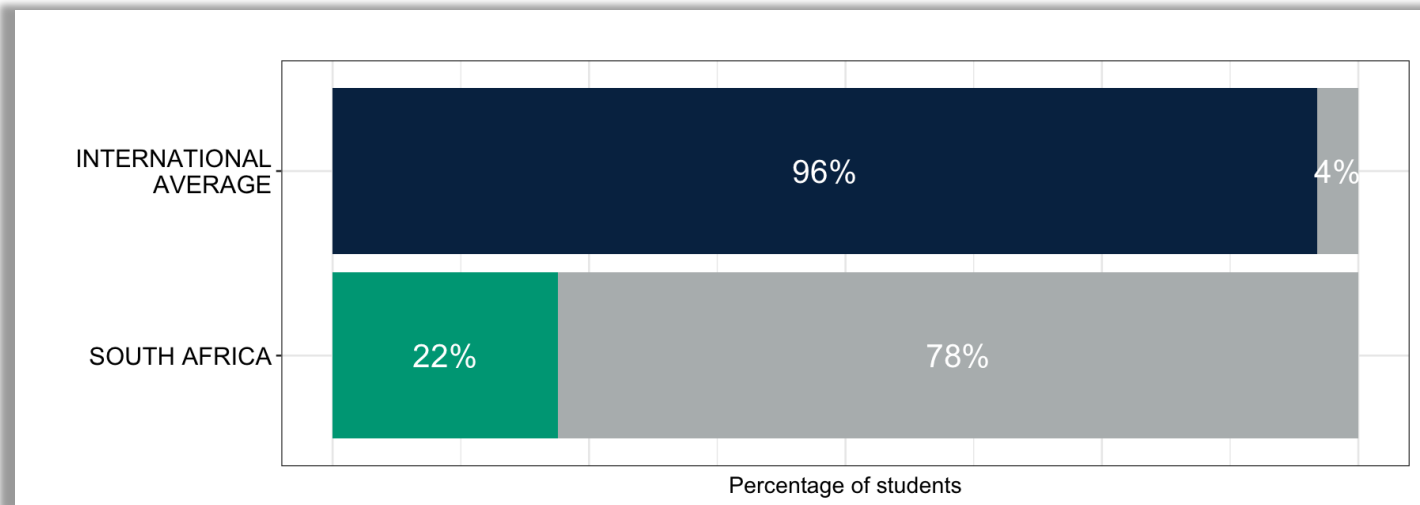
- Gender (control)
- Contextual factors (e.g. home resources)
- Language factors (e.g. language use at home)
- Test language

PIRLS 2016 results in South Africa



The low benchmark in PIRLS indicates that students can understand and extract basic information from the text

PIRLS 2016 results in South Africa

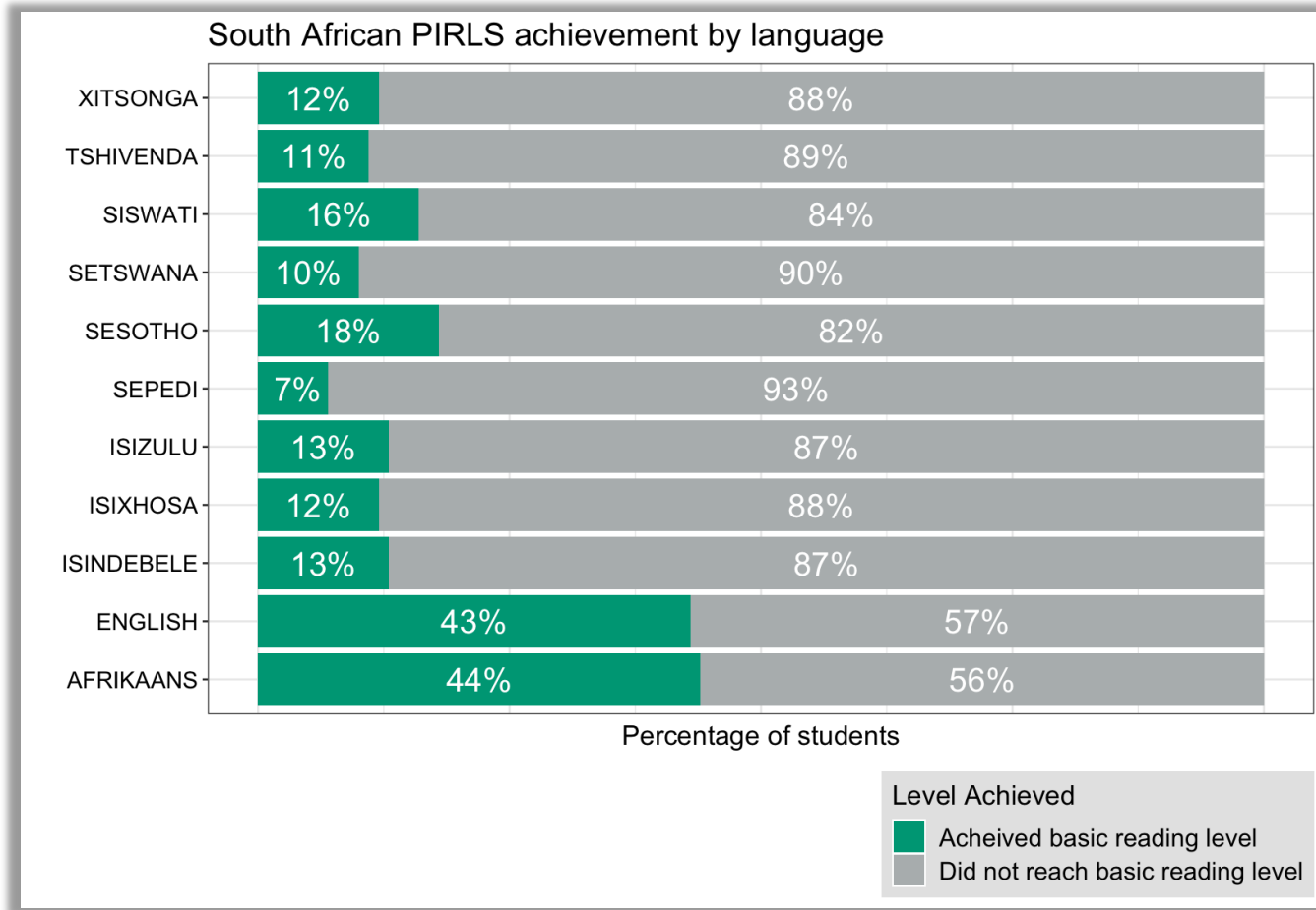


- Much lower overall performance compared to international average
- Wide attainment gaps between language groups within South Africa

