

Language Testing Research Colloquium
Denver, 20 March, 2009

In the ear of the beholder:
Dependence of comprehensibility on language
background of speaker and listener

Ryan Downey & Alistair Van Moere
Pearson Knowledge Technologies

Overview

- Context of the study
- Methodology
- Results
- Discussion

Call-centre Recruitment



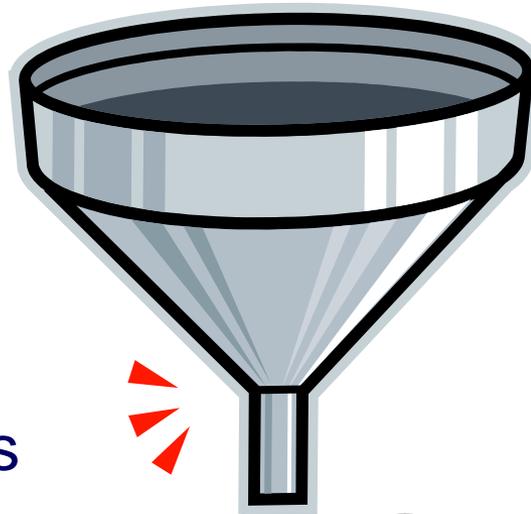
English proficiency

Personality

Cultural awareness

Negotiation skills

Computer skills



Traditional Approach

Oral Proficiency Interviews



Human examiners' ratings:

Grammar

Vocabulary

Pronunciation

Fluency

Automated Approach

 **VERSANT™**

Reading

Repeat Sentences

Short Answers

Sentence Builds

Story Retells



IRT-scaled scores:

Grammar

Vocabulary

Pronunciation

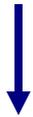
Fluency

Pronunciation construct

- In these interviews, pronunciation is assessed by listening to speech and evaluating to what extent the speaker has mastered the sounds of the language. Human raters produce scores with reference to a checklist.
- In Versant Tests, features of test-taker's speech signals - such as duration and spectral properties of phones - are evaluated with regard to a model of native-speaker productions, and scaled scores are produced.

Pronunciation construct

Interview Test



✓	went / vent
✗	three / sri
✗	egg / aeg
✓	live / leave
✓	apple / upple
	Retroflex 't'

Versant Test



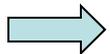
Linguistic Analysis

- Linguistic analysis of speech revealed that breakdowns in communication were due to traditional features of pronunciation:

- /r/ /l/ /s/ /sh/ /z/ /t/ /d/

- BUT also due to:

- Word-level stress; the (mis)articulation of every syllable
 - E.g. resident -- rest tent ; comfortable -- comforTABLE
- Sentence-level stress; rhythm
 - E.g. court documents -- court arguments
- Speech rate; fluidity



Native listeners conduct top-down processing on phonemic pairs, but they struggle with irregular supra-segmentation and incorrect stress at word and sentence level

Stakeholders



Agents' speech

Expert linguists in India:
-Examiners
-Voice trainers

Expert linguists in
the US

Customers in the US
Customers in the UK

Non-native speaker
customers

Previous research

- Brown (1995)
Testing for tour guides: industry workers were harsher raters than language instructors
- Yoshida-Morise (1998)
Japanese speaking examiners more tolerant of “Japanese-isms” in test-taker speech
- Xi (2008)
Indian raters performed comparably to ETS raters when judging both Indian and non-Indian responses

Comprehensibility

- Defined in this study as:
“The degree to which a speaker’s utterance was understood by the listener, and the degree of effort it took to understand the utterance”
- Judgment based on initial reaction to speaker’s utterance; relevant to the goals of the study in this test use domain

Research Questions

1. Do Indian expert linguists and US expert linguists have different assessments of Indian speech?
2. Do Indian expert linguists have different assessments of the comprehensibility of Indian speech than naïve listeners (customers)?

Participants: Candidates

- Spoken samples from 80 Versant English Tests
 - 20 speakers of “Indian English” dialects
 - 60 “Other L2 speakers of English” (Asia, Europe)

	Indian English (<i>n</i>)	Other English (<i>n</i>)
High	5	15
Med-High	5	15
Med-Low	5	15
Low	5	15

Participants: Raters

- 4 Groups of raters
 - 10 “US English Experts” (Native-speakers)
 - 10 “Indian English Experts” (Indian voice trainers)
 - 10 Naïve listeners (US)
 - 5 Naïve listeners (UK)

Total Data = 72,345 judgments

35 raters * 8 speech samples * 80 candidates * 3 traits
> 30 seconds of total speech per candidate

Procedure

- Raters were screened and recruited
- Raters worked from home and listened to language samples on the phone or computer
- Raters were given a small set of language samples for self-norming
- Submitted ratings according to criteria, 0 - 6

Comprehensibility

- Listen to the speech sample once
- Rate how much of a speaker's utterance you understood and how much effort it took you to understand it

4	Good Comprehensibility I understood everything in the utterance but it required some time and effort for it to make sense.
2	Poor Comprehensibility I missed many key words and I would have needed to ask for clarification/repetition to understand

