



PEARSON

Jo Fond Lam (CINOP)

Ryan Downey (Pearson)

The validity of the reading comprehension test for immigrants: the Dutch case

Outline

- Background: Dutch policy and social integration
- Test Development
- Validity
 - Literacy & Reading Comprehension
 - Reliability, machine scoring
 - Construct relevance
- Discussion

Background



Dutch policy

Before 1995:

- No policy in relation to integration or citizenship

1998 Law on integration of Newcomers:

- Responsibility of the Dutch society
- Obligatory lessons of Dutch language and knowledge of Dutch society
- No exams but certificates

2006 Law on integration:

- Responsibility of the migrant
- Social Integration exam includes test of spoken Dutch and test of Dutch society/culture
- Social Integration exam in home country and in the NL

Social integration 2011

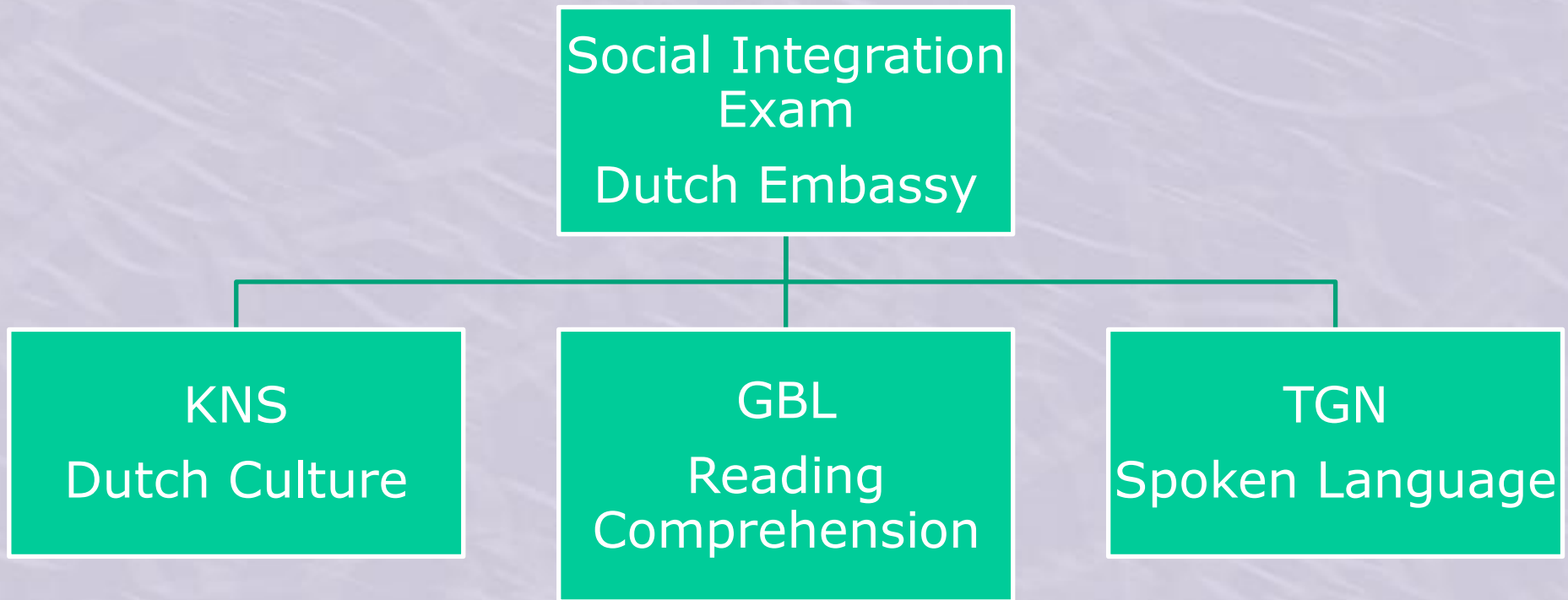
2011 Law requires (basic) reading skills of newcomers

- Social Integration exam includes tests of Speaking, Culture, and Reading

Dutch government encourages creating individual conscience:

- “What does immigration really mean?”
- Acquire basic knowledge about the Dutch society and language
- Goal is improving integration, not exclusion!

