

# Using Timing Information Associated with Response Data in Large Scale Assessment

Kentaro Yamamoto

ETS

16 May 2019, Dublin, Ireland

6/6/2019



# Where are the Timing Information in PIAAC and PISA

- Some timing information are already in the data,
  - Scored responses
  - **Overall timing**
  - **Timing of first action**
  - Number of actions
- Also it can be harvested from logfiles
  - BQ timing
  - Time stamping of selected actions

# Overarching Goal of Using Timing Information

- Use timing data to increase reliability and validity of measurement that support measurement framework through:
  - Identification of data fabrication together with cognitive responses
  - Identification of errors added to the response
  - Interpretation of measurement constructs by examining the interaction with item types

# Respondents

- Omitted responses or early abandonment of survey
- Random responses (guessing, response styles)
- Intentional erroneous responses

# Interviewer

- Fake interview
- Shortened data collection
- Replicating data
- Multiple cases together
- Particular responses
- Increased non-interview

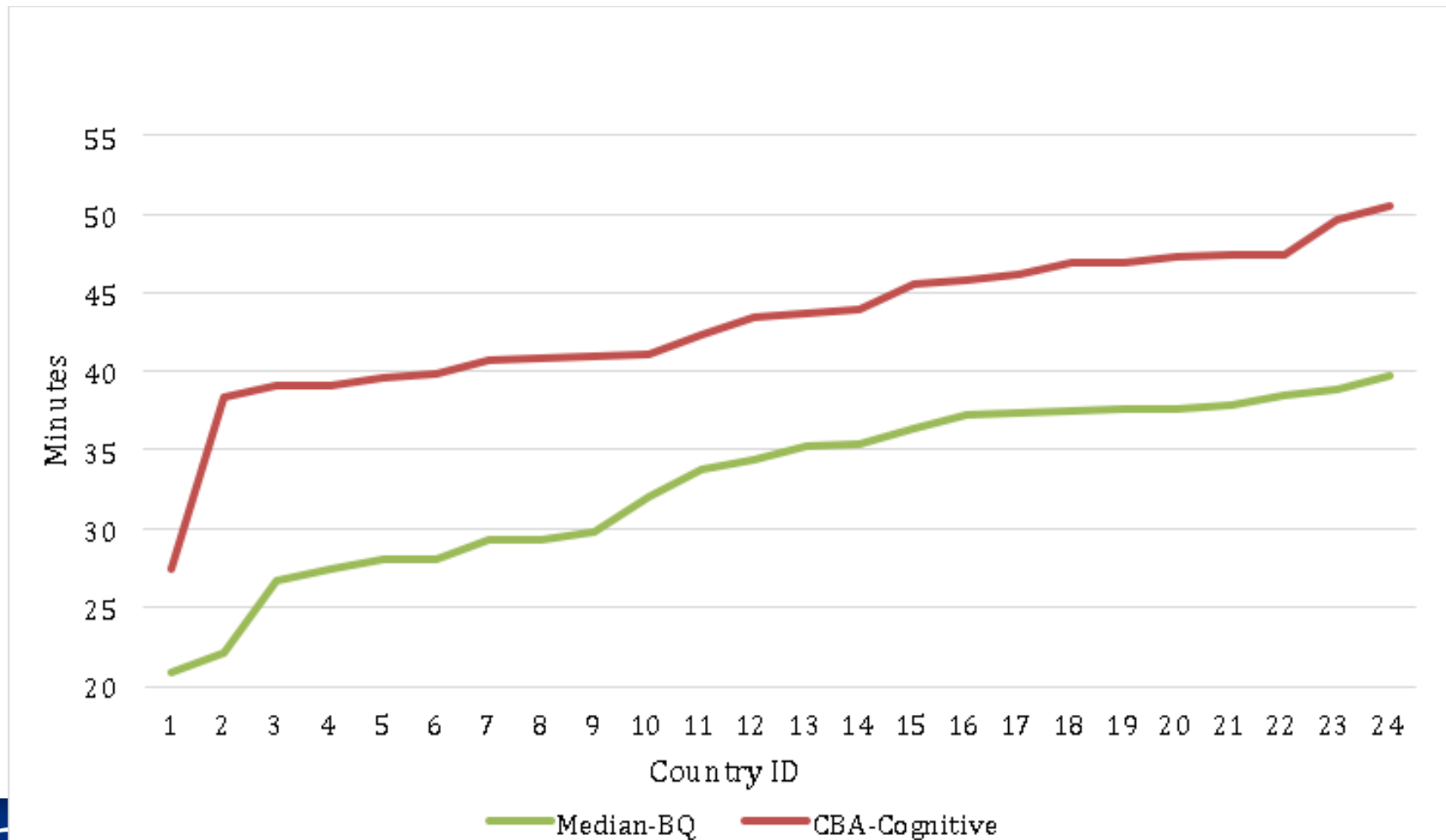
# Data Collection Contractor

- Duplication of data
- Synthetic data
- Dropping of data
- Insufficient sampling information

# Item Instrumentation

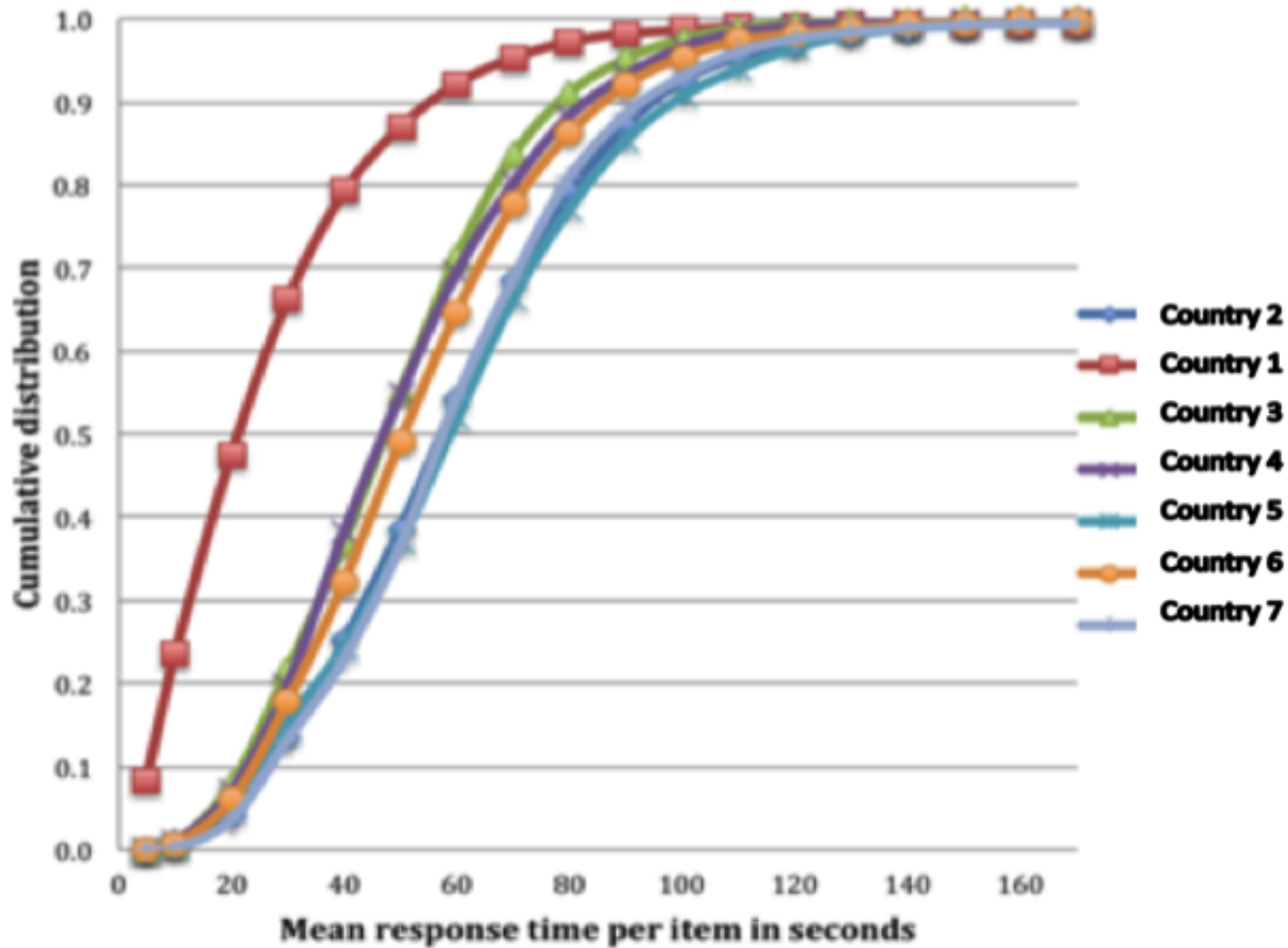
- Response time by item by country/language tend to show consistency over cycles and it can be used for QC of instruments, data, administration and data processing.