Pronunciation

- Ability to produce consonants, vowels and stress in a native-like manner in sentence context; also, knowledge of the phonological form (phonemes and lexical stress) of words

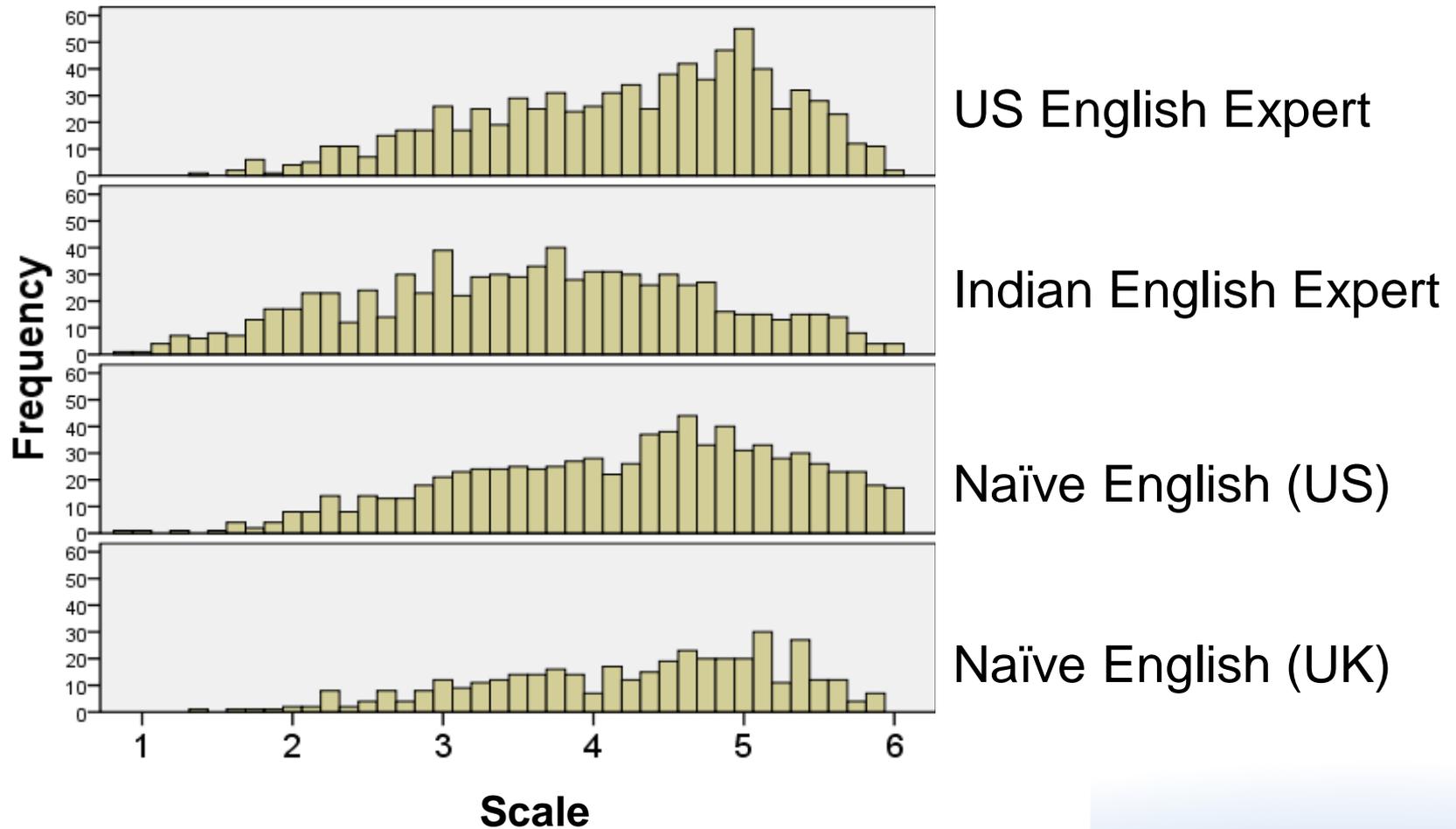
4	Good Pronunciation; but clearly non-native Most consonants and vowels are pronounced correctly. Some consistent errors might render a few words unclear. A few consonants and vowels may be distorted, omitted or incorrectly stressed.
2	Intrusive Pronunciation Many consonants and vowels are mispronounced, resulting in a strong intrusive foreign accent. Stress placement is unclear; unstressed vowels may be distorted or omitted; a few syllables may sometimes be added or skipped.