Social Integration exam in home country



Social integration exams in home country

Continents	%	Top 3 Countries
Africa	30%	53% Morocco, 12% Ghana, 11% Egypt
Australia	0%	50% Australia, 50% New Zealand
Asia	45%	15% Thailand, 14% Turkey, 10% Indonesia
Europe	15%	20% Russia, 13% Ukraine, 10% Serbia & Montenegro
North-America	3%	38% Mexico, 16% Dominican Republic, 10% Cuba
South-America	7%	44% Brazil, 17% Suriname, 13% Colombia
Total	100%	

Outline

- Background: Dutch policy and social integration
- Test Development
- Validity
 - Literacy & Reading Comprehension
 - Reliability, machine scoring
 - Construct relevance
- Discussion



PEARSON



Test Requirements for KNS, TGN and GBL

Practical

- Available 24 x 7 x 365
- Available in any country with an Embassy
- Administered by embassy staff with minimal training
- Immediate results
- Secure
- Delivered over phone
- No additional equipment (keyboard, mouse, monitor, etc.)

Other requirements

- Referenced to Common European Framework of languages
 - (Speaking and Reading)
- Reliable, fair, free of bias, validity arguments

Test Requirements

Mandate by Government of the Netherlands:

A new test of basic literacy and reading comprehension

Reading CEFR descriptors at Level A1

A1	Candidate can understand familiar names, words and very simple sentences, e.g., on notices and poster or in catalogues
----	--

Reading comprehension and oral responses

- All items written to target reading comprehension:
 - below A1
 - at A1
 - above A1
- Dutch as a foreign language instead of second language
- 2000 most frequent lemmas in Corpus Gesproken Nederlands
- Simple constructions (simple past, present)

Dutch Reading Test (GBL)

Receive test paper

Call testing system and respond orally to all items into the phone

Section	Task	# items
A	Word rows	4
B	Sentences	8
C	Short Passages	3
D	Fill-in-the-gap sentences	28
E	Passage comprehension	12
	TOTAL	55

Example: Word row

Voorbeeld Woordrij A
en
de
ook
bal
wil
zee
ham
boom

Example: Sentence

Voorbeeld	
Zin A	
A	Mijn zoon is een lieve jongen.

Example: Short passage

Voorbeeld

Tekst A

Tamar gaat naar de tandarts.

Zij heeft pijn aan haar kies.

Tamar is bang voor de tandarts.

De tandarts kijkt in haar mond.

Tamar heeft gelukkig alleen maar een gaatje.

Ze is blij als de tandarts klaar is.

Ze heeft nu geen pijn meer.

Example: Fill-in-the-gap sentence

Voorbeeld		
Zin A		
A	Het is mooi weer. De zon ...	loopt regent schijnt

Example: Passage comprehension

Voorbeeld

Tekst A



Mevrouw Jansen wacht samen met mevrouw Bol op de bus naar de stad. Er is vandaag markt in de stad. Elke week gaan mevrouw Jansen en mevrouw Bol samen naar de markt. Wat gaan ze vandaag kopen? Niets. Ze gaan naar de markt, omdat het leuk is. Dat doen ze elke woensdag.

Vragen

- Op welke dag gaat mevrouw Jansen naar de stad?
- Hoe gaat mevrouw Jansen naar de stad?
- Met wie gaat mevrouw Jansen naar de stad?

Scoring traits of GBL

Trait	Interpretation	Scoring
Accuracy	Did the candidate read each word in context well enough to be understood as the target?	Candidate ability estimate (in logits) based on the # word errors
Rate	How quickly did the candidate read correctly?	# number of words read correctly / time to read X 60 aka, words correct per minute
Expressiveness	How expressively did the candidate read, on a 6 point scale?	Candidate ability estimate (in logits) based on models from human scores
Comprehension	Did the candidate produce the correct answer?	Candidate ability estimate (in logits) based on correct/incorrect

All subscores returned on a 10-40 scale and then combined into a final Overall score

Scoring traits of GBL

Task	Accuracy	Expressiveness	Rate	Compreh.
Word rows	X			
Sentences	X	X	X	
Short passages	X	X	X	
Fill-in-the-gap sentences	X		X	X
Passages with short answer questions				X

Outline

- Background: Dutch policy and social integration
- Test Development
- Validity
 - Literacy & Reading Comprehension
 - Reliability, machine scoring
 - Construct relevance
- Discussion

Data

Field Test Participants

- 1741 Dutch as second language learners (=NT2-ers)
- 125 native languages
- 78 candidates at 3 Embassies
- 202 native speakers of Dutch

Validation Studies (set aside from field test population)