# Median Time of BQ and Cognitive Modules by Country





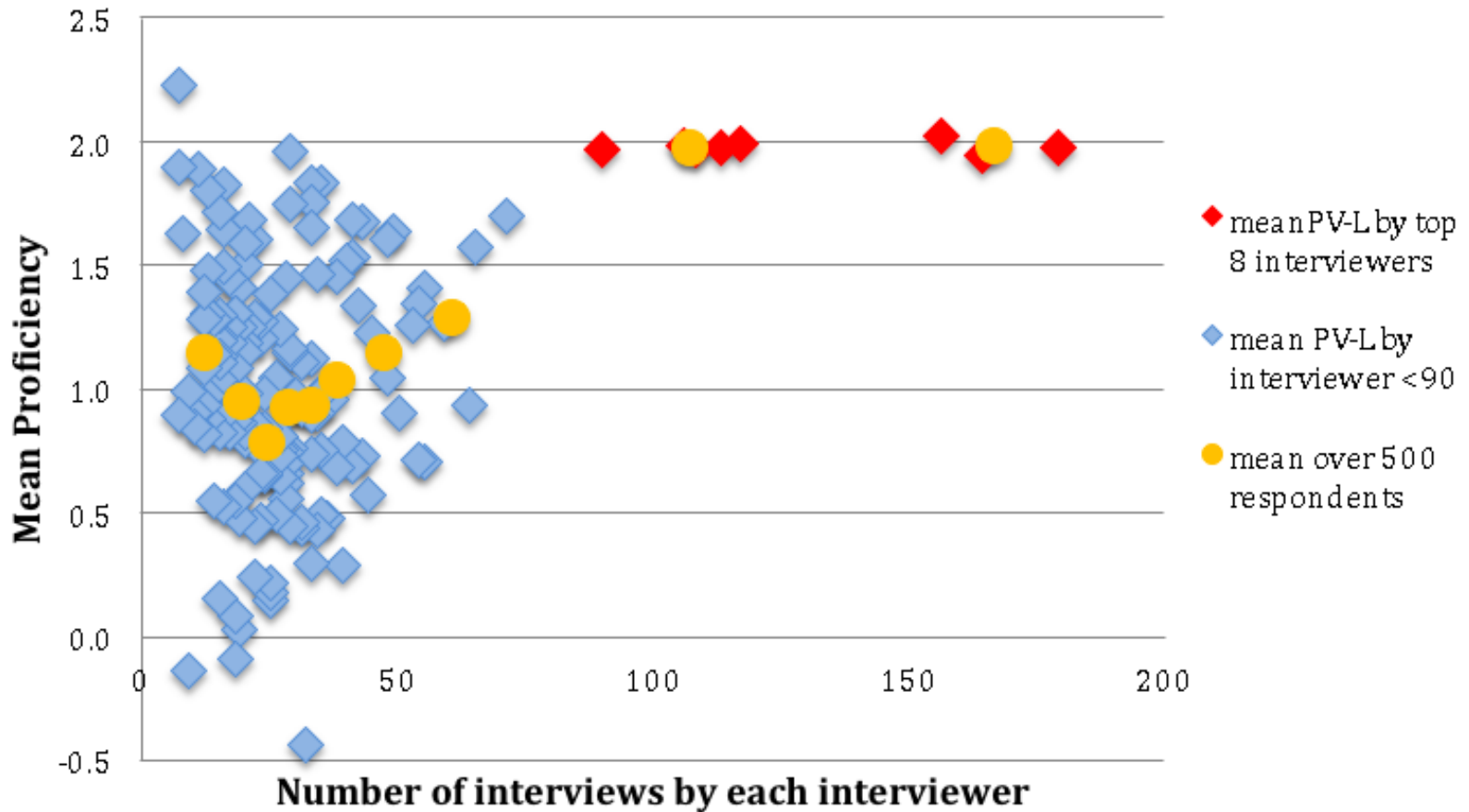
# Cumulative distribution of mean response time



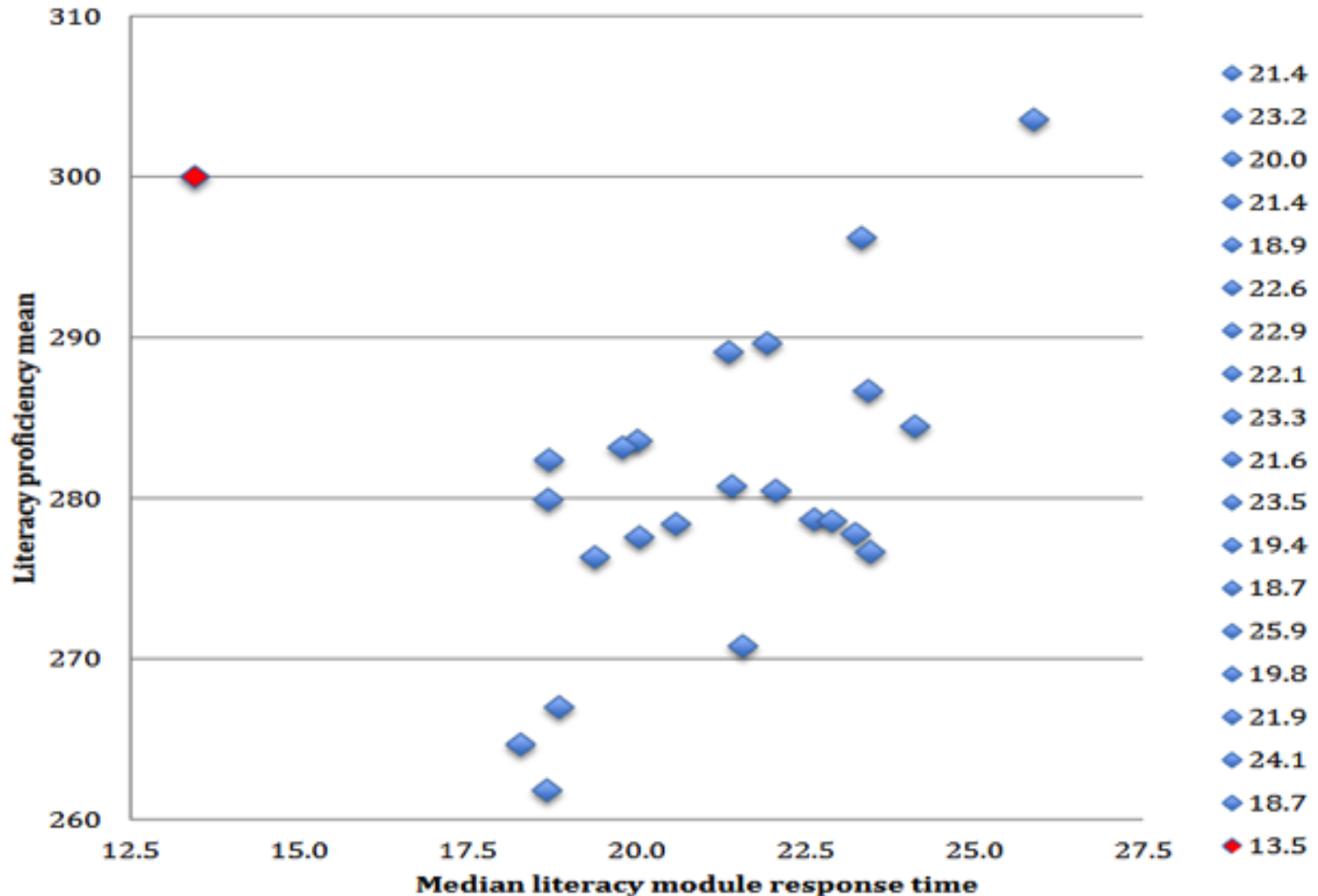
# Number of items with significant deviation from international parameters in Countries 1 and 8 in Language A

	<b>Country 1</b>	<b>Country 8</b>
<b>Literacy</b>	<b>14</b>	6
<b>Numeracy</b>	<b>17</b>	3
<b>Problem solving</b>	<b>3</b>	0
<b>Total</b>	<b>34</b>	9