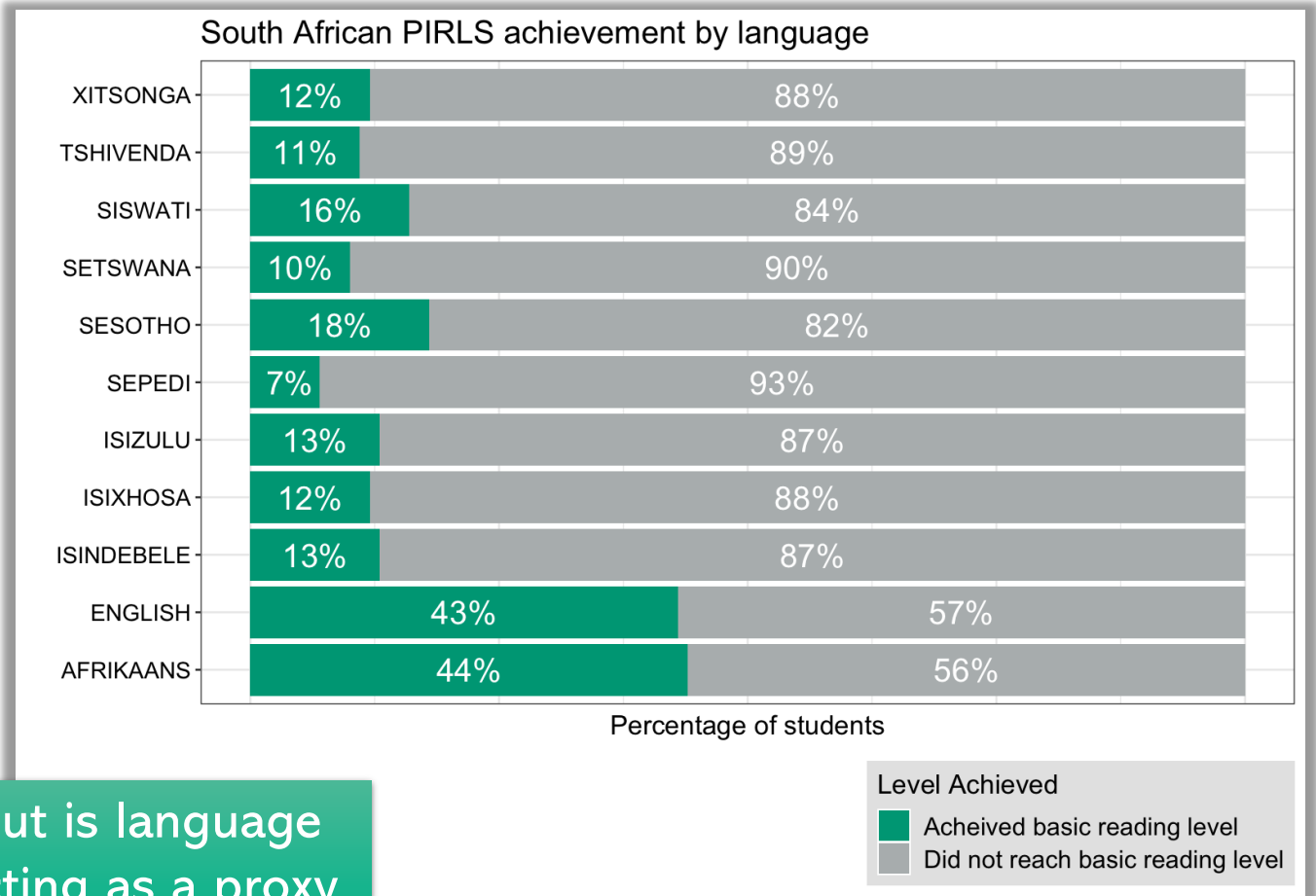




PIRLS results across language groups



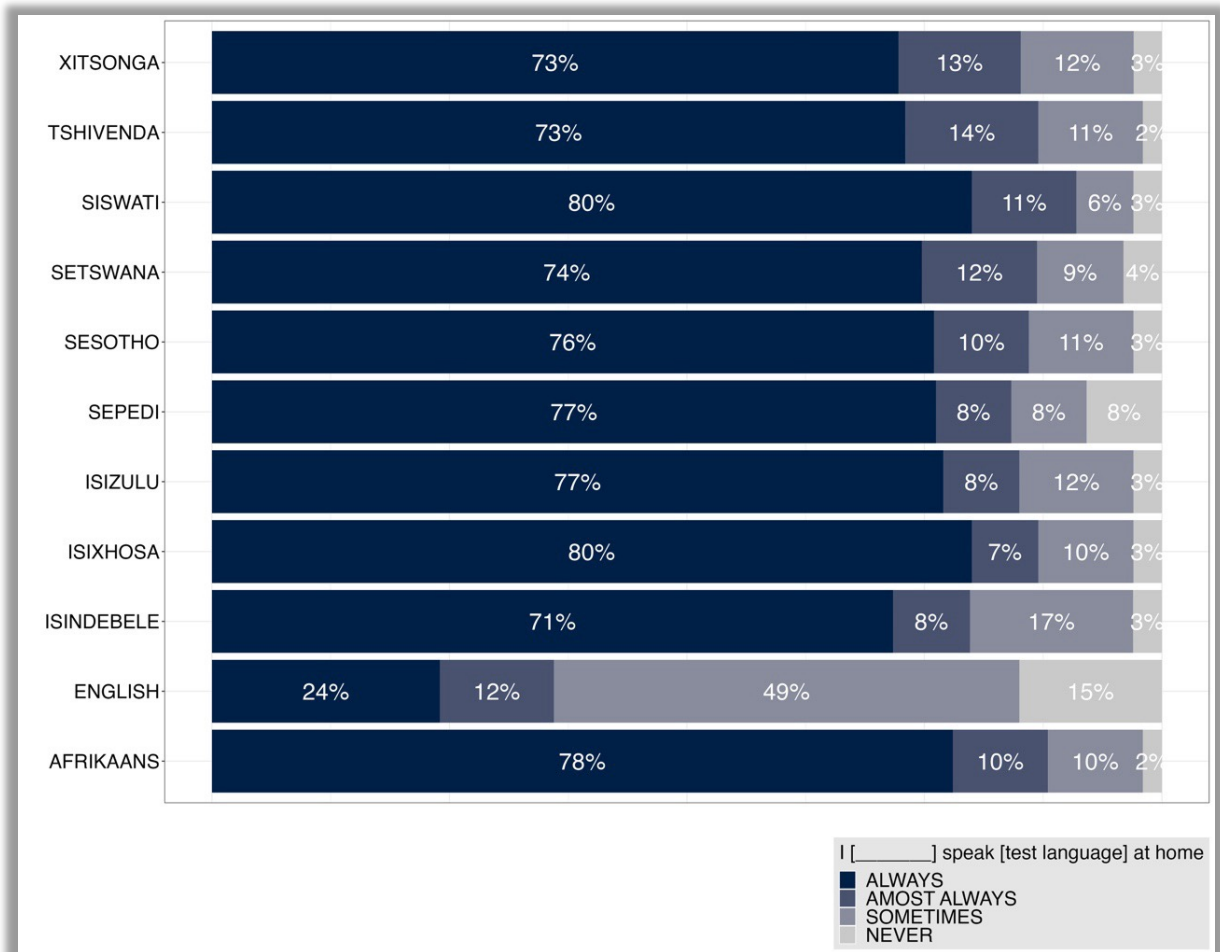
PIRLS results across language groups



But is language acting as a proxy for other factors?

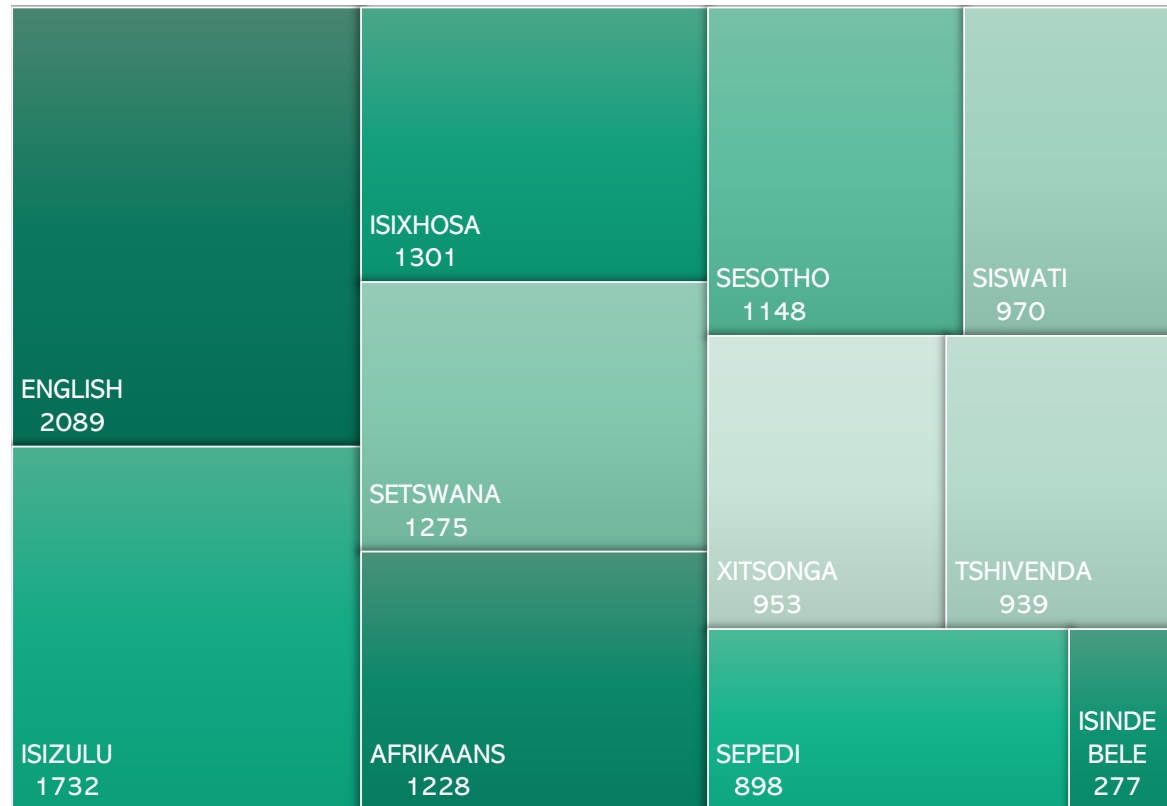
Test language use at home

Most groups have 80-90% home language speakers, except English which is closer to 36%



Data

SOUTH AFRICAN NATIONAL SAMPLE PIRLS LITERACY 2016



- Students (N = 12,810)
 - All 11 language groups
 - Grade 4 (~10 years)
- Achievement data
 - PIRLS Literacy 2016 overall reading scores
- Contextual data
 - Home, school and student questionnaire data

Does test language matter?

Model	Theme	Variables included	N-used	% variance explained (R ²)
Model 0	Gender	1	12807	6%
Model 1	Home factors	5	5586	23%
Model 2	Student factors	10	5157	36%
Model 3	School factors	15	4151	41%
Model 4	Language use	20	2479	45%
Model 5	Test language	21	2479	54%

Does test language matter?

Model	Theme	Variables included	N-used	% variance explained (R ²)
Model 0	Gender	1	12807	6%
Model 1	Home factors	5	5586	23%
Model 2	Student factors	10	5157	36%
Model 3	School factors	15	4151	41%
Model 4	Language use	20	2479	45%
Model 5	Test language	21	2479	54%

Does test language matter?

Model	Theme	Variables included	N-used	% variance explained (R ²)
Model 0	Gender	1	12807	6%
Model 1	Home factors	5	5586	23%
Model 2	Student factors	10	5157	36%
Model 3	School factors	15	4151	41%
Model 4	Language use	20	2479	45%
Model 5	Test language	21	2479	54%

Does test language matter?