Fluency

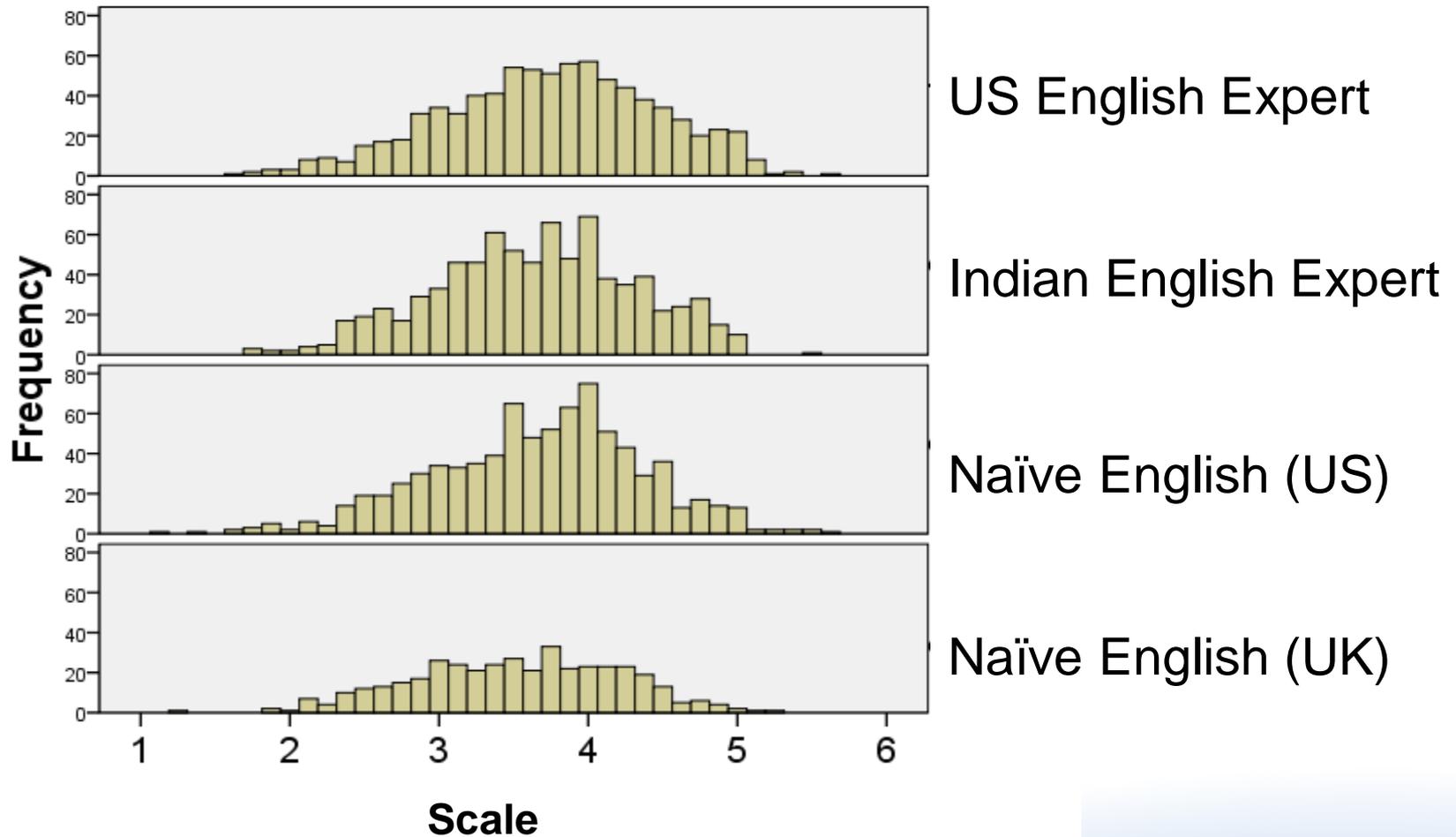
- Smooth, rapid speaking; evidenced by appropriate rhythm, phrasing, pausing, and word emphasis in continuous speech

4	Good Fluency Speech has acceptable speed, but may be uneven. Long utterances may exhibit more than one hesitation; but most words are spoken in continuous phrases. Few repetitions or false starts per utterance.
2	Limited Fluency Speech has irregular phrasing or sentence rhythm. Poor phrasing, staccato or syllabic timing, and/or multiple hesitation, repetitions, or false starts render spoken performance notably discontinuous.