# Mean literacy proficiency of respondents by interviewer



# Mean CBA Literacy Proficiency and Median Response Time to Literacy Module by Country

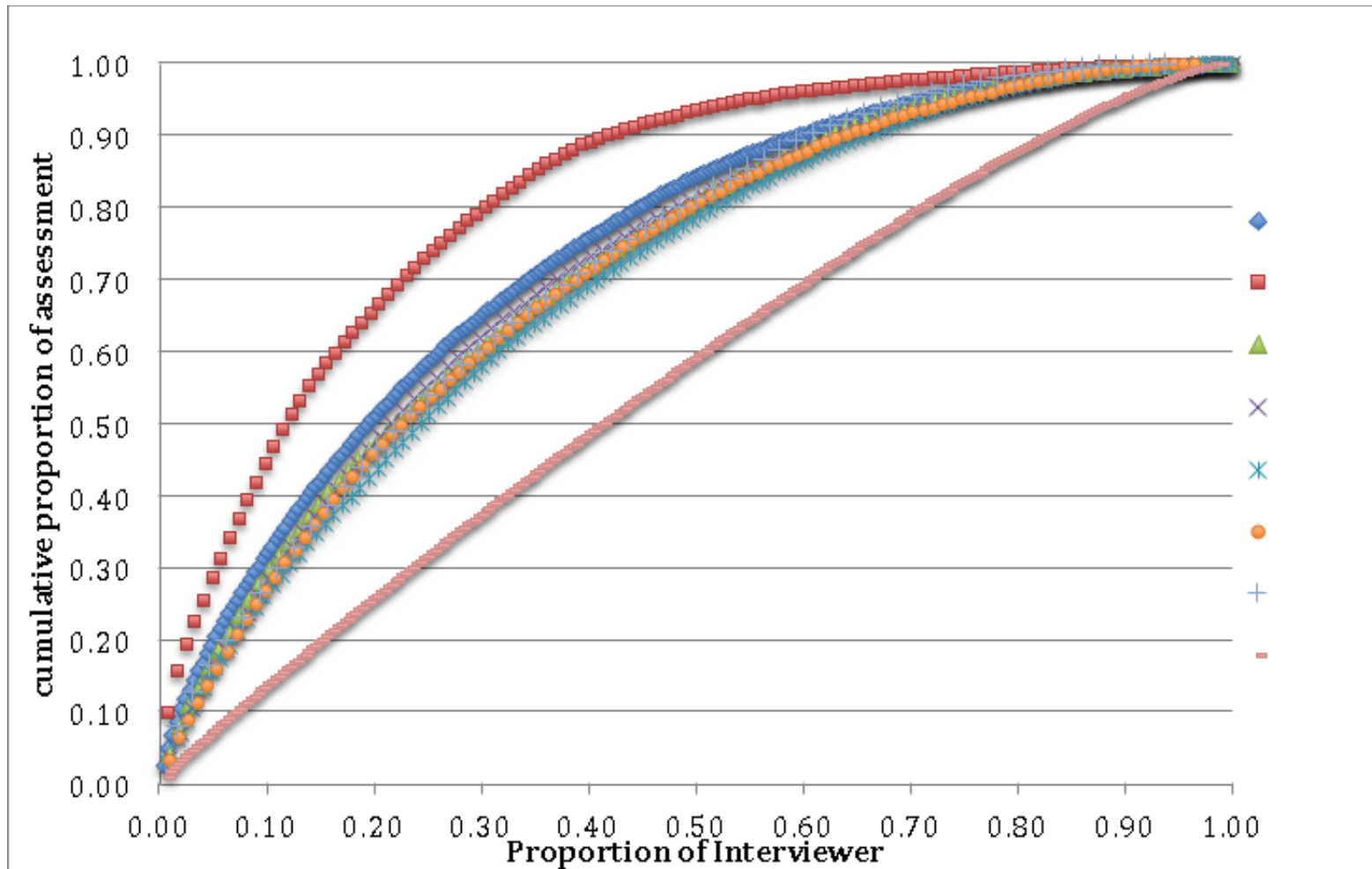


# Summary

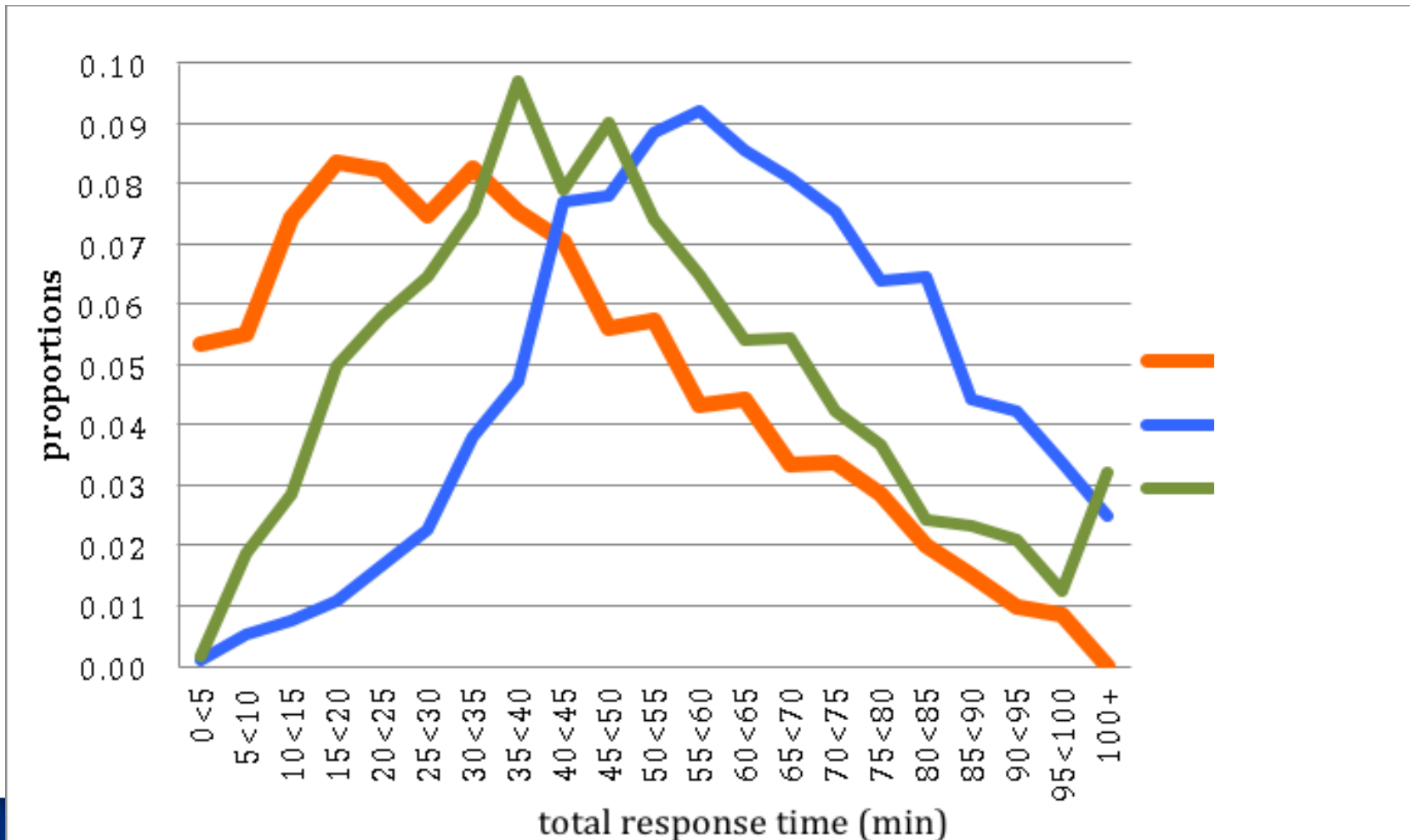
- Duplicated data
- Short response time
- Lack of variance of performance data within interviewer
- Geographically localized anomaly
- Incongruity of responses within a respondent
- Lack of overall comparability of data within a country
- Lack of comparability of data across countries

Resulted in elimination of data from entire region, respondents with average response time less than 10 seconds, and all duplicated cases. Altogether 1220 cases were eliminated.

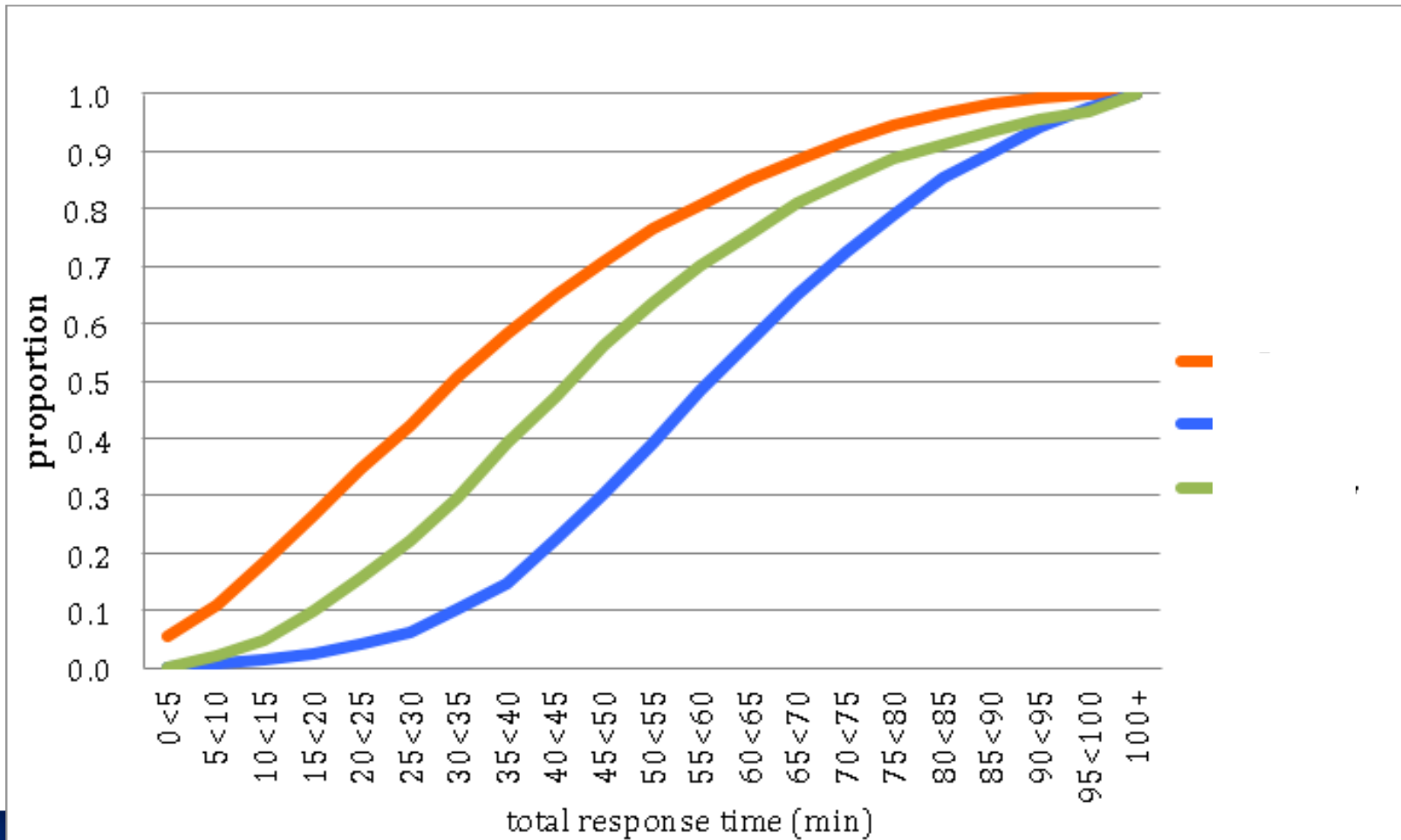
# Cumulative Proportion of Assessment by Interviewer