Model	Theme	Variables included	N-used	% variance explained (R ²)
Model 0	Gender	1	12807	6%
Model 1	Home factors	5	5586	23%
Model 2	Student factors	10	5157	36%
Model 3	School factors	15	4151	41%
Model 4	Language use	20	2479	45%
Model 5	Test language	21	2479	54%

Test language matters

54%

of variance explained
by language and
contextual factors
together

Significant negative effects
seen for rural school location,
low confidence in reading,
and test language

- Test language had significant impact on achievement, beyond home language use
- Students taking test in an African language scored **51-130 points lower** than the English group (after controlling for other factors)

Some inconsistencies and issues

In South Africa, language of instruction, race and school quality are almost impossible to disentangle, and the nuances are not accounted for from an international instrument like PIRLS



Inconsistent reports of home language

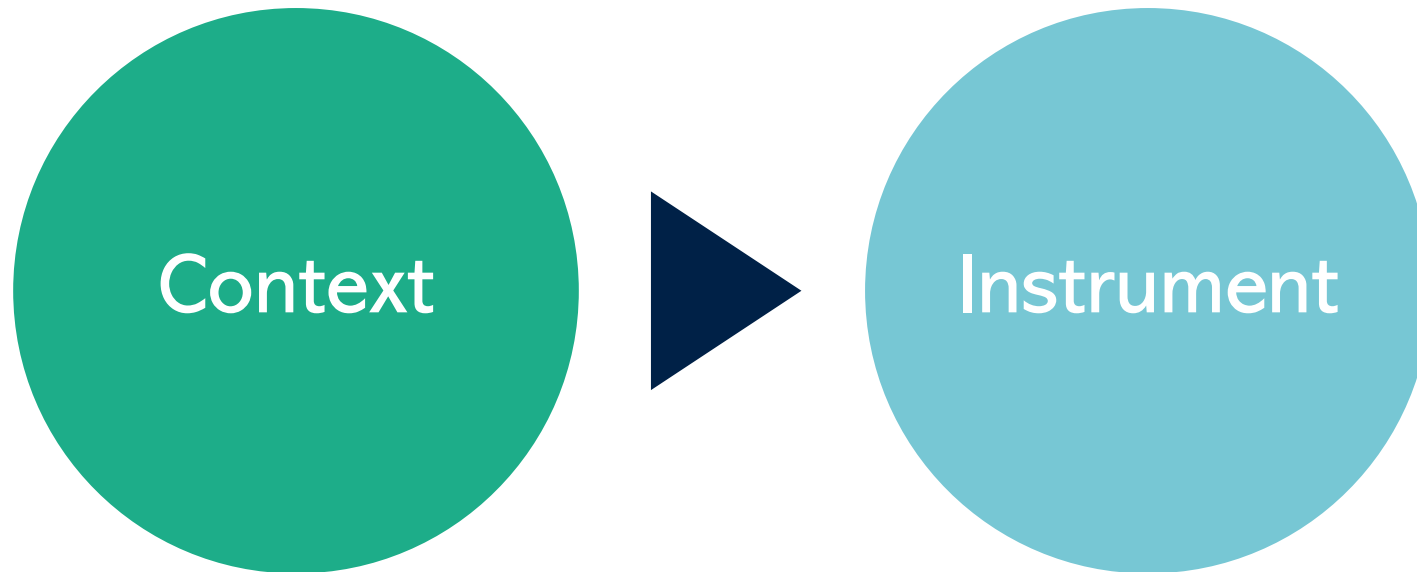


Language and school quality



School location

Evaluating the instrument



The research study

What can be learned from PIRLS Literacy 2016 about the reading achievement of South African students across different languages?

CONTEXTUAL ANALYSIS

Paper 1: Language Matters
Exploring how contextual and linguistic factors influence reading achievement across language groups

PSYCHOMETRIC ANALYSIS

Paper 2: Fit for Purpose
Evaluating the psychometric properties and targeting of PIRLS across language versions

Paper 3: DIF Happens
Examining comparability and differential item functioning across language versions

Fit for purpose or off the mark: PIRLS 2016 in South Africa (Paper 2)

South Africa has very low performance on PIRLS 2016, substantially below the international centrepoint with a wider range of within country variability than any other country

Research questions

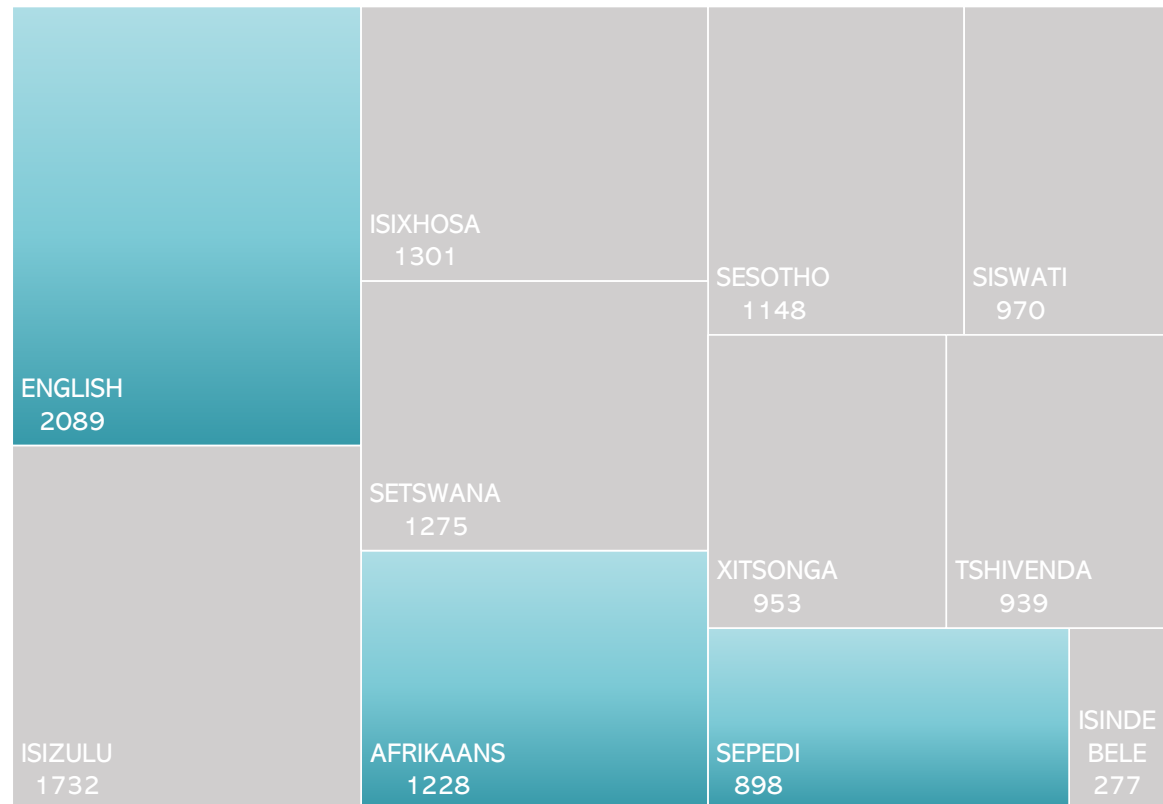
1. To what extent do the models used by PIRLS 2016 fit students' data?
2. To what extent is the difficulty level of items in PIRLS Literacy 2016 suitably targeted to the ability of students?

Method

Item response theory (IRT) framework
2PL & GPCM model
Checking fit of the model and items
Investigating targeting of the items to student ability

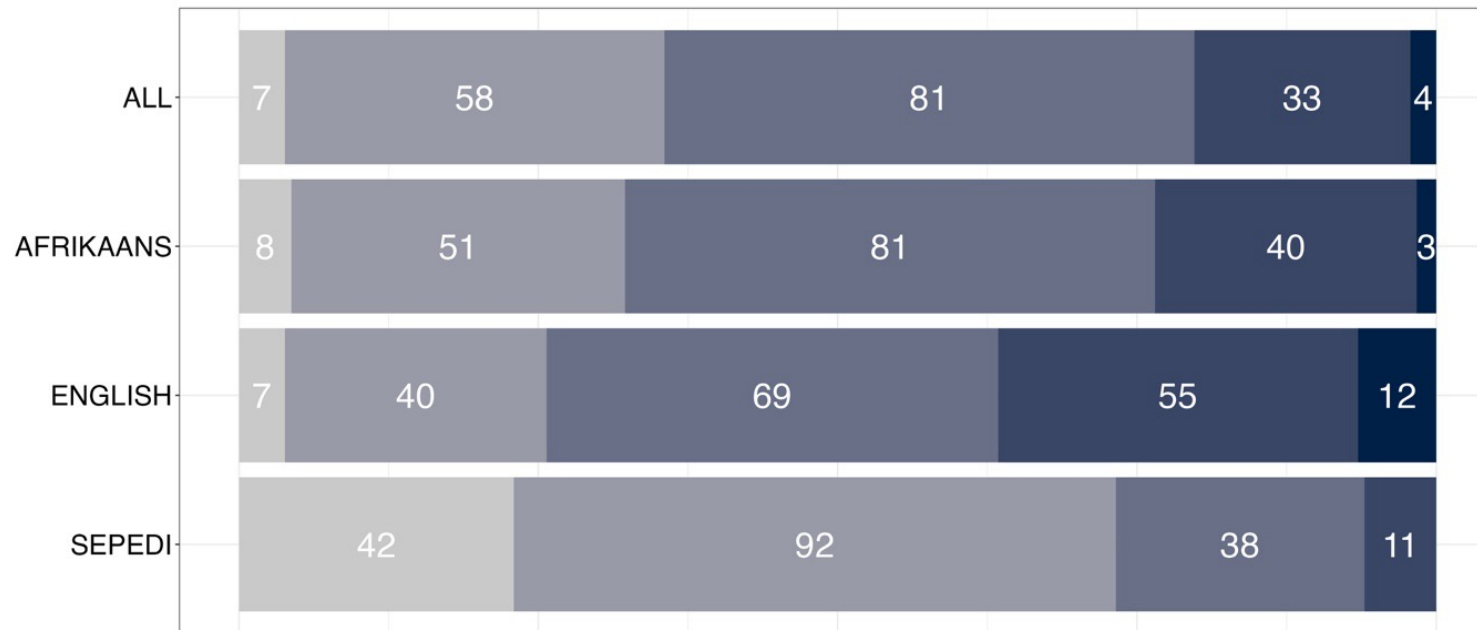
Data

SOUTH AFRICAN NATIONAL SAMPLE PIRLS LITERACY 2016



- Students (n = 4,150)
 - Three language groups
 - English (high performing)
 - Afrikaans (high performing)
 - Sepedi (lowest performing)
- Instrument data (n = 183)
 - 12 passages
 - ~15 items each
 - 90 multiple choice, 93 constructed response