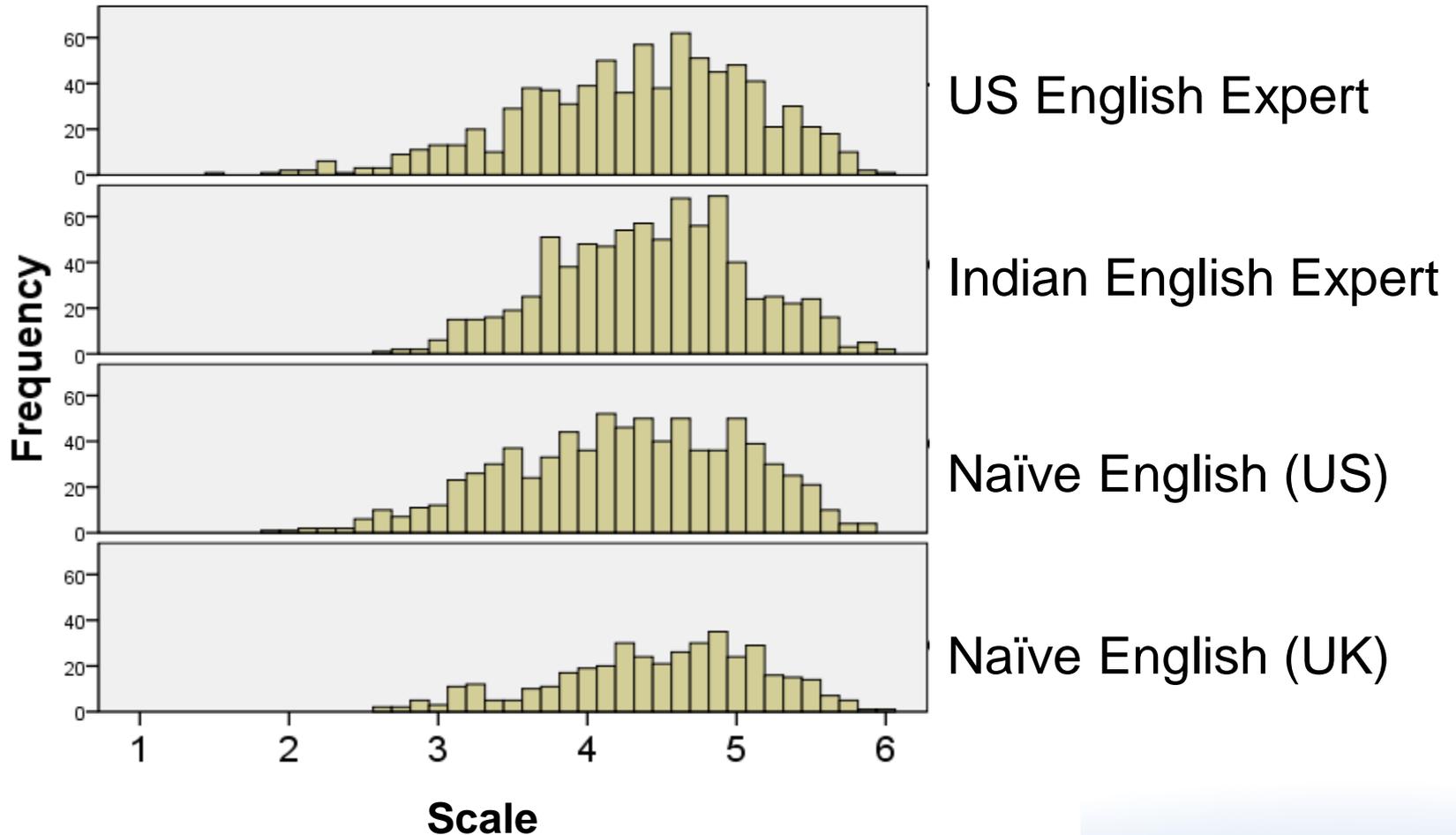
Results - Comprehensibility



Results - Pronunciation



Results - Fluency



Analysis

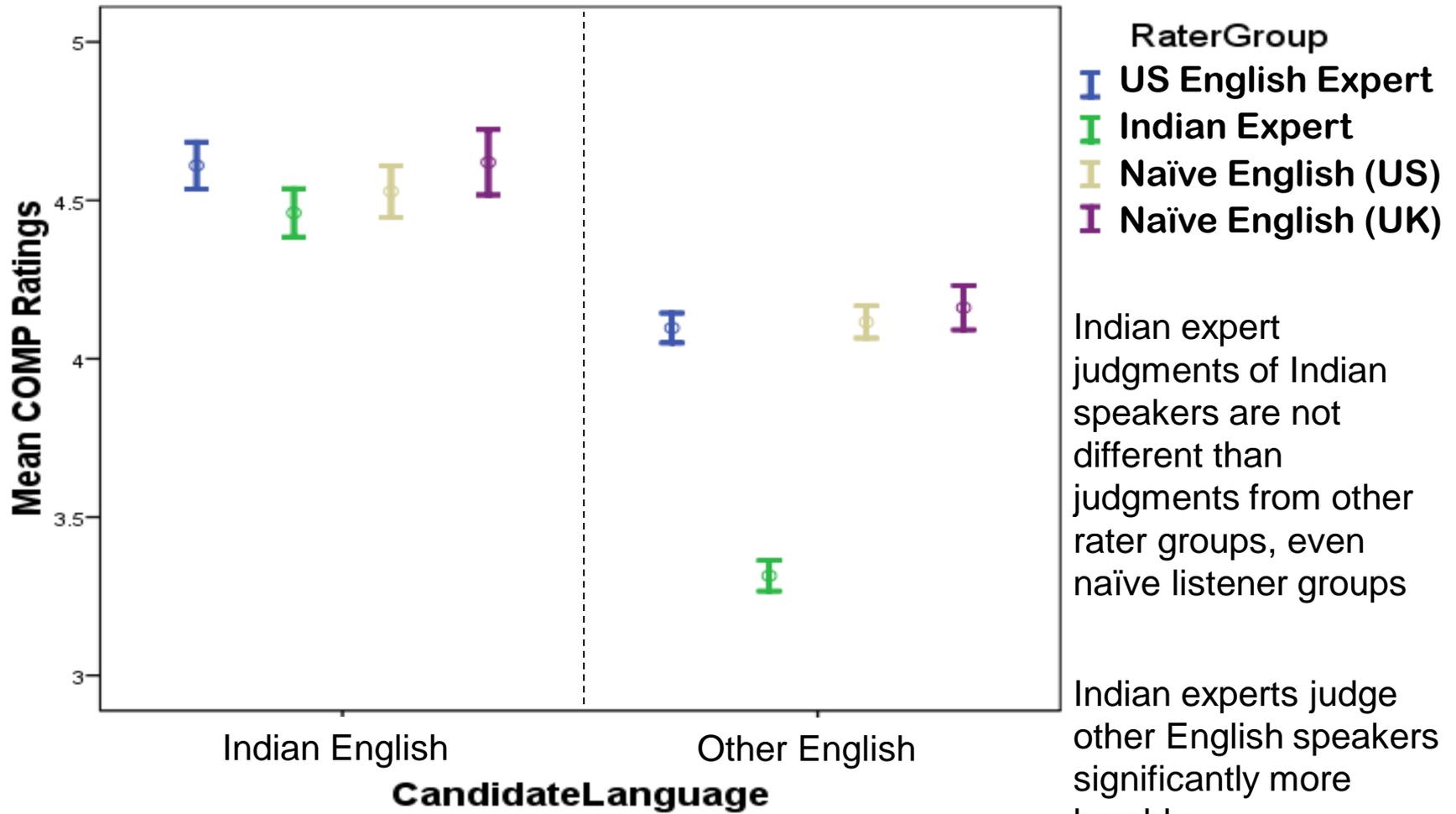
- Separate two-factor ANOVAs for each Trait (Comprehensibility, Pronunciation, Fluency)