# Distribution of Respondents Based on the Total Time on CBA Cognitive Items (min)



# Cumulative distribution of respondents based on the total time on CBA cognitive items (min)





# 144 Responses collected by the Interviewer #92

Core

Literacy

Numeracy

144 respondents

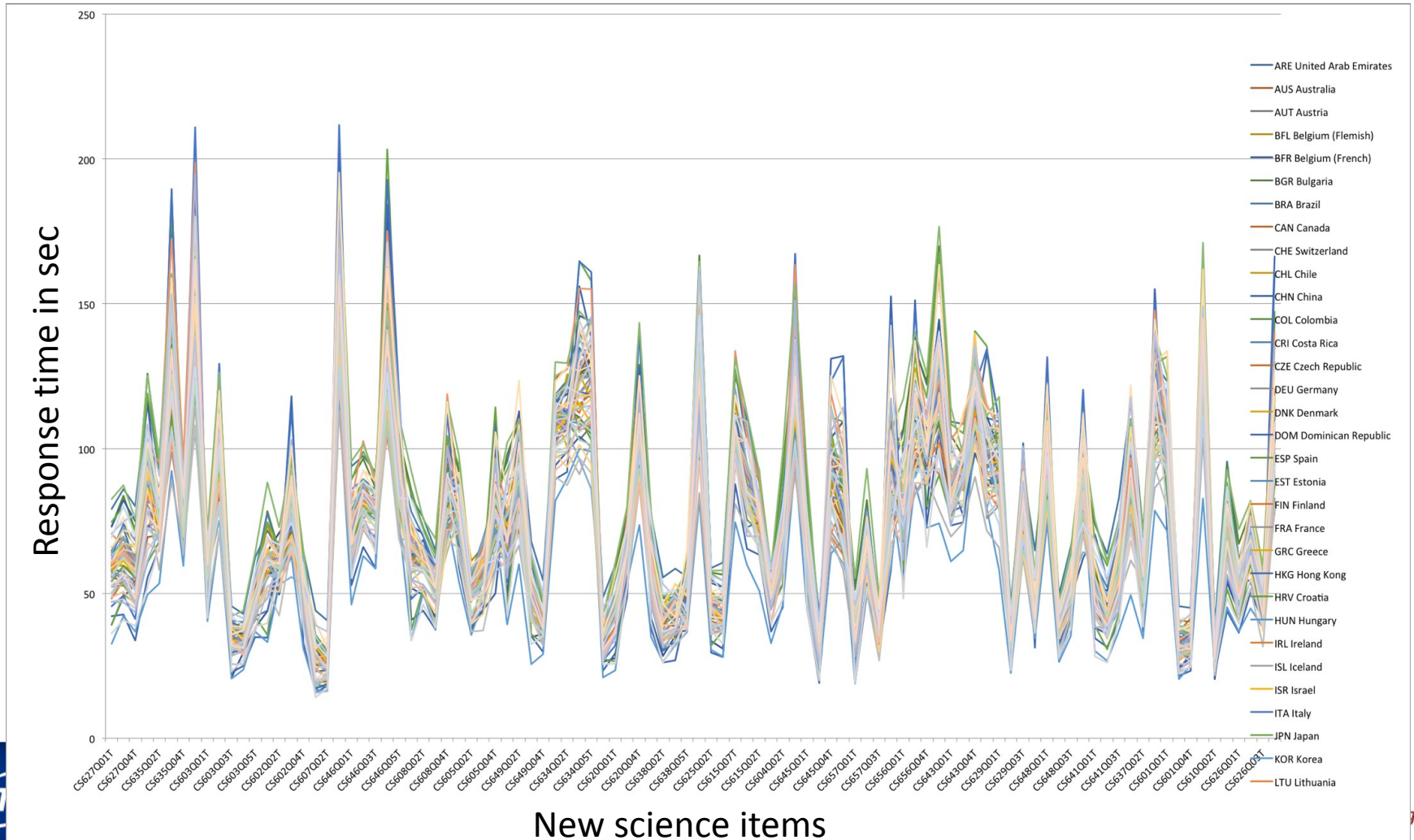
The table displays a grid of 144 rows (respondents) and multiple columns (items) grouped into three sections: Core, Literacy, and Numeracy. Each cell in the grid is color-coded: green indicates a correct response, red indicates an incorrect response, and black indicates a response was omitted. The grid shows a high density of correct responses (green) across all sections, with some scattered incorrect (red) and omitted (black) responses.

# Summary

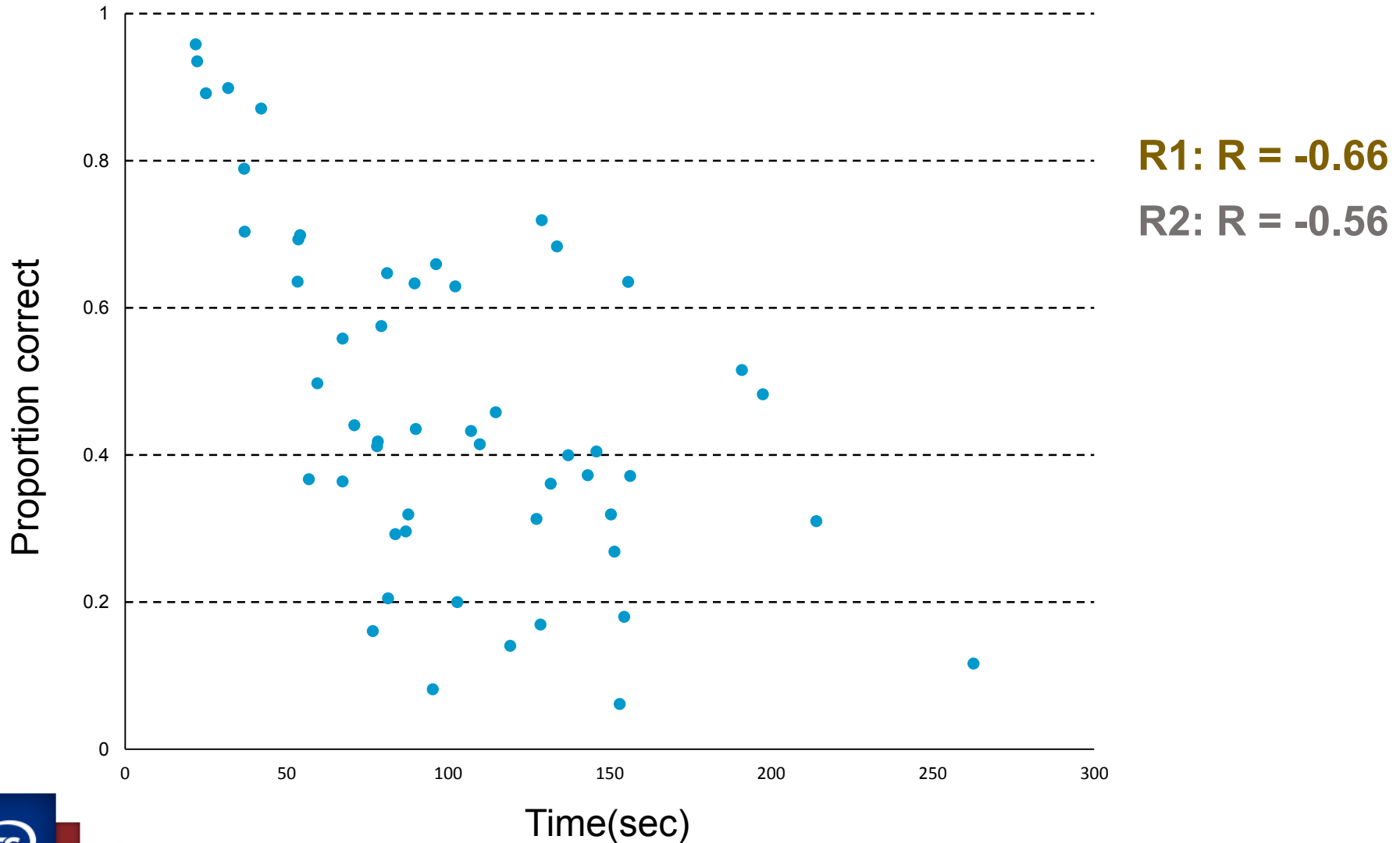
- 20% of data was collected by 6 interviewers
- Short response time interacted with interviewer ID
- Lack of variance of performance data within interviewer
- Large number of non-response within a few interviewers
- Incongruity of responses across domains within a respondent
- Lack of overall comparability of data within a country
- Lack of comparability of data across countries
- BQ data did not show anomaly

Resulted in elimination of cognitive responses of 1042 cases collected by 7 interviewers.