- 160 NT2-ers in NL
- 52 candidates at Embassies

Trait-level correlations

Score	Accuracy	Express.	Rate	Compreh.
Overall	0.90	0.87	0.88	0.95
Accuracy		0.77	0.81	0.79
Expressiveness			0.89	0.70
Rate				0.70

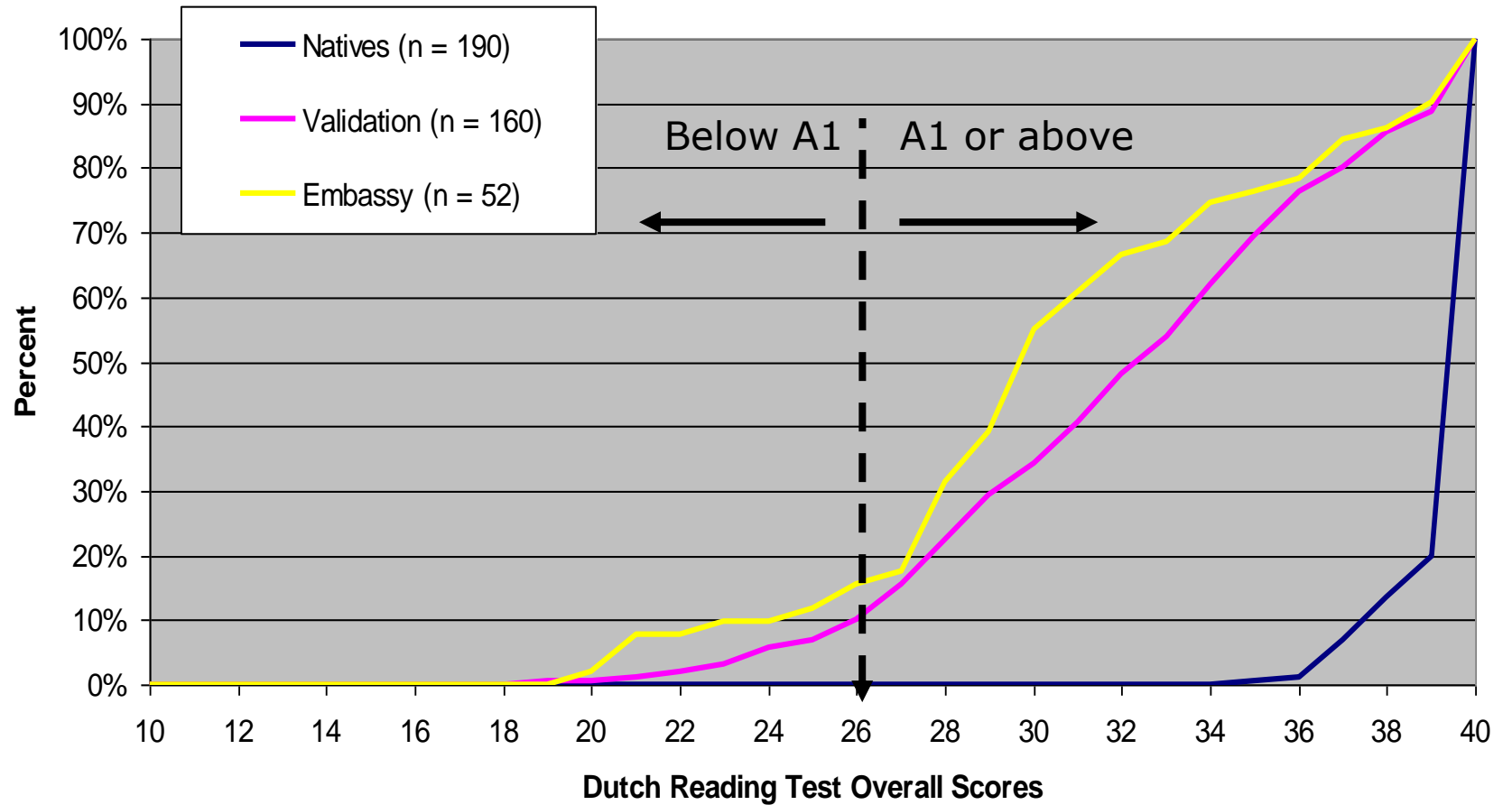
Reliability: human-machine

Aspect	Split-half reliability		Machine-Human Correlation
	Human	Machine	
Overall	0.93	0.94	0.95
Accuracy	0.95	0.91	0.89
Expressiveness	0.94	0.94	0.84
Rate	0.95	0.96	0.98
Comprehension	0.93	0.89	0.91