% correct responses

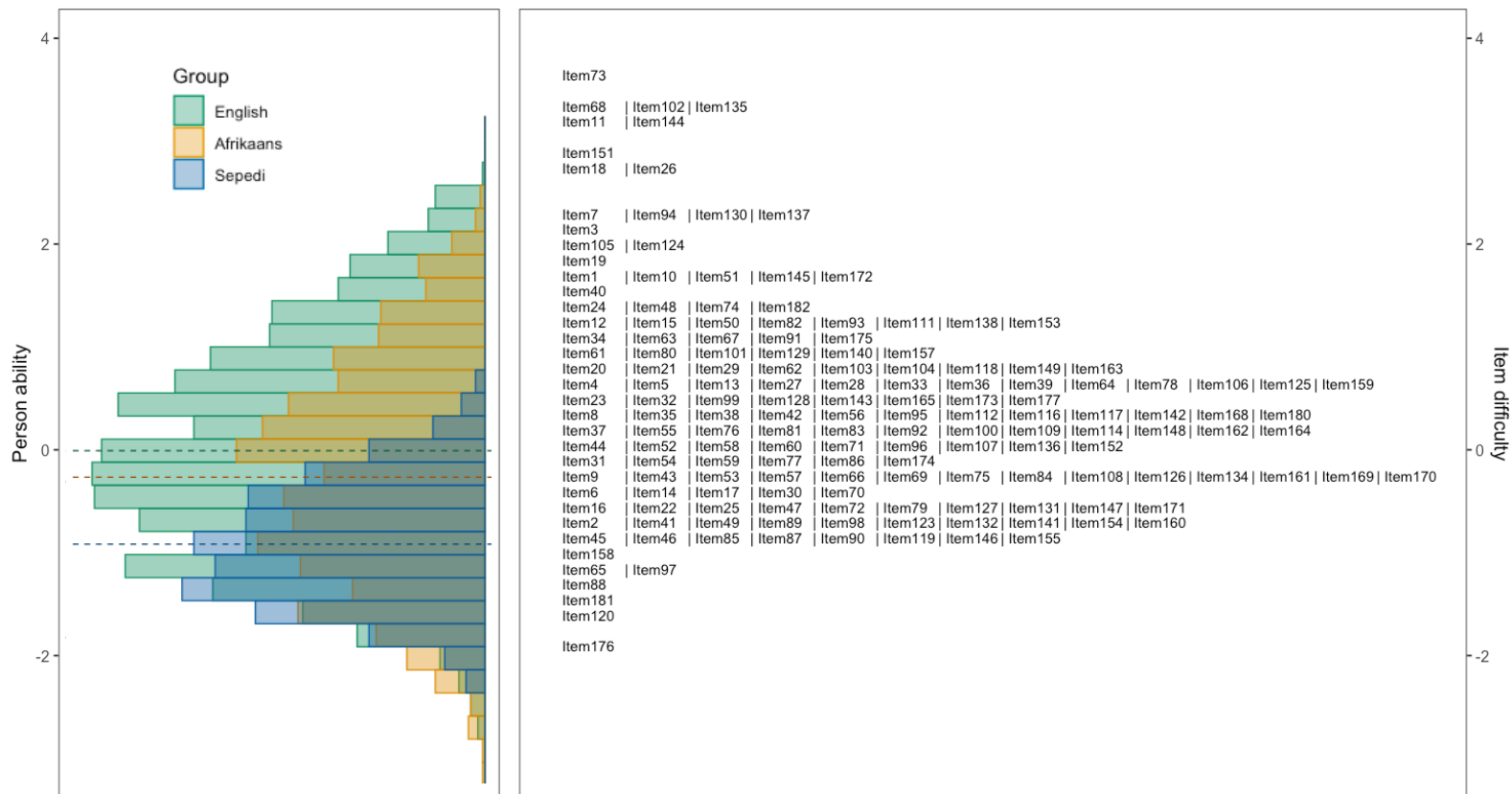


Number of items (n = 183 items)

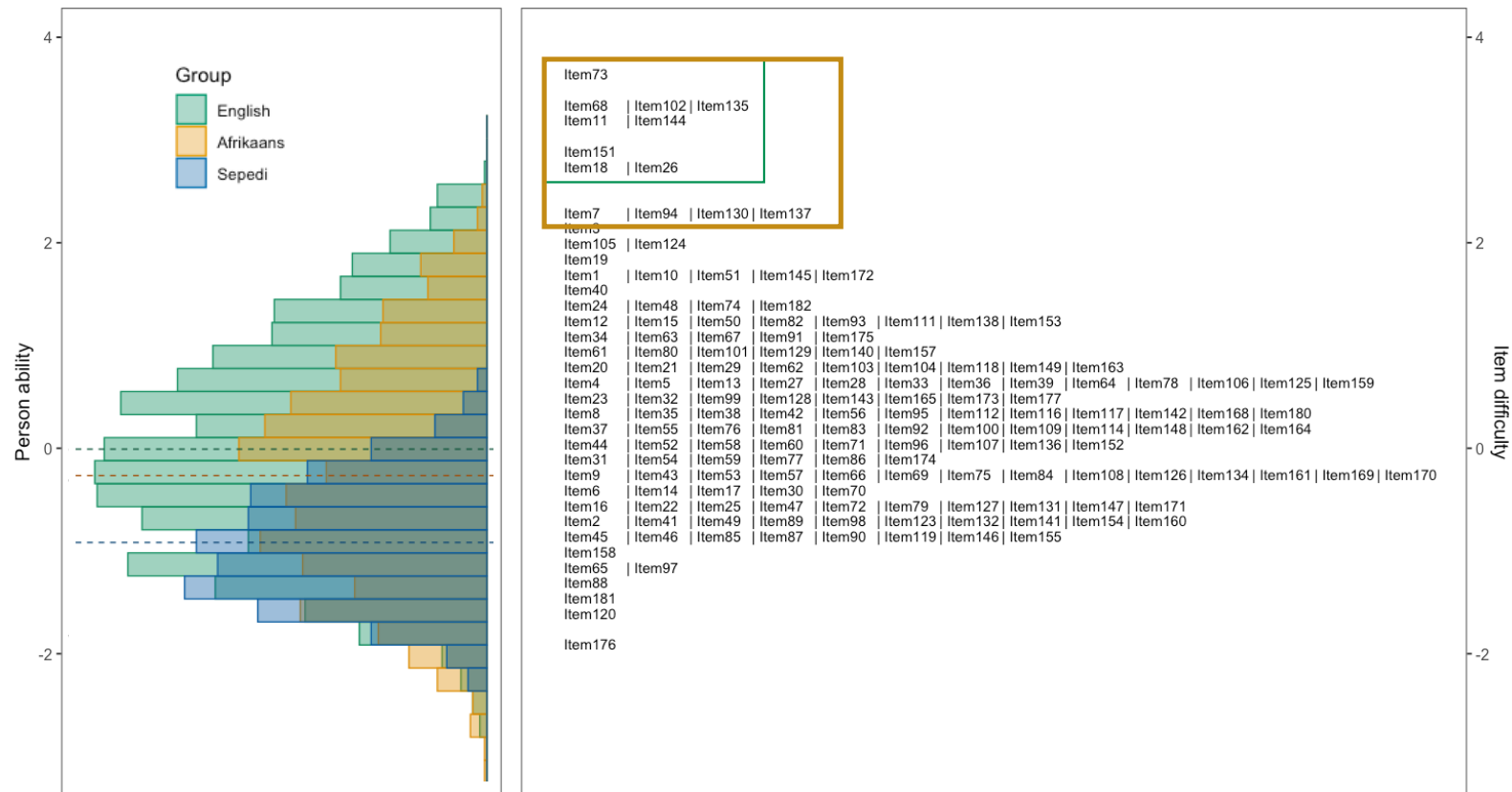
Percentage of students responding correctly

- 0-20% correct
- 21-40% correct
- 41-60% correct
- 61-80% correct
- 81-100% correct