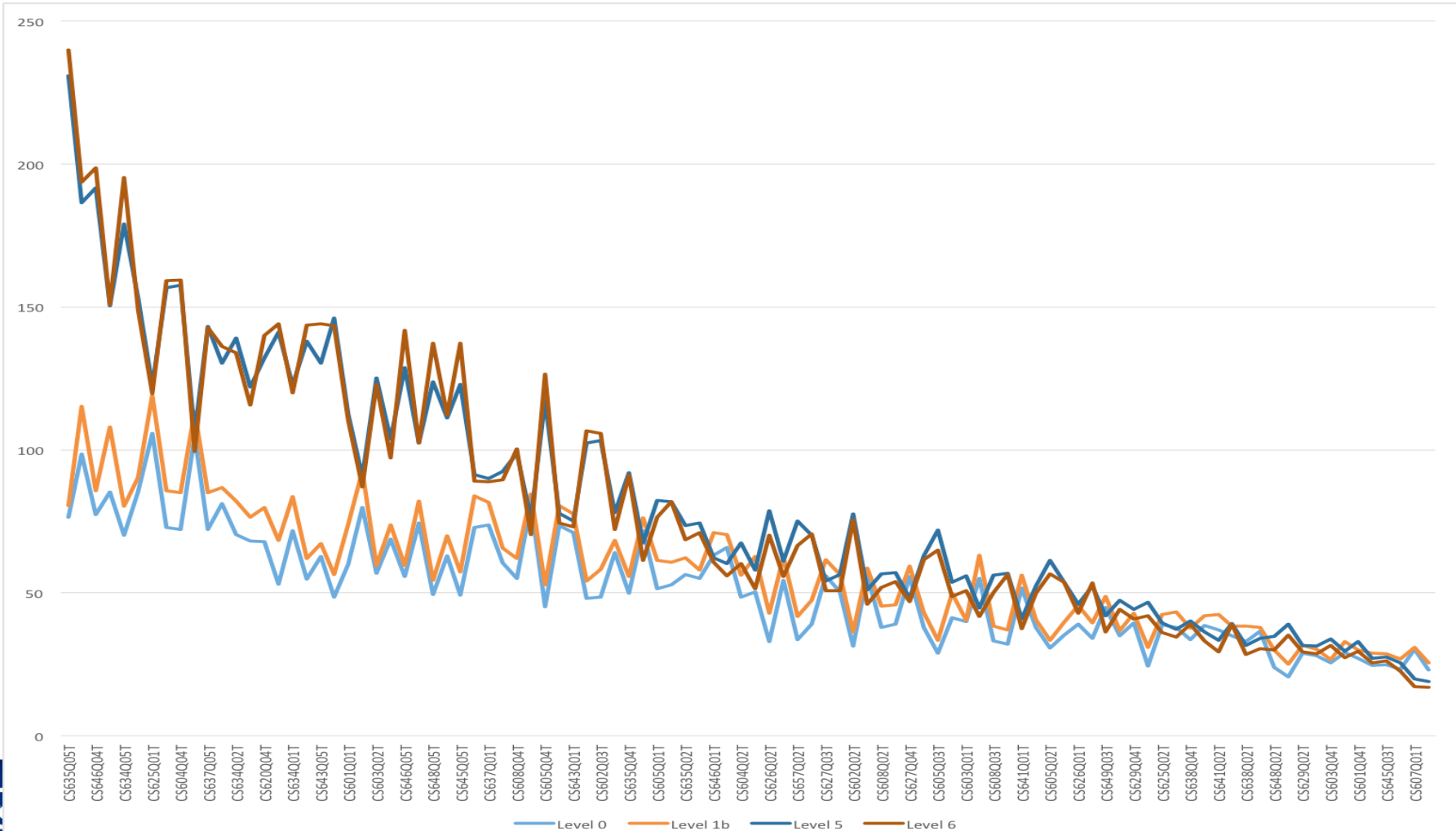
# Mean Response Time by Item – SCIENCE New



# Relationship of Timing and Difficulty of Literacy Items (PIAAC 1&2)



# Average Response Time by Skill Level (PV1) – SCIENCE New Items

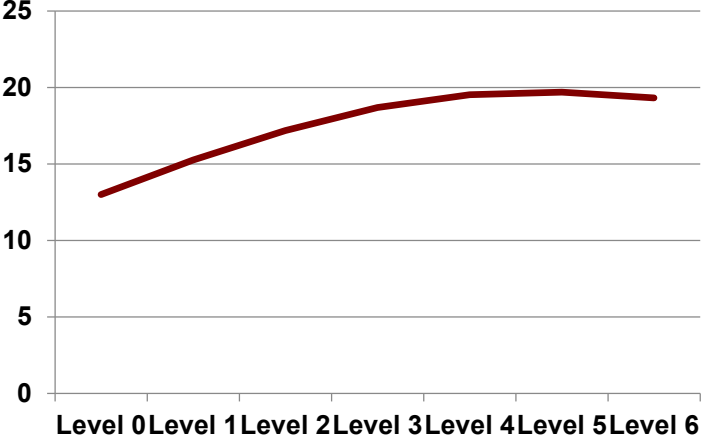


# Cluster Level Response Time by PV1 Skill Level (min)

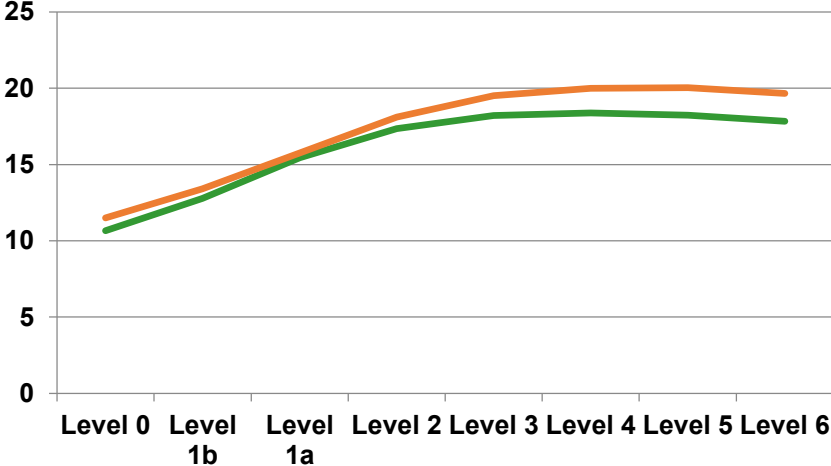
	Level 0	Level 1		Level 2	Level 3	Level 4	Level 5	Level 6
<b>Mathematics</b>	13.0	15.3		17.2	18.7	19.5	19.7	19.3
<b>Reading</b>	10.7	12.8	15.5	17.4	18.2	18.4	18.2	17.8
<b>Science</b>	11.5	13.4	15.8	18.1	19.5	20.0	20.0	19.7
<b>CPS</b>	18.5	20.1		21.3	21.9	21.8		