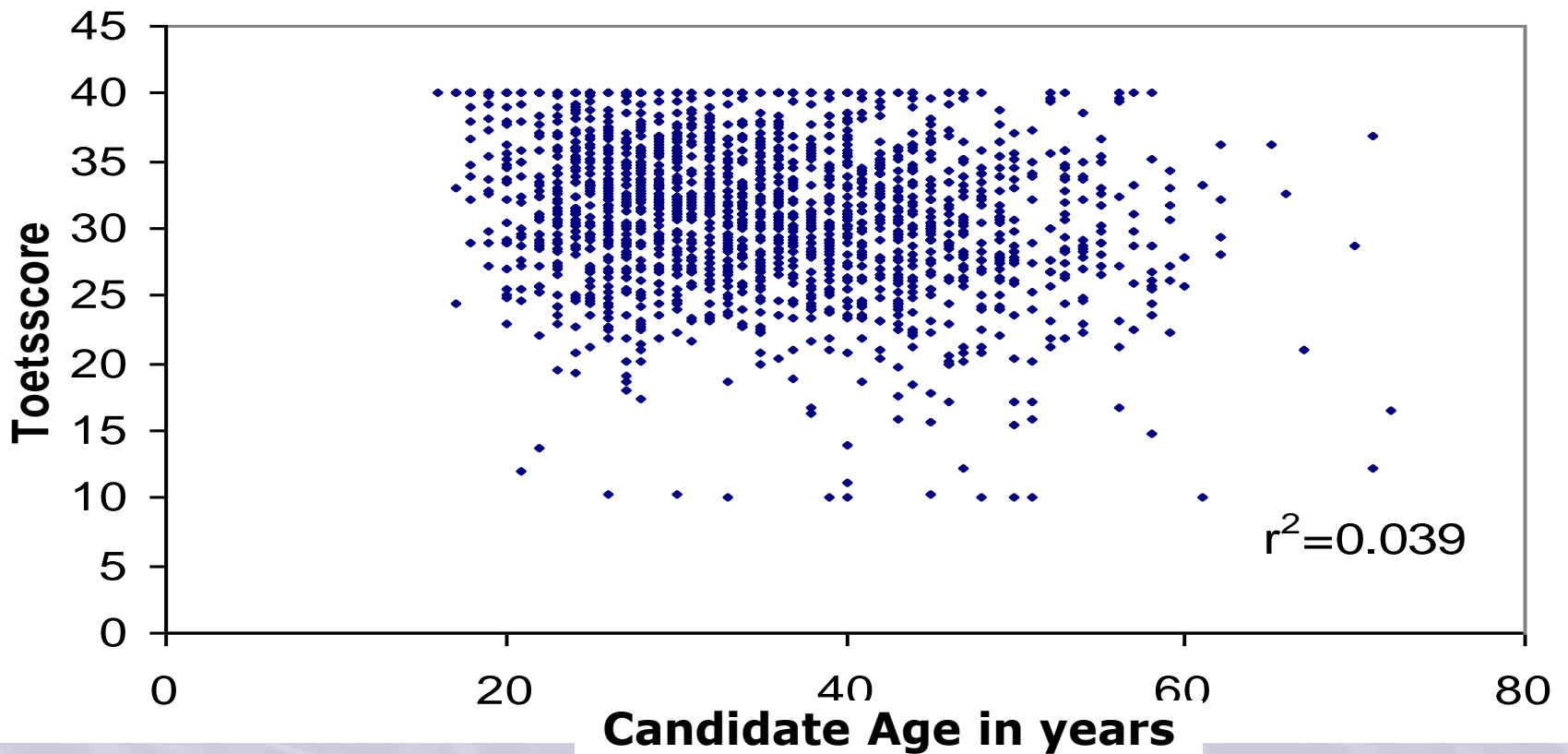
Reliability: human-machine

Aspect	Human-machine		
	April 2011 (n = 25)	January 2012 (n = 25)	Validation (n=160)
Overall	0.94	0.95	0.95
Accuracy	0.89	0.88	0.89
Expressiveness	0.93	0.95	0.84
Rate	0.95	0.95	0.98
Comprehension	0.92	0.93	0.91