<p>Factor 1: Rater Group</p> <p>US English Expert</p> <p>Indian English Expert</p> <p>Naïve English (US)</p> <p>Naïve English (UK)</p>	<p>Factor 2: Candidate Language</p> <p>Indian English</p> <p>Other English</p>
--	--

	Comprehensibility	Pronunciation	Fluency
Rater Group	$p < .05$	$p < .05$	NS
Rater Group * Candidate Language	$p < .05$	$p < .05$	NS

NS = Not Significant at alpha .05

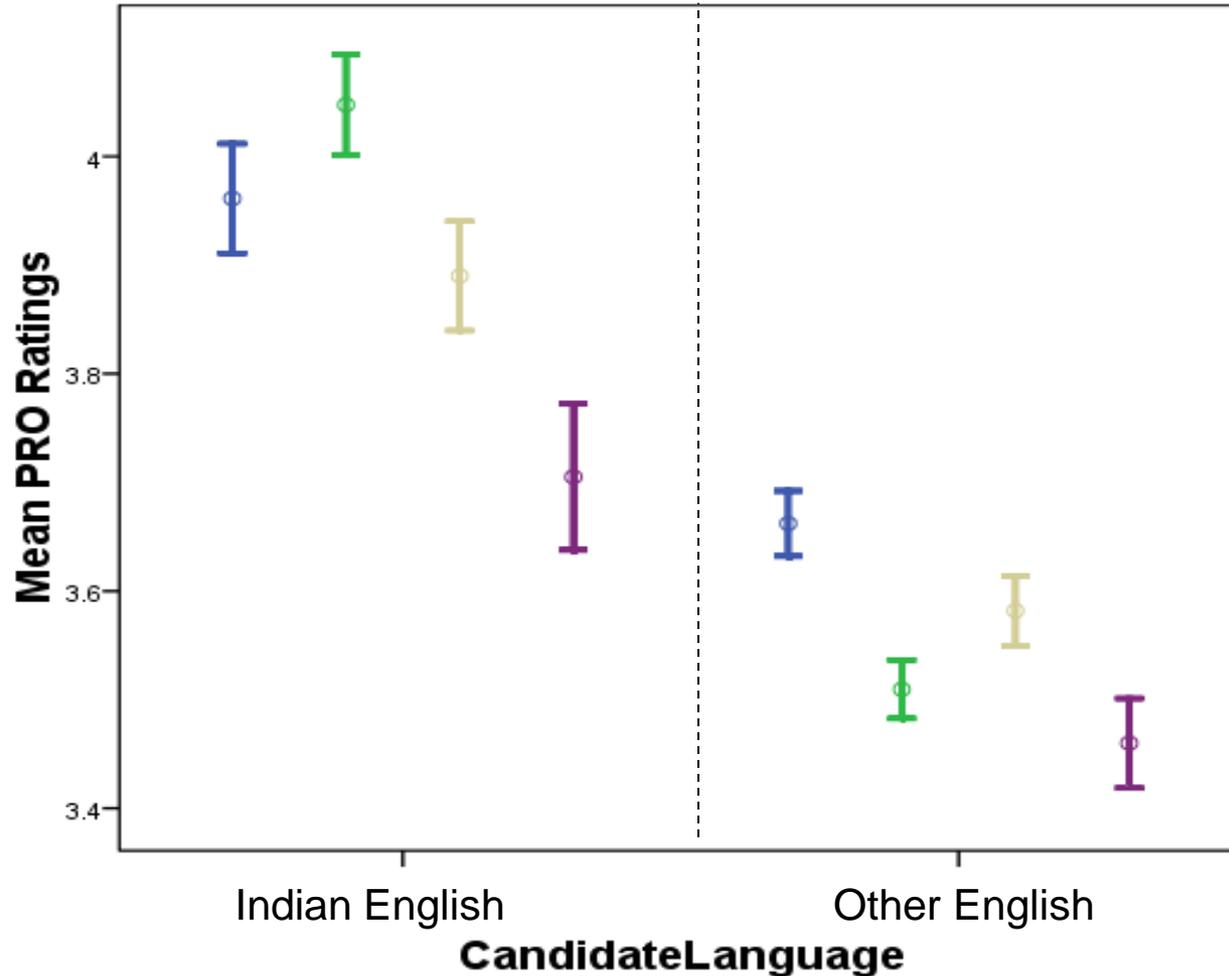
Comprehensibility



Indian expert judgments of Indian speakers are not different than judgments from other rater groups, even naïve listener groups

Indian experts judge other English speakers significantly more harshly

Pronunciation

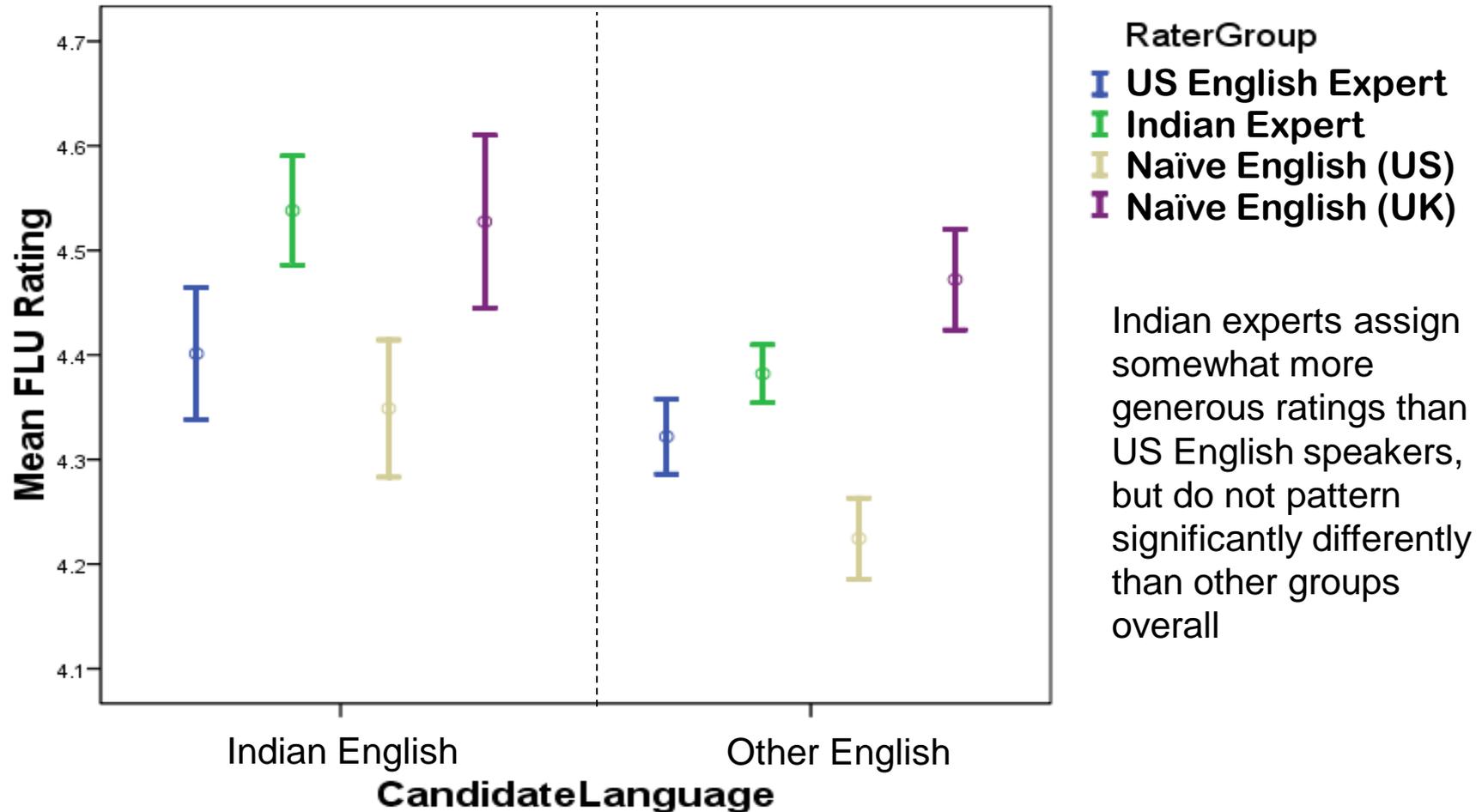


- RaterGroup
- US English Expert
 - Indian Expert
 - Naive English (US)
 - Naive English (UK)