# Cluster Level Response Time by PV1 Skill Level (min)

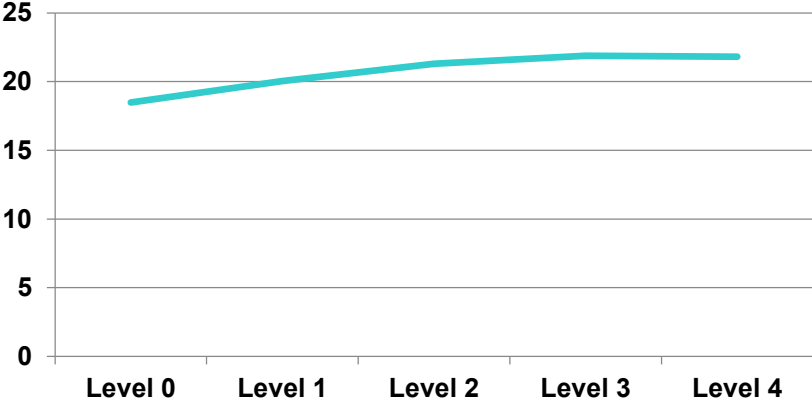
**Math**



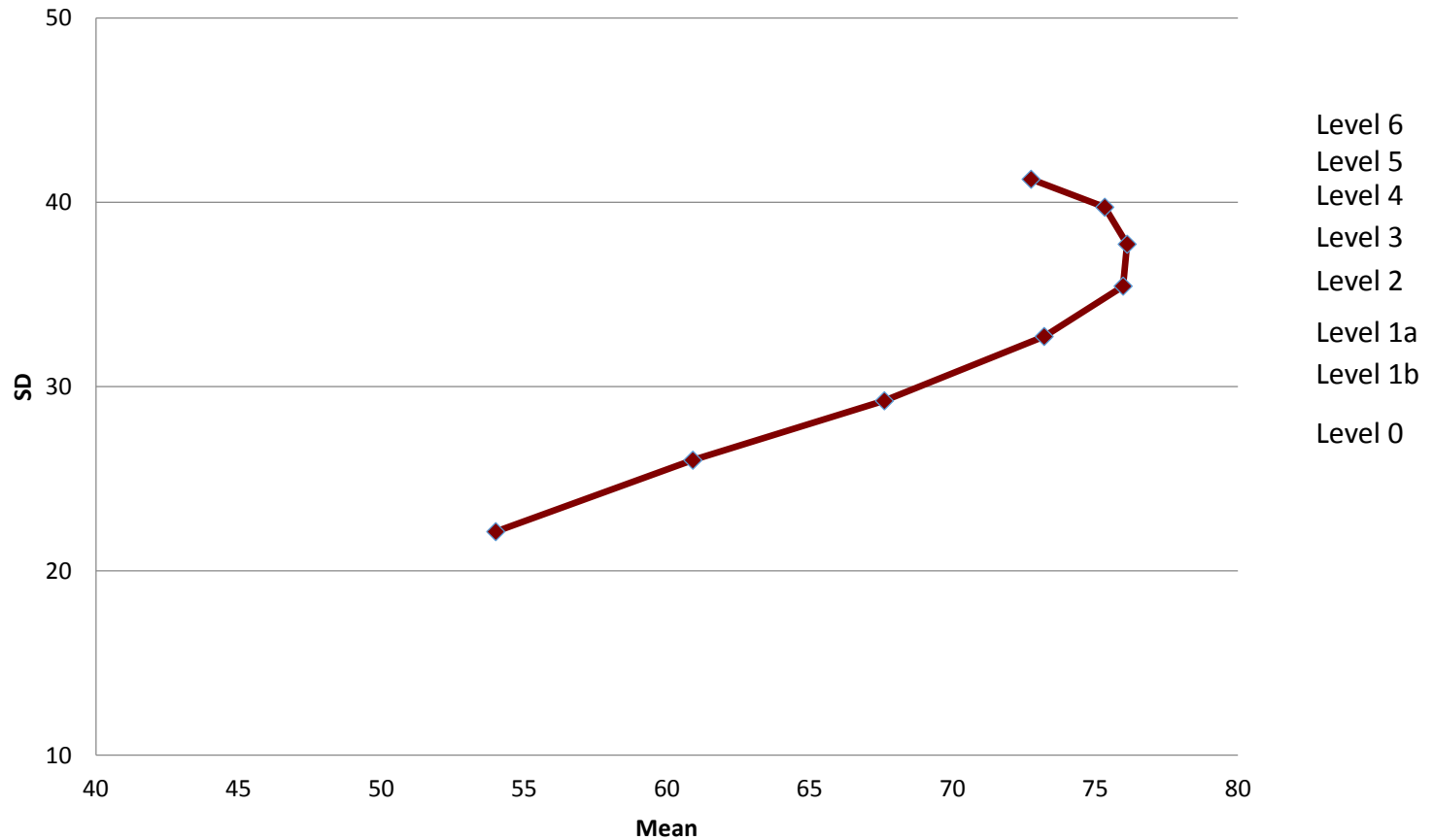
**Reading Science**



**CPS**

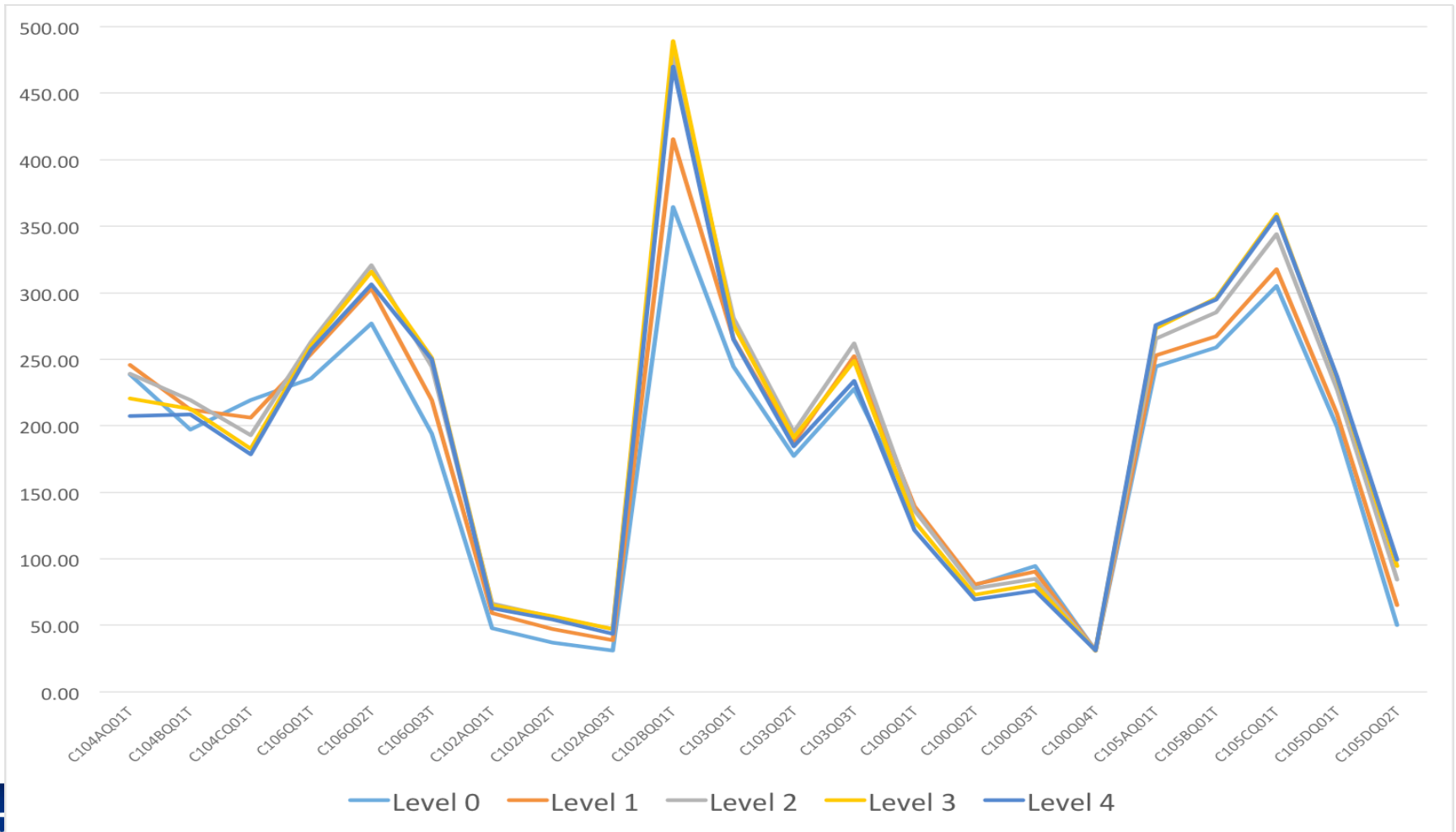


# SD and Mean Time (sec) of SCIENCE Trend Items by Skill Levels (PV1)

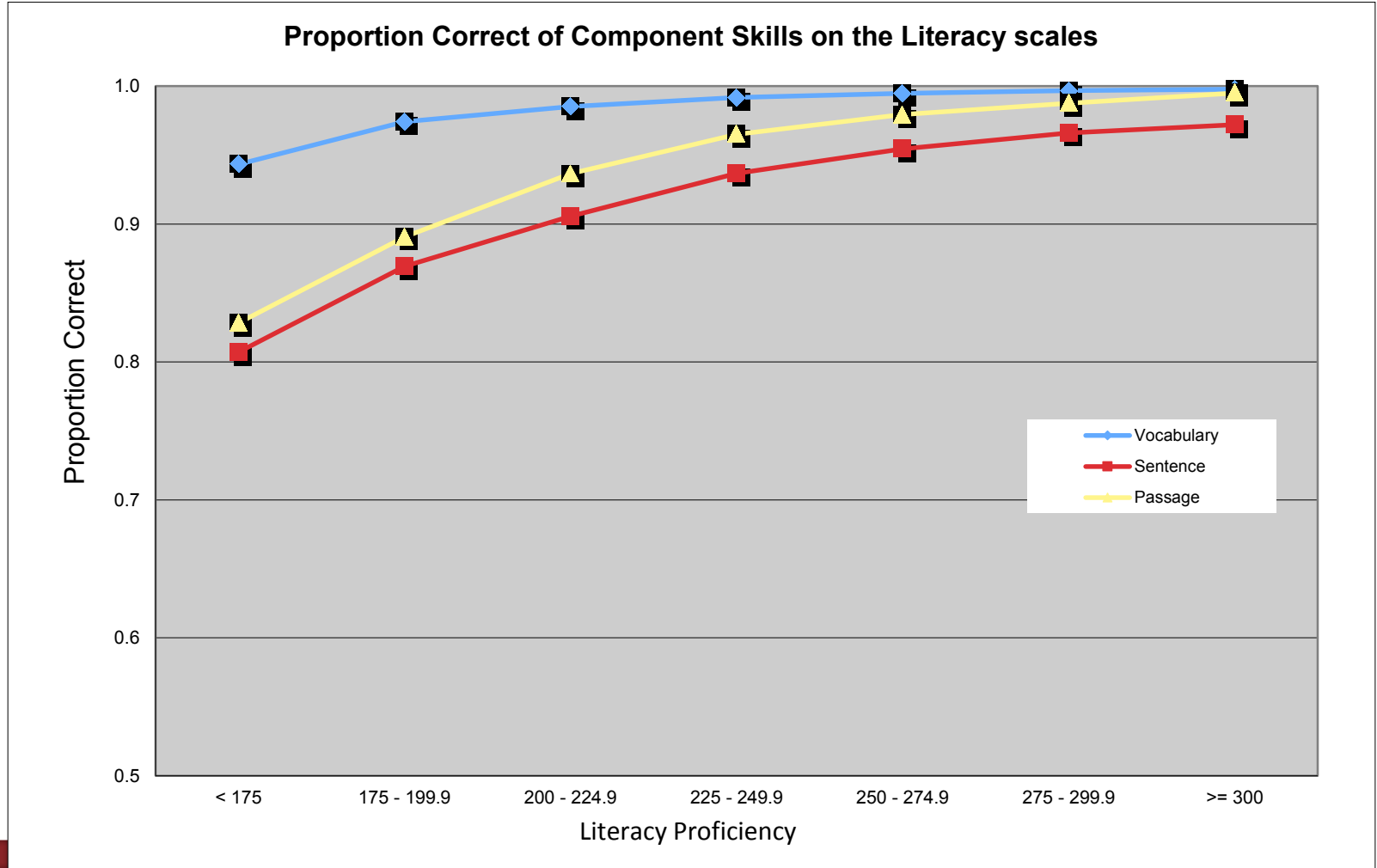




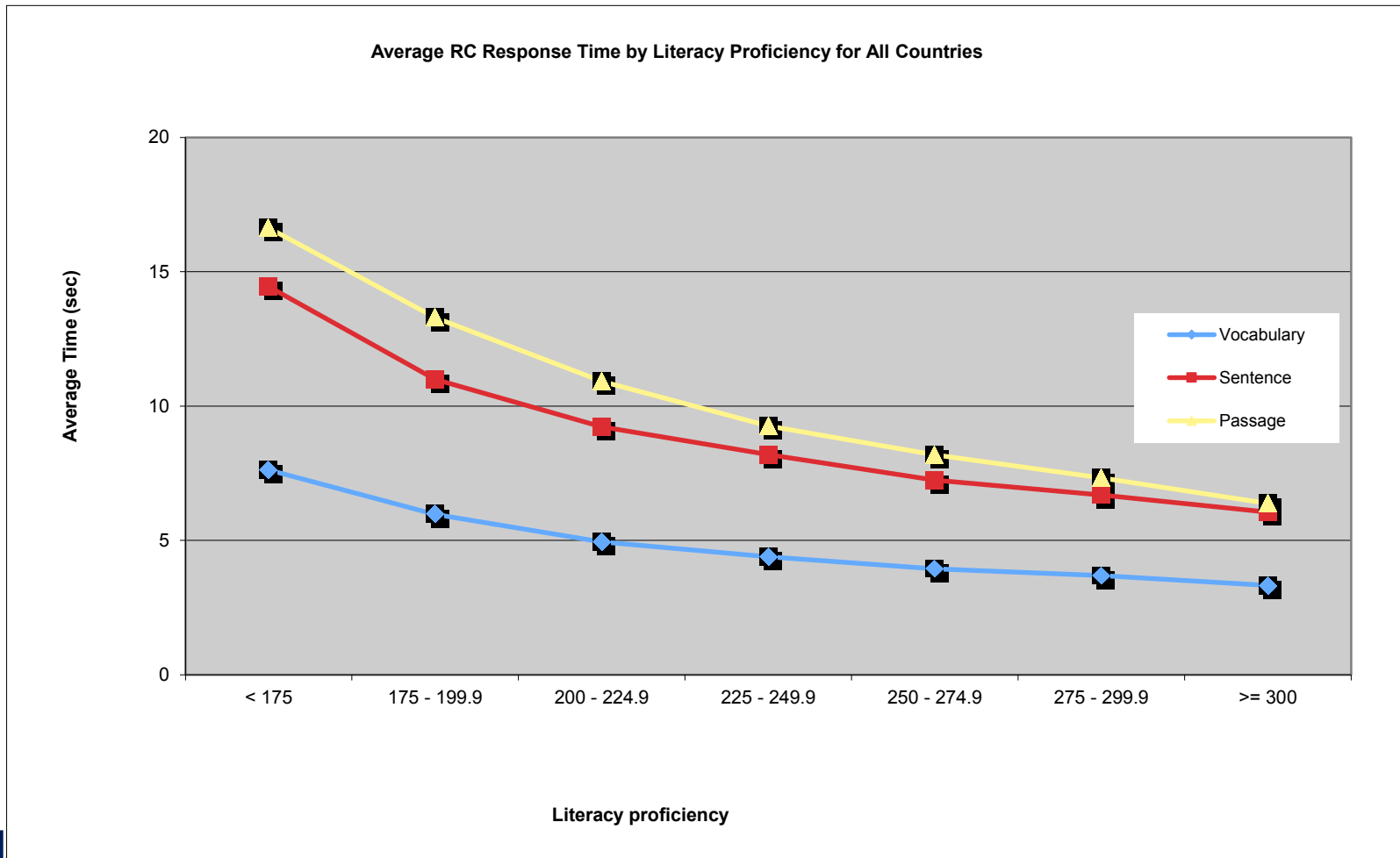
# Average Response Time by Skill Level (PV1) – CPS Items



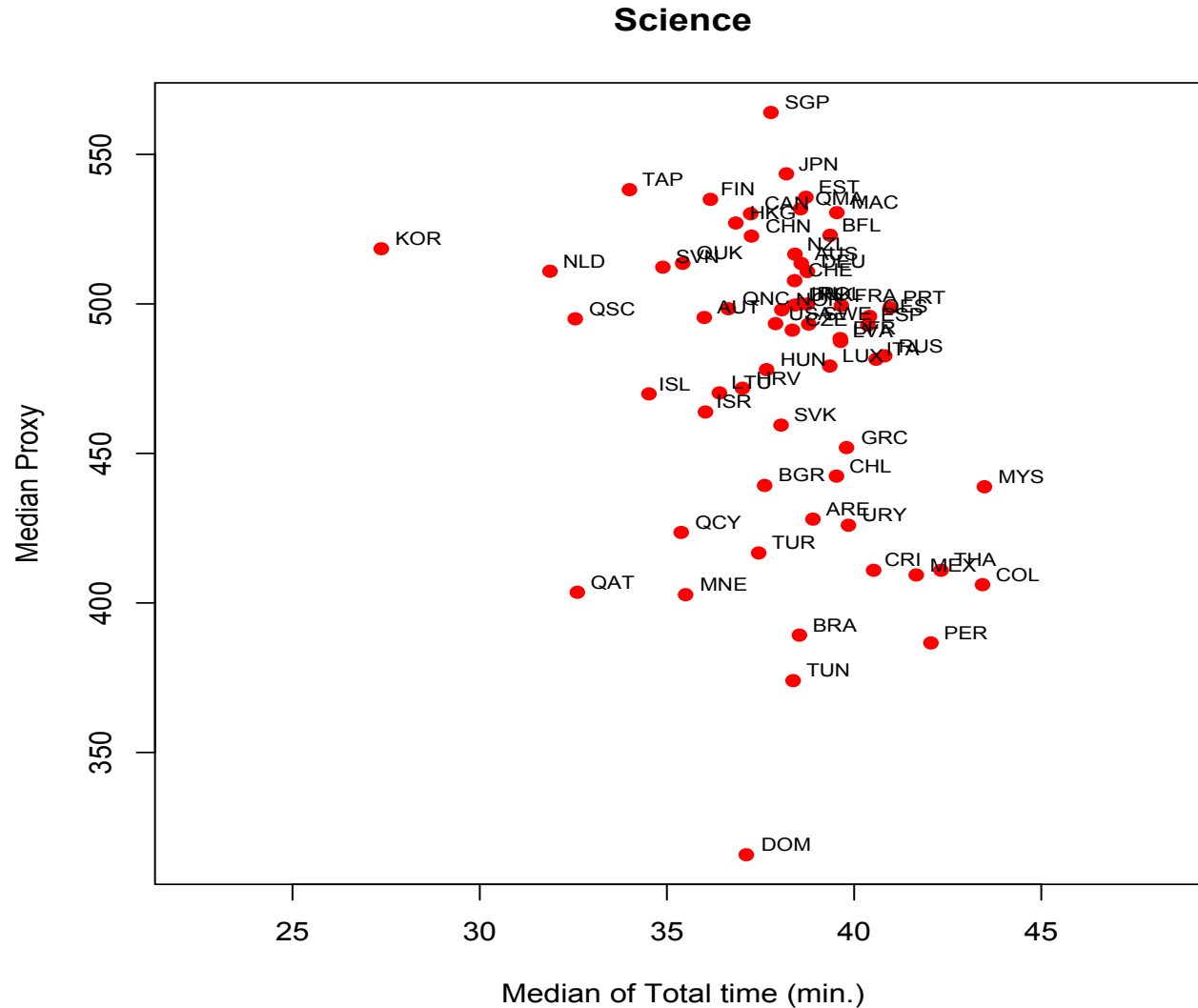
# Reading Components Accuracy on the Literacy Scale – PIAAC(1&2) countries



# Reading Components Response Time on the Literacy Scale – PIAAC(1&2) countries

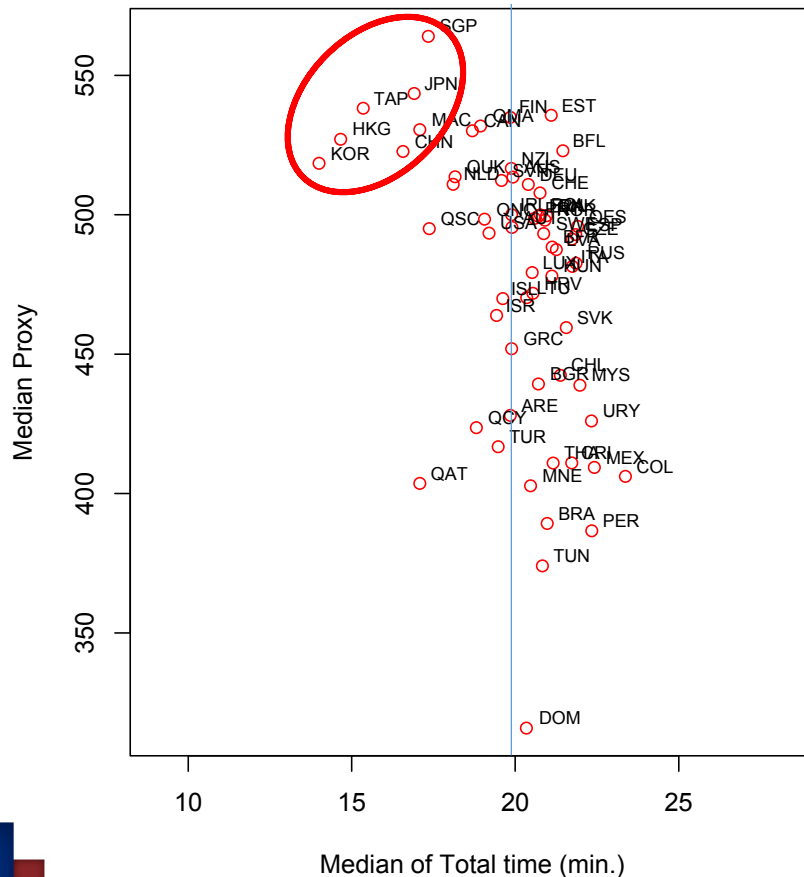


# Country Mean Proxy(WLE) by Median Response Time

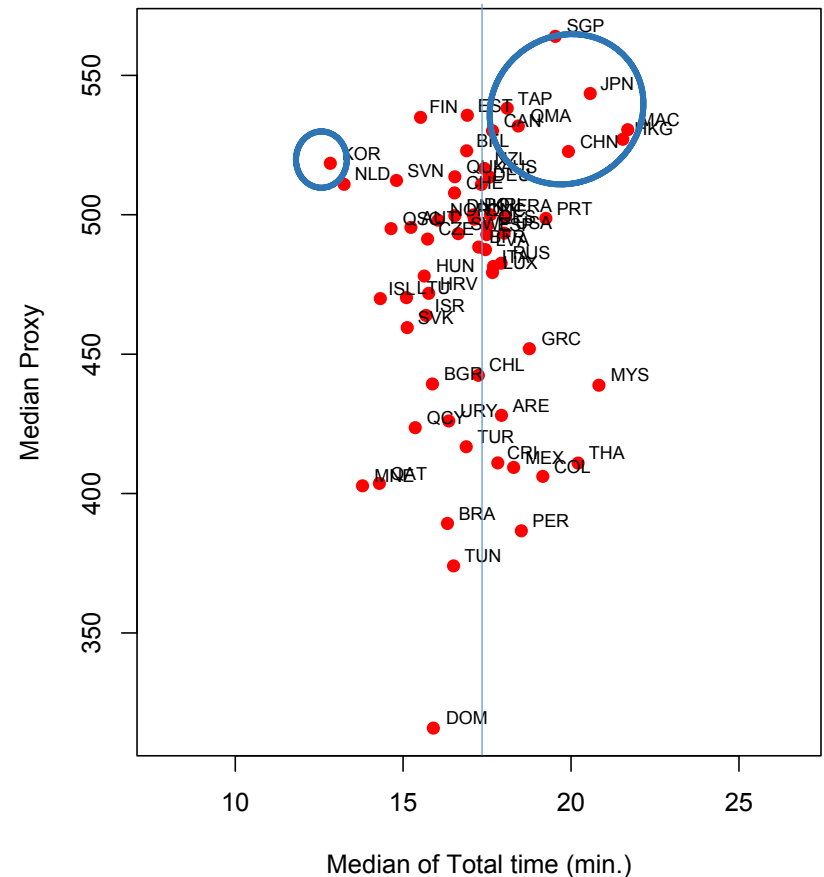


# Country Mean Proxy(WLE) by Median Response Time by Type of Items

Science (MC items)

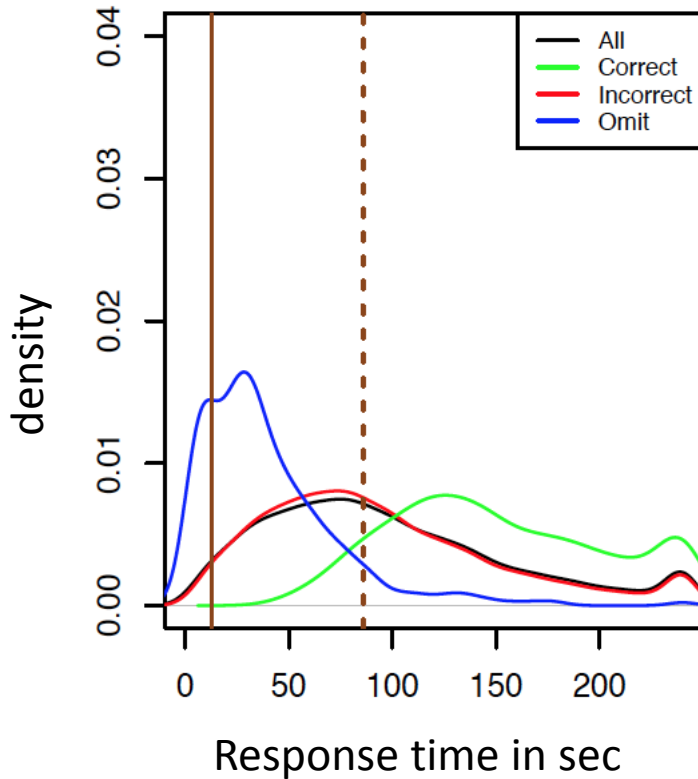


Science (HC items)

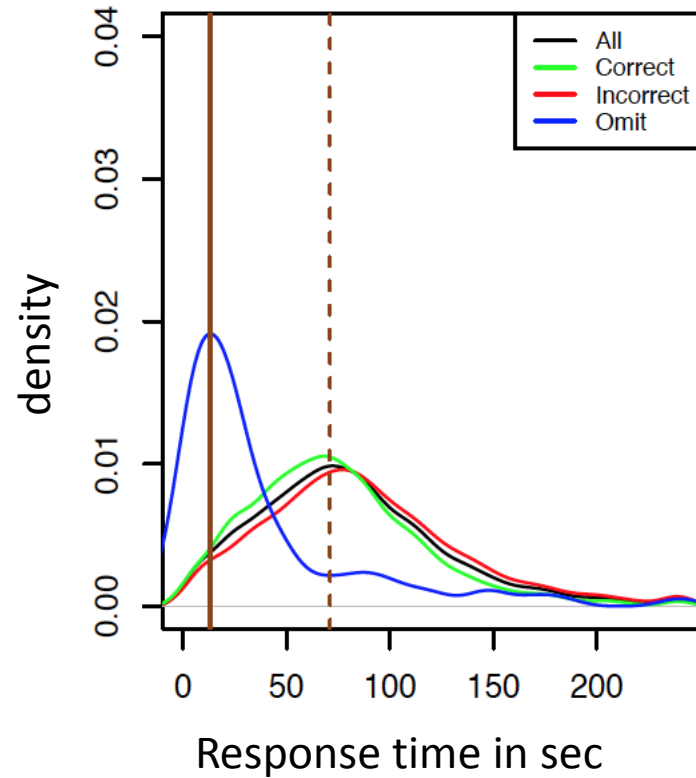


# Response Time Distributions by Response Type

S646Q05



S510Q01



# Summary of Findings

- Within-country variability across items is much larger than across-country variability
- Countries are more varied in median time for Human Coded items
- Less variability in Math across proficiency levels compared to Reading and Science
- Item response time interact with item demands and students' ability and can be opposite depending on the amount of cognitive demand
- Remarkable regularity of response time by items across countries on some items

# How to use Response Time for Psychometric Modeling

- Challenges to incorporate all interactions to settle on one model for all domain and item types.
  - Response Time interact with assessment domain
  - RT interact with language of assessment
  - RT interact with cognitive demand of items
  - RT interact with proficiency skills uniquely



Thank you