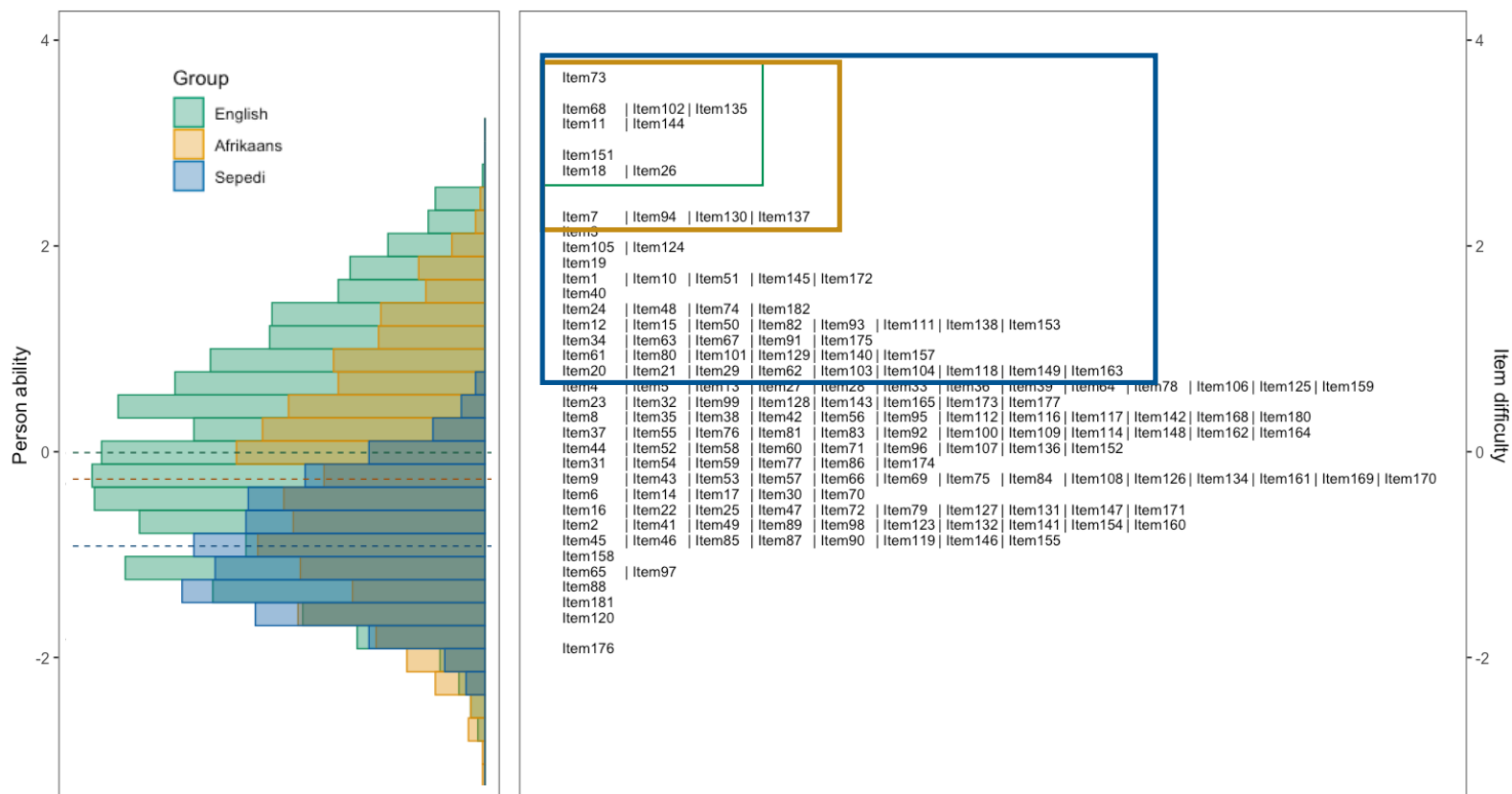
Checking targeting by group



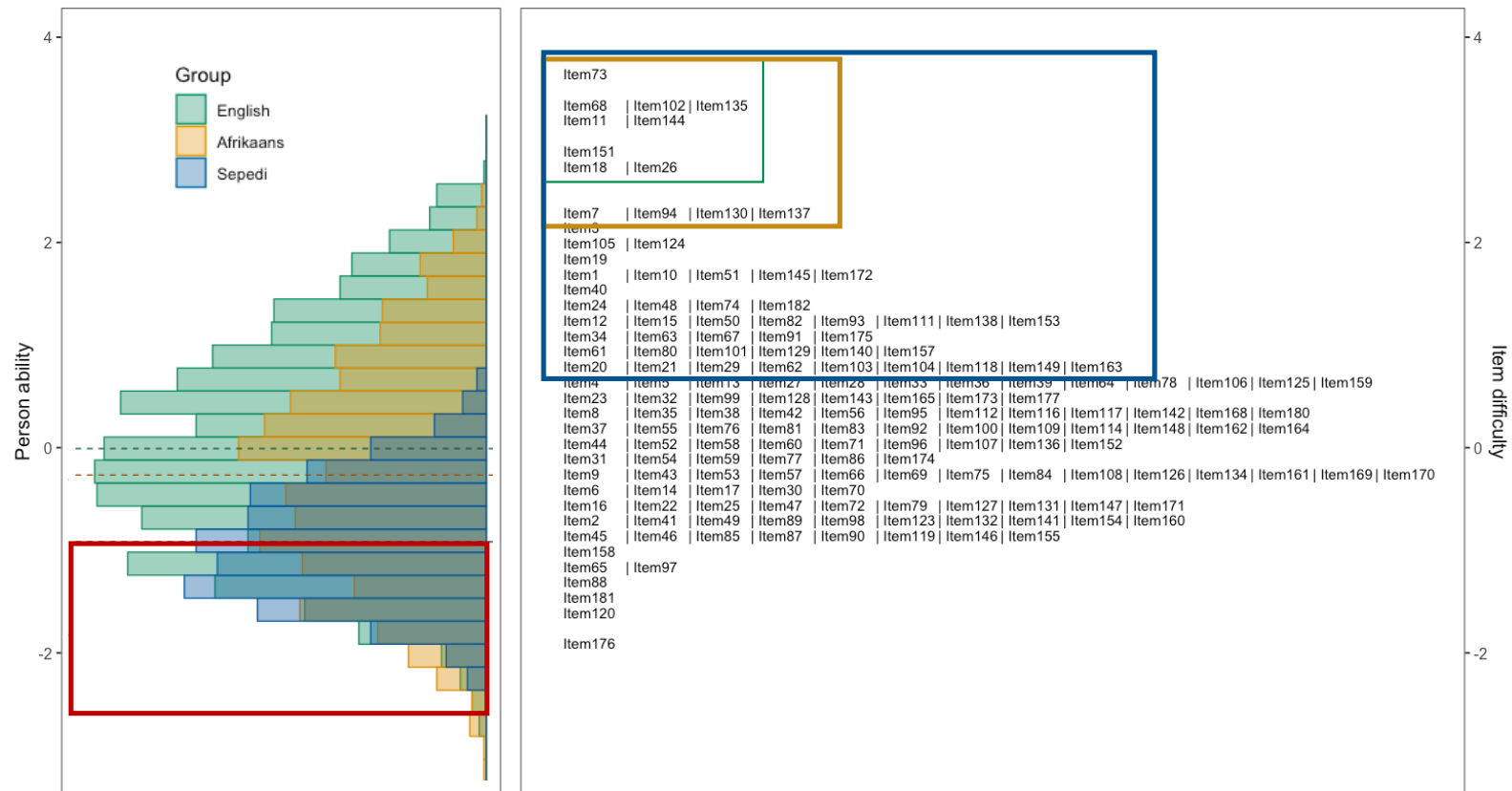
Checking targeting by group



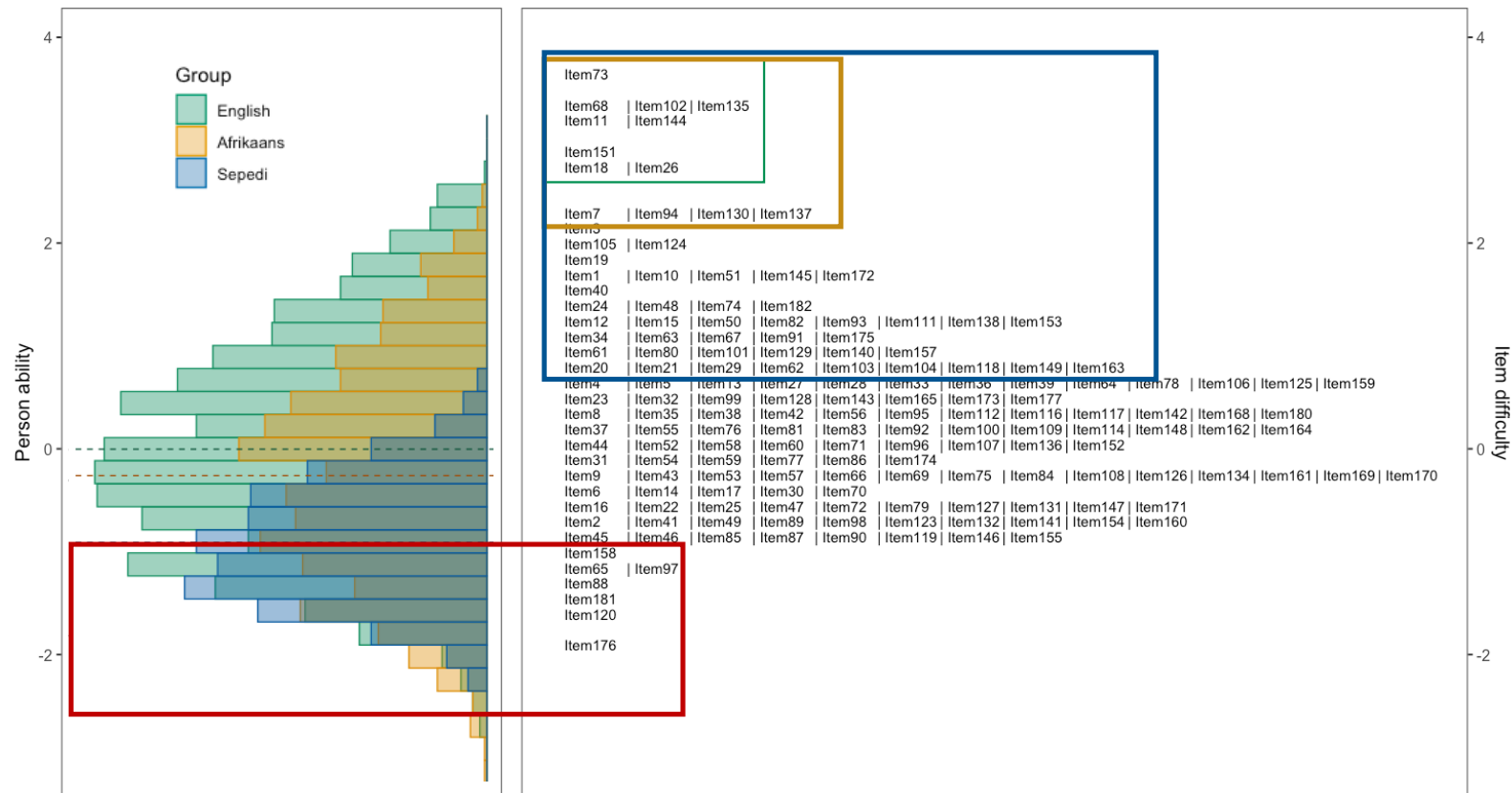
Checking targeting by group



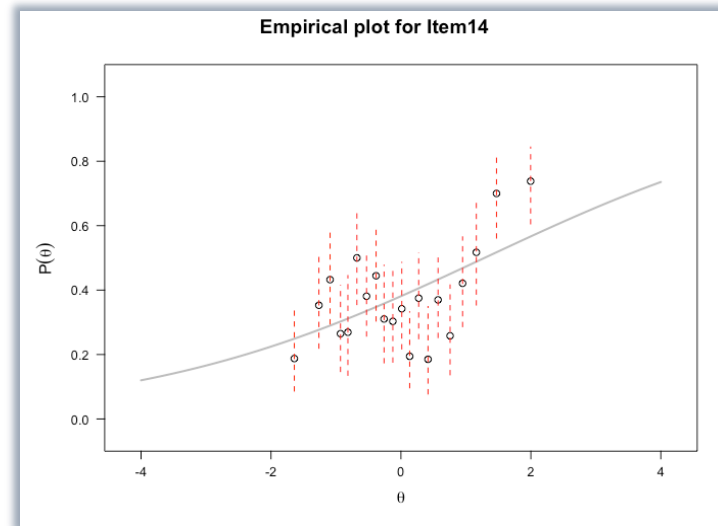
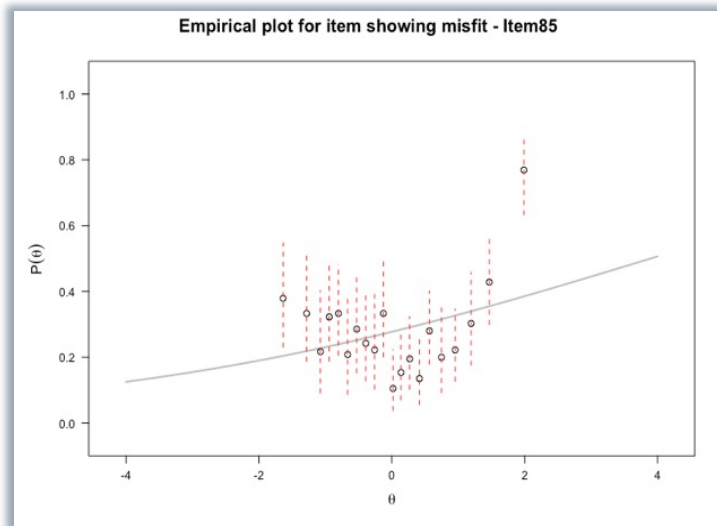
Checking targeting by group



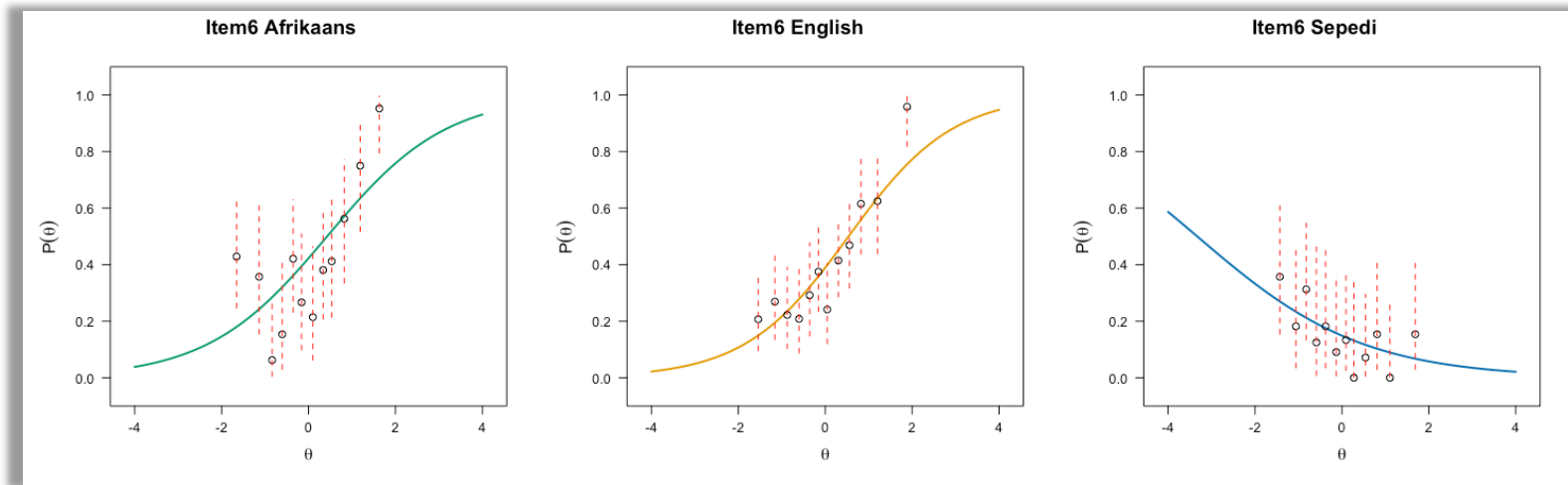
Checking targeting by group



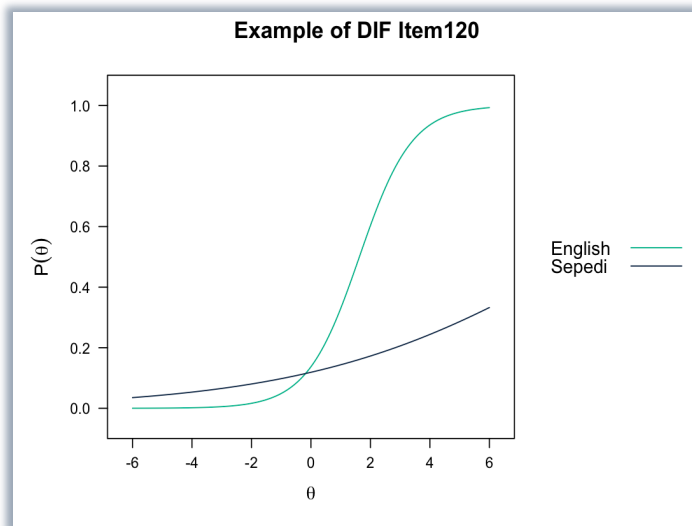
Fit issues



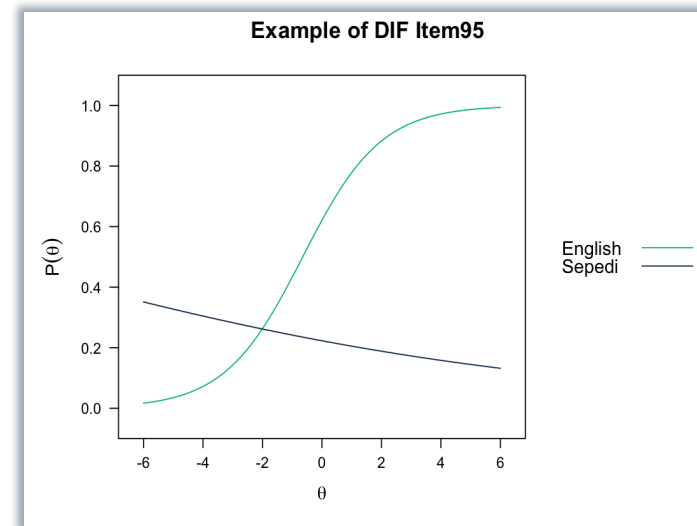
Fit issues



Fit issues



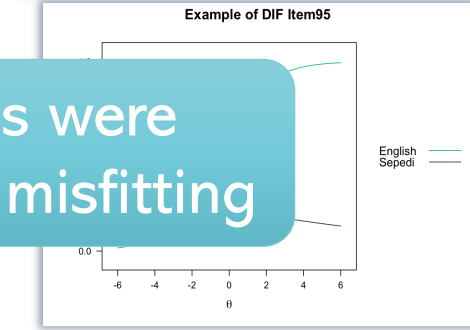
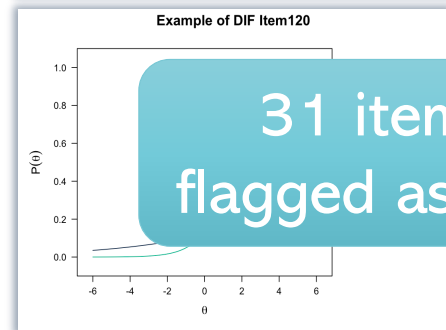