Indian experts are relatively more generous to Indian candidates than naïve listeners with regard to pronunciation judgments

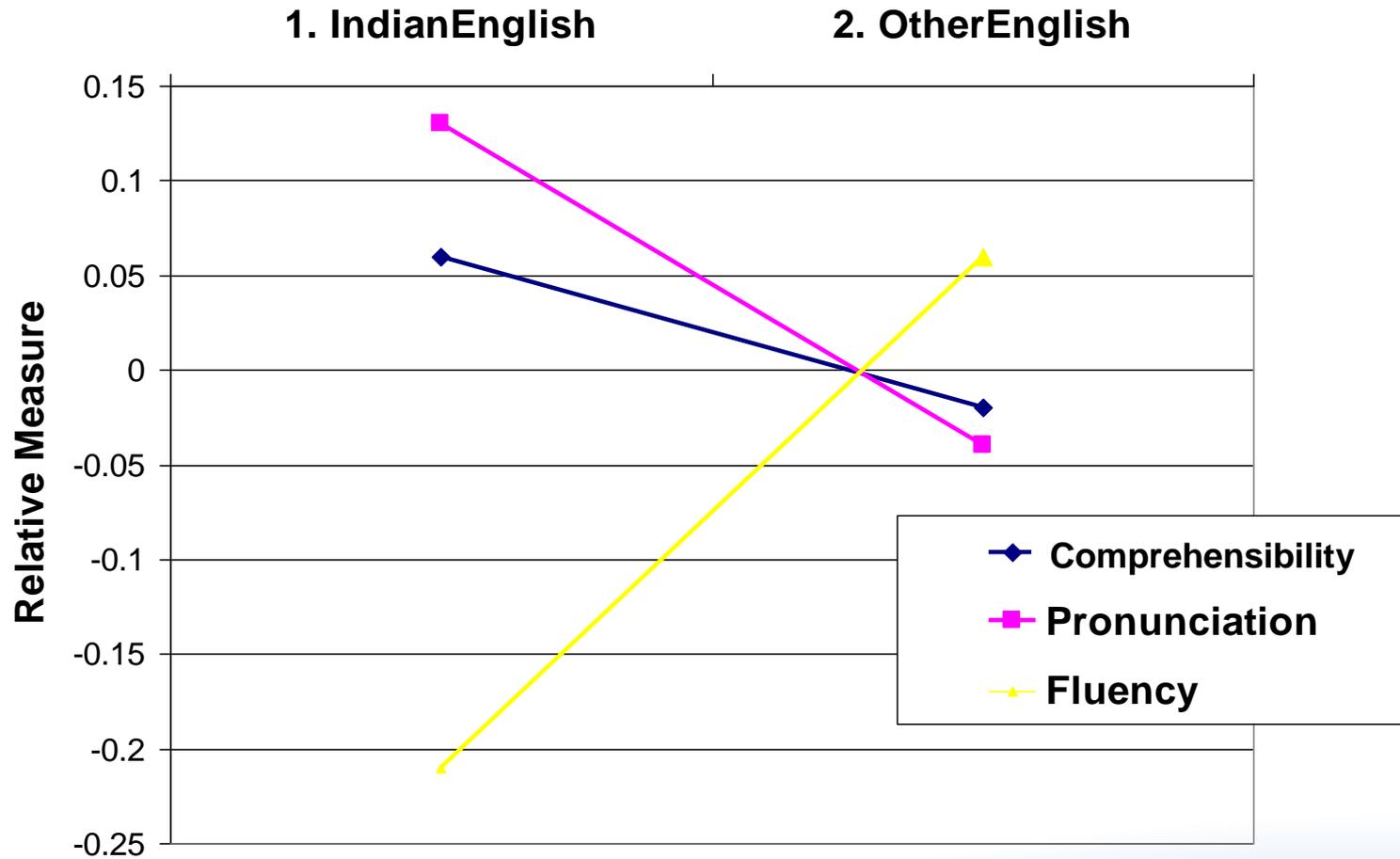
Their ratings of other English speakers are not significantly different than judgments from naïve listeners

