Transitions in language assessment: Topics en vogue, in the doghouse, inert despite lip service & back from the dead

Talia Isaacs, UCL Institute of Education

University College London, UK

EALTA, Dublin, Ireland

May 31, 2019

Transitions in language assessment:

Topics en vogue, in the doghouse, inert despite lip service & back from the dead

Inspiration: EALTA theme; image made me think of…

16th EALTA conference
Transitions in Language Assessment

31st May 2019 - 2nd June 2019
University College Dublin

https://www.ucd.ie/alc/ealta2019/

Seasonal transitions

"You know it's summer in Ireland when the rain gets warmer" (Hal Roach, comedian)

Photo attribution: J+B=Us "Rain in Temple Bar"
https://search.creativecommons.org/photos/fe81e9ed-b8bc-47ae-814e-1a0ec4771c54 3

Transitions in language assessment:
Topics en vogue, in the doghouse, inert despite lip service & back from the dead

If the conference is EALTA, why not call the theme "Transitions in language testing & assessment?"

16th EALTA conference
Transitions in Language Assessment