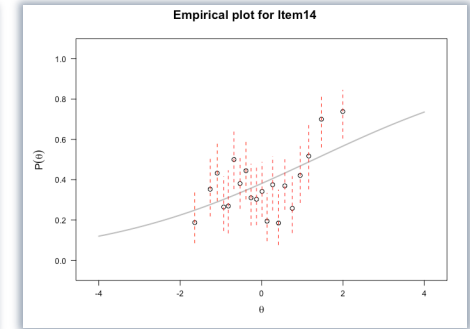
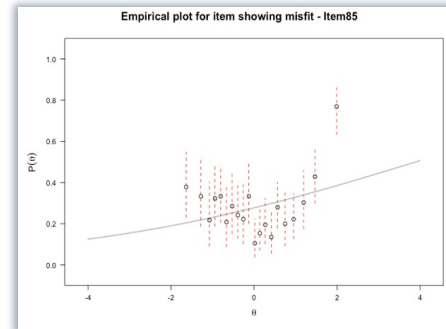
48 items had low discrimination
($a < .50$)



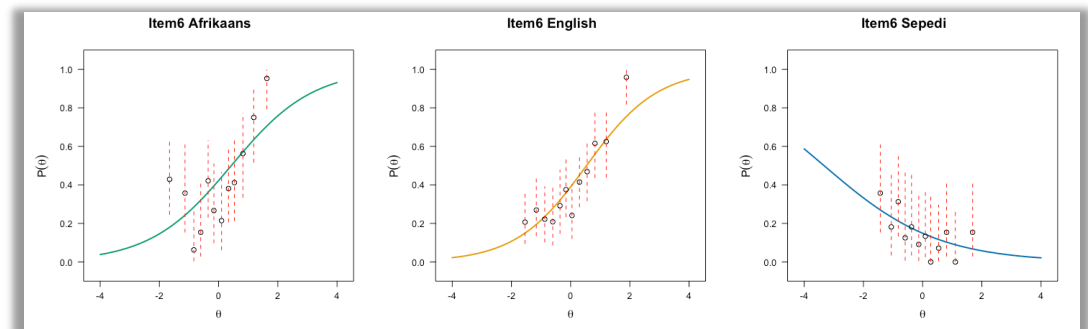
14 items discriminated negatively
(12 for Sepedi)

Fit issues

- 14 items had <15 correct responses
- 21 items had poor fit with the model expectations
- 7 items loaded on no factors in dimensionality checks



31 items were flagged as misfitting



The research study

What can be learned from PIRLS Literacy 2016 about the reading achievement of South African students across different languages?