Fluency



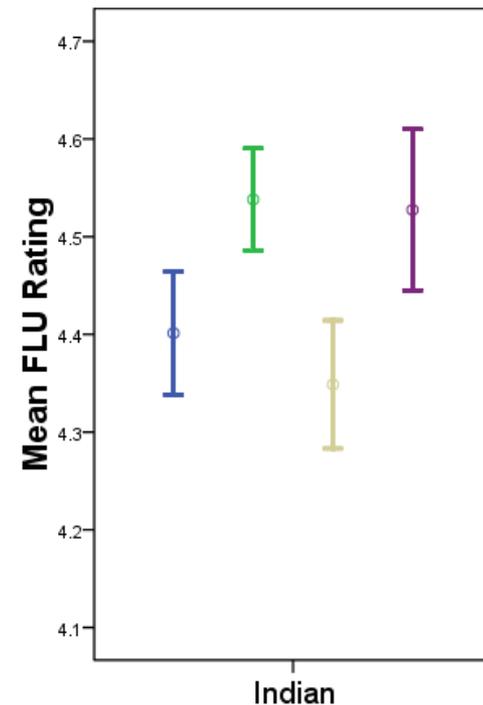
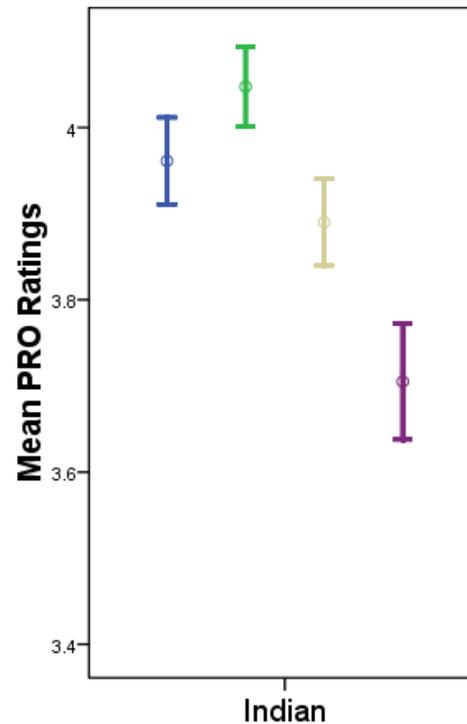
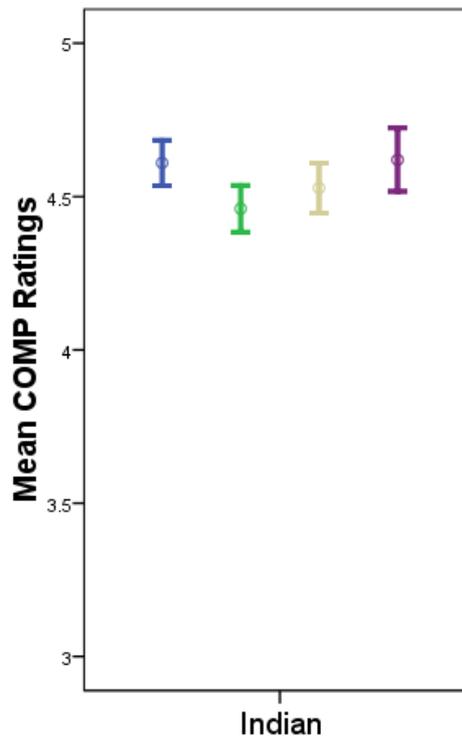
Indian experts assign somewhat more generous ratings than US English speakers, but do not pattern significantly differently than other groups overall

Bias/interaction analysis



Summary

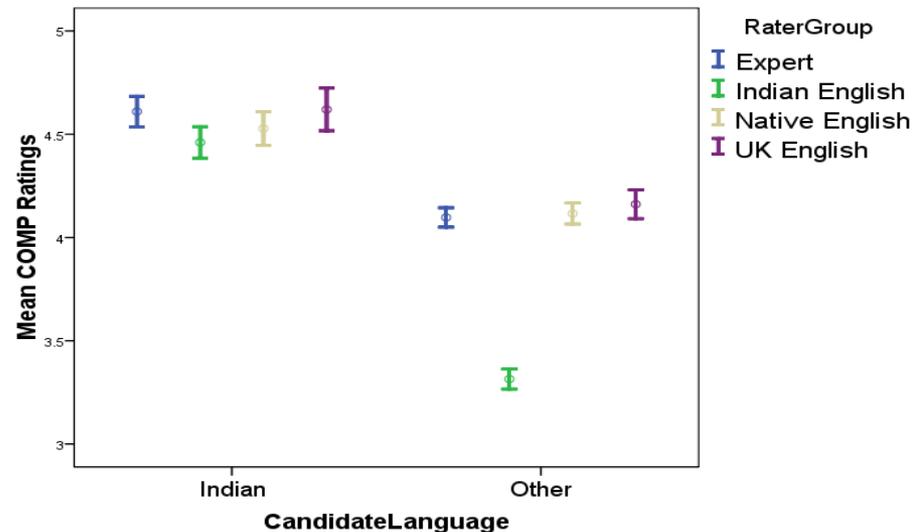
1. Do Indian expert linguists and US expert linguists have different assessments of Indian speech?



No, they do not (unexpected).

Summary

2. Do Indian expert linguists have different assessments of comprehensibility of Indian speech from naïve listeners (customers)?



No, they do not (unexpected). But they do show a bias against other L2 English speakers.

Findings

- Indian listeners as a rater group:
 - Do NOT perform differently to other English NS rater groups when listening to Indians
 - DO perform differently to other English NS rater groups when listening to Other L2 English
- Indian speakers are perceived generally as being disfluent in comparison to their comprehensibility / pronunciation

Implications

- Findings have intuitive appeal – Indian listeners show bias against non-Indian L2 speech
- Despite Indian raters having a preference for Indian speech, overall they behave like other trained raters when listening to Indians (cf. Xi 2008)

Thank you

Ryan.Downey@pearson.com

Alistair.VanMoere@pearson.com