CDF of scores from different groups

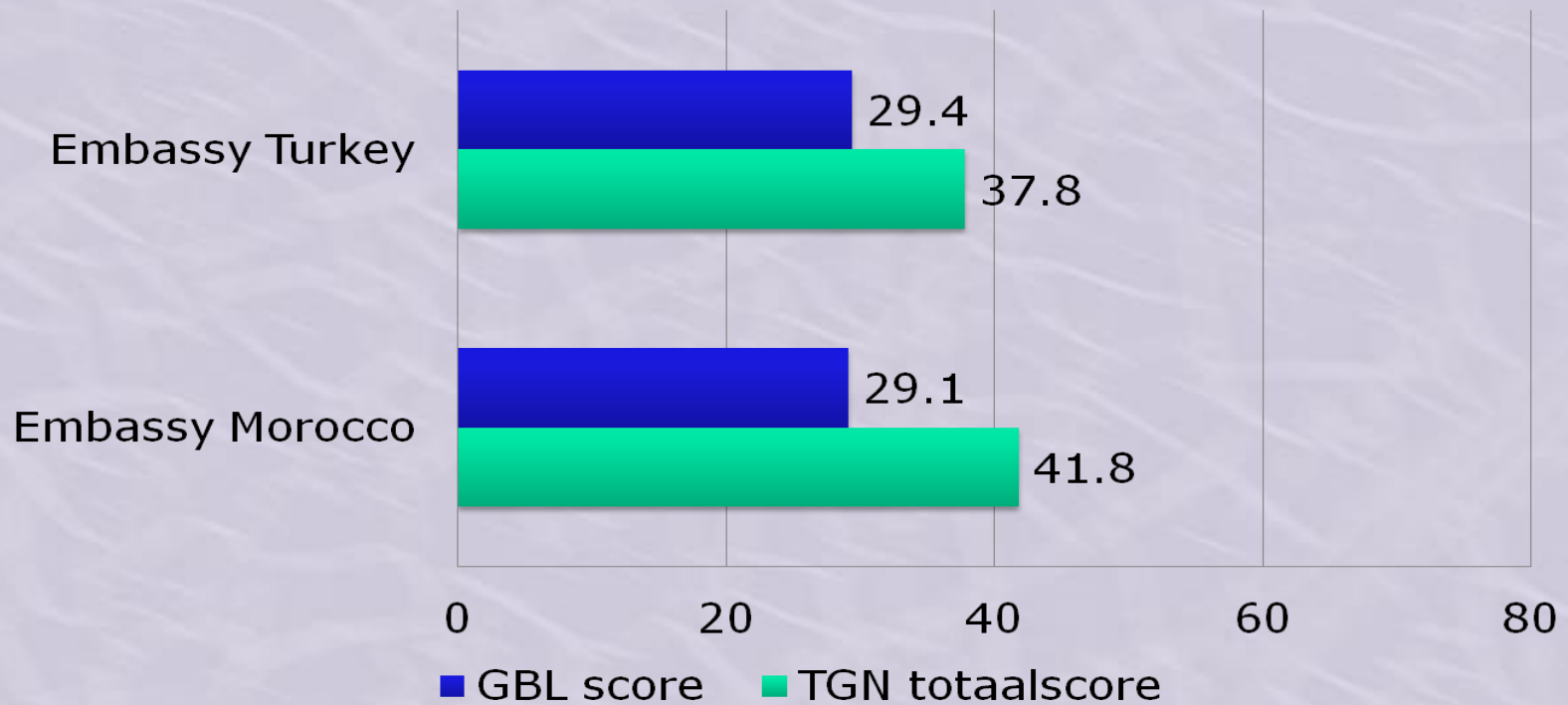


Age



Embassy validation set (n = 52):
Reading Test (blue) and **Speaking Test** (green)

Embassy validationset: GBL and TGN score



Construct-irrelevant influences

Shadow transcription study (N=53)

Score	Pronunciation-based Transcriptions	Machine Scoring	Traditional Transcriptions
Overall	30.7	31.3	31.9

Outline

- Background: Dutch policy and social integration
- Test Development
- Validity
 - Literacy & Reading Comprehension
 - Reliability, machine scoring
 - Construct relevance
- Discussion

Discussion

Language Testing and...

Communication!

- Keep the dialogue open
- Learn from each other
- Education, cooperation, validation, communication...





Thank you!

Jo Fond Lam (jlam@cinop.nl)

Ryan Downey (ryan.downey@pearson.com)

The validity of the reading comprehension test for immigrants: the Dutch case