CONTEXTUAL ANALYSIS

Paper 1: Language Matters
Exploring how contextual and linguistic factors influence reading achievement across language groups

PSYCHOMETRIC ANALYSIS

Paper 2: Fit for Purpose
Evaluating the psychometric properties and targeting of PIRLS across language versions

Paper 3: DIF Happens
Examining comparability and differential item functioning across language versions

DIF happens: The impact of differential item functioning across language versions (Paper 3)

Substantial reading achievement gaps across language versions means there is a need to determine the comparability of the test difficulty across versions

Research questions

To what extent are the Afrikaans, English, and Sepedi versions of PIRLS Literacy 2016 comparable:

- 1) At an item level?
- 2) At a passage level?
- 3) At the overall test level?

Method

ITEM LEVEL

IRT-based differential item functioning (DIF) analysis
Likelihood-ratio tests to test for DIF, IRT-LR (Thissen, 2001)
Effect sizes determined using Meade's (2010) taxonomy

PASSAGE & TEST LEVEL

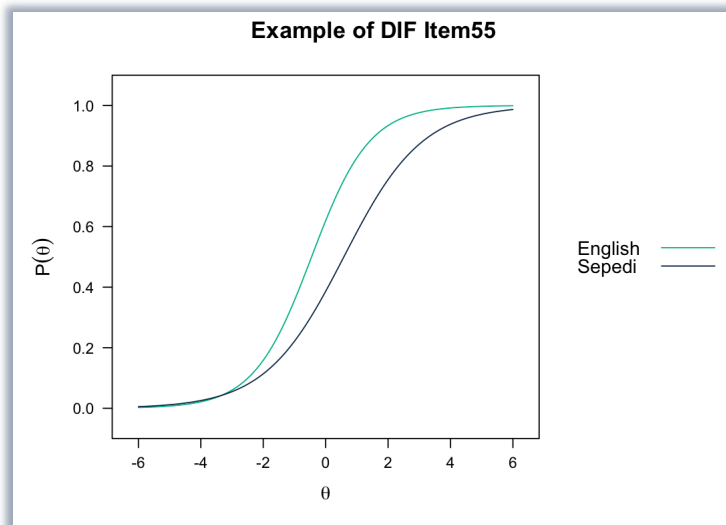
Differential response functioning (DRF) framework (Chalmers, 2018)

Used signed (directional) and unsigned (absolute) DRF statistics for effect size

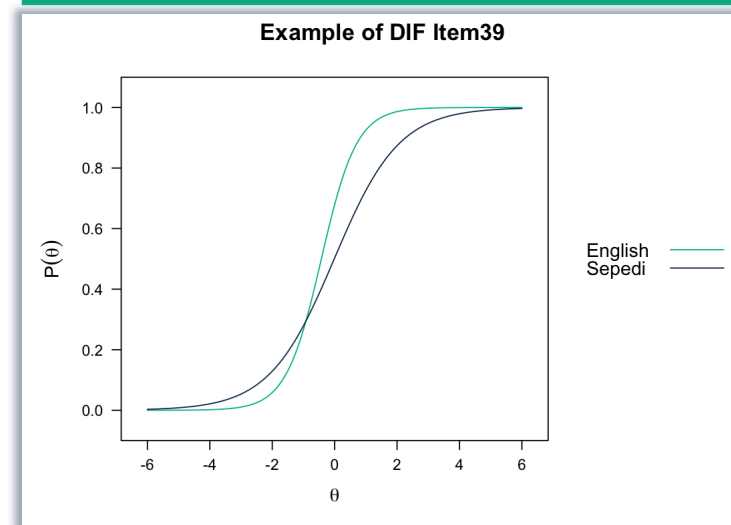
Significant differences in item difficulty

- 28% of items (42 of 152) displayed DIF - Sepedi focal
- 20% of items (30 of 152) displayed DIF - Afrikaans focal

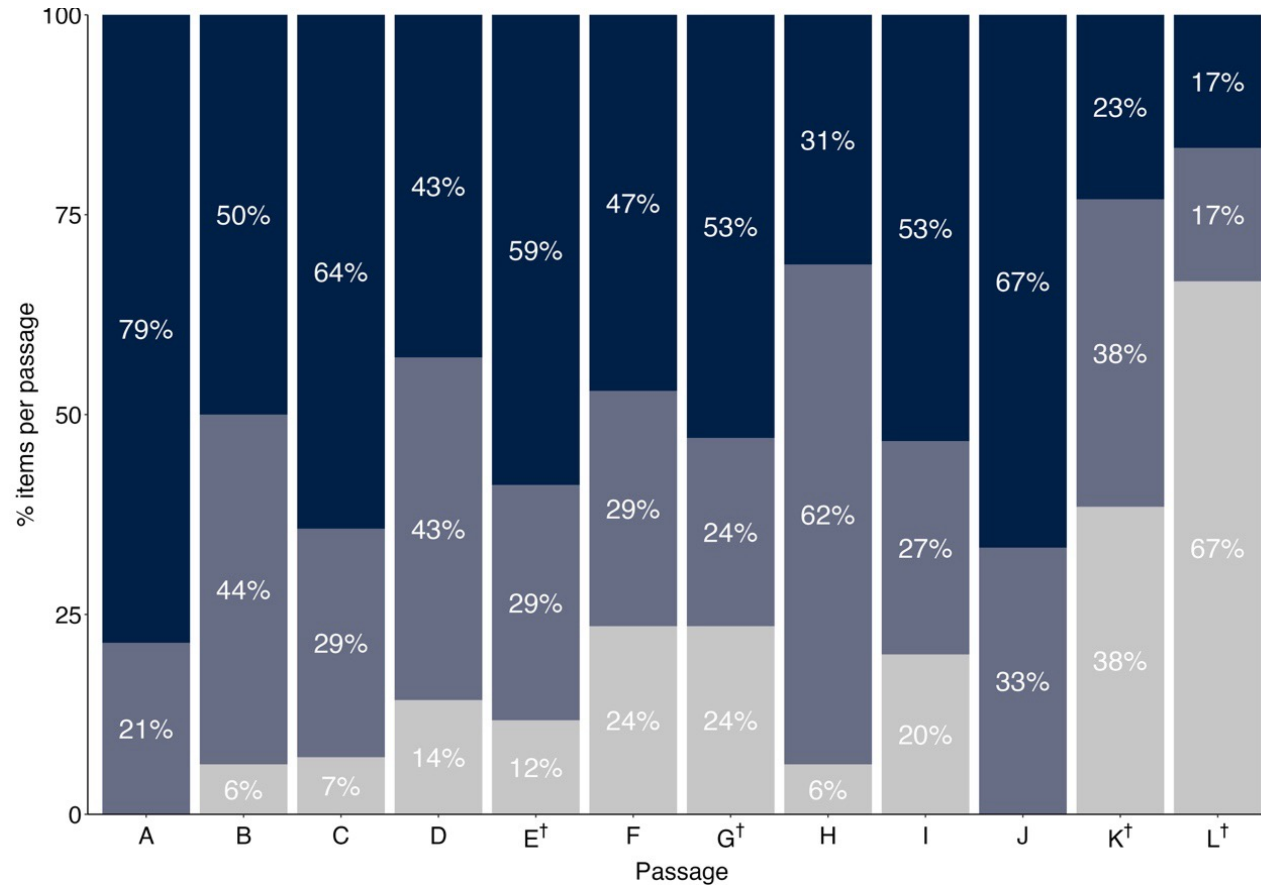
9 of 14 large uniform DIF items advantaged English students



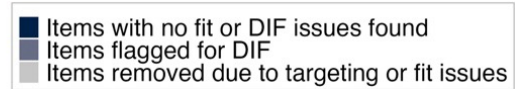
Typically, the non-uniform DIF items advantaged lower-ability Sepedi students, and higher ability English students



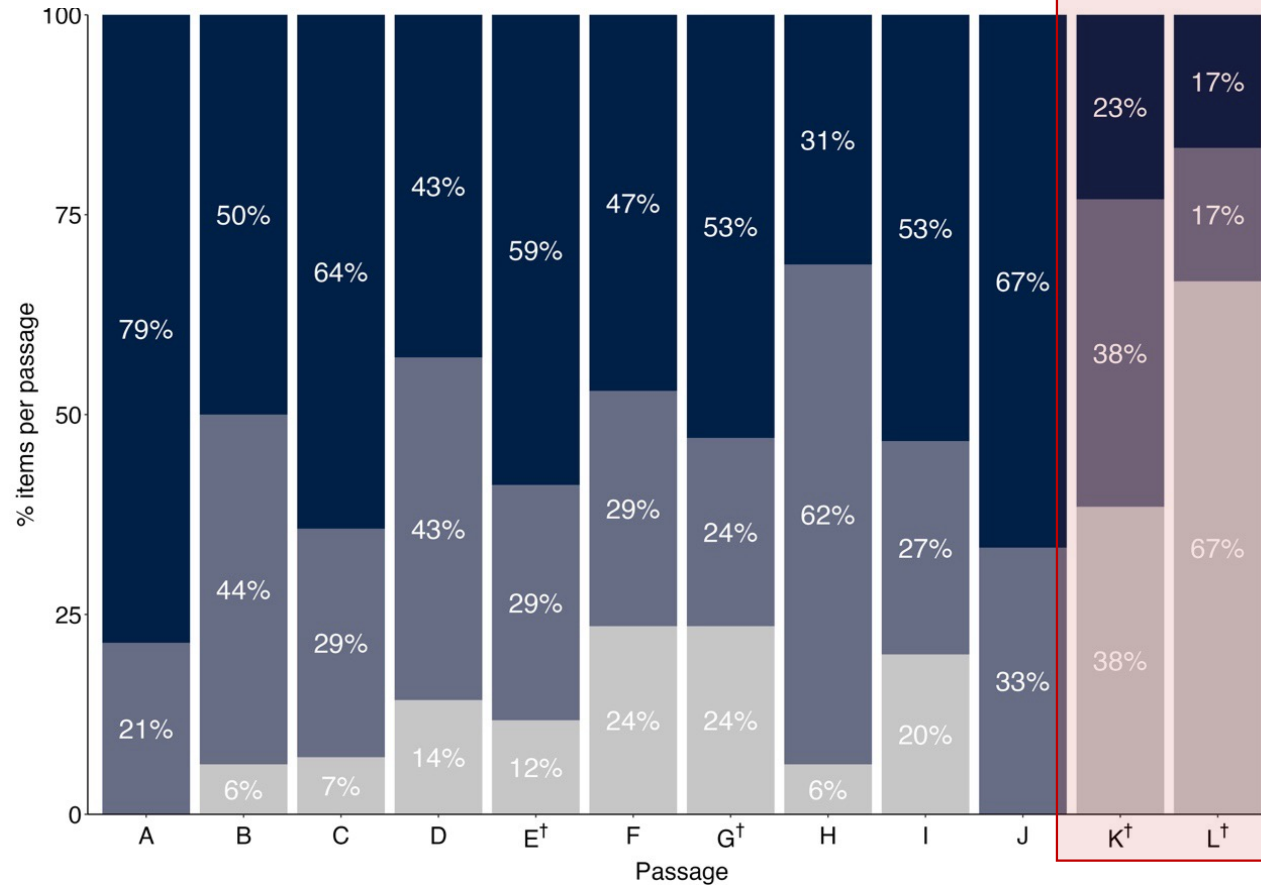
Looking across passages



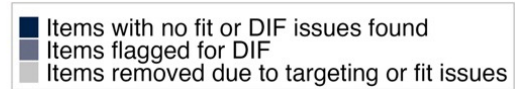
† indicates linking passages (with main PIRLS test)



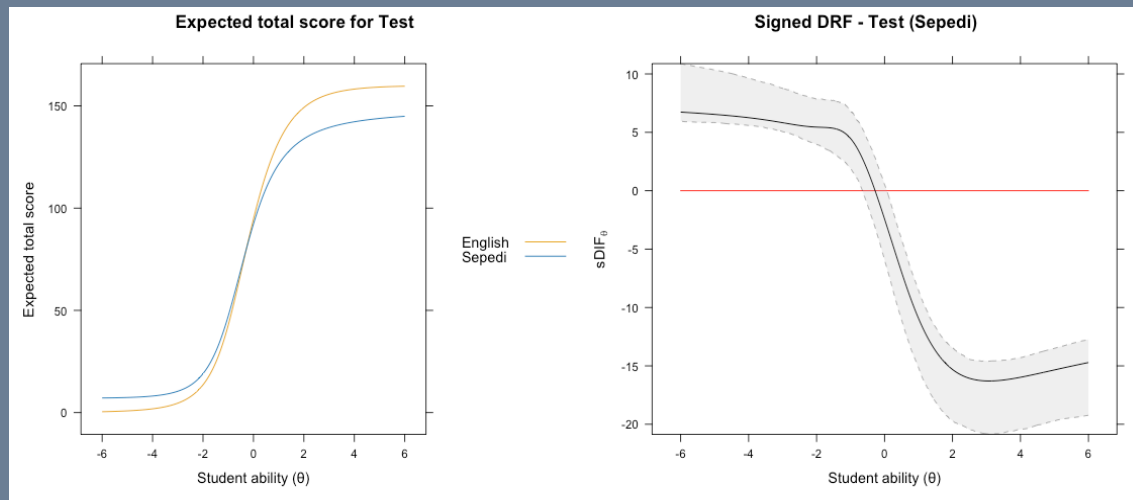
Looking across passages



† indicates linking passages (with main PIRLS test)



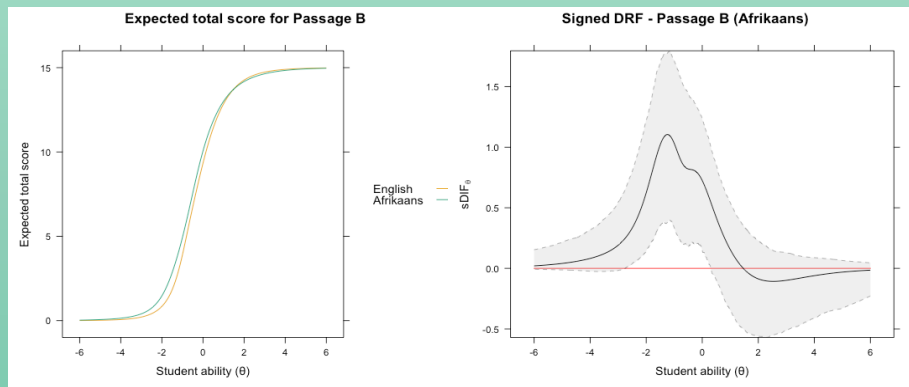
Test level impact of DIF



While no consistent directional bias was found, significant comparability issues exist overall between the English and Sepedi versions

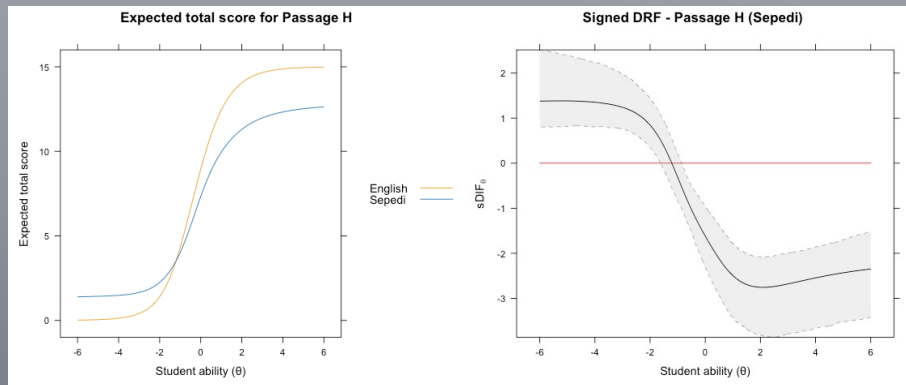
- Absolute difference of 5.29 points between the English and Sepedi versions
- No significant directional divergence in expected scores across groups
 - sDRF results result suggest the non-uniform DIF effects that cancelled each other out at the overall test level

Examples of sDRF plots – Passages



sDRF of .67, 95% CI (.33 to 1)

This passage was consistently easier for Afrikaans students compared to English students of matched proficiency



sDRF of -1.13, 95% CI (-1.63, -.67)

Although the direction of advantage changed ($\sim \theta = -1$), on average across the full range of ability these passages disadvantaged Sepedi students

Passage level impact of DIF

- Considered signed and unsigned DRF statistics
- uDRF gives an indication of **absolute** divergence
 - 11 of 12 passages significantly different for Sepedi group
 - 4 of 12 passages significantly different for Afrikaans group
- sDRF gives an indication of **direction** of advantage
 - 5 passages advantaged English group
 - 1 passage advantaged Sepedi group
 - 1 passage advantaged Afrikaans group

A stylized tree graphic with a grey trunk and branches, and several green leaves, positioned on the left side of the slide.

But what does this all mean?

Missing the mark

- PIRLS overall difficulty level substantially misaligned with student abilities (despite using easier PIRLS Literacy version)
- More severe targeting issues for African language groups, especially Sepedi
- Impacts the validity, reliability and fairness of the test
- What does this mean in practice?
 - *Within* country comparisons are problematic
 - Issues with linking passages also mean that cross country comparisons are problematic
- Why is this a problem?

Why is this a problem?

- Cross country comparison is the main purpose of PIRLS
- South Africa has used and continues to use PIRLS for within country comparisons
- It is also used to inform targeted interventions and track progress
- Informing language policy reform is highlighted as a main reason for participating in PIRLS

PIRLS is used for two primary purposes in South Africa: monitoring progress towards national reading development goals; and informing education policy and targeted resource allocation

Implications

For South Africa:

- Results show evidence of comparability issues
- Results raise questions about the validity of PIRLS in South Africa, particularly for low performing groups
- PIRLS by language has limited value for policy decisions

Broader contributions:

- Issues with cross country comparability
- Psychometric scrutiny for developing countries is rarely done
- Many of the issues found are likely to be the same for other low performing countries
- Concerns raised about effectiveness of adaptive methods

Strategic priorities for South Africa

1. Improve African language resources
2. Enhance teacher education for reading pedagogy
3. Address school resource inequalities

To achieve these, we need valid, comparable and reliable data about learning progress within the country

Moving forward



Testing
agencies



Participating
countries



Researchers

Shared responsibility



Thank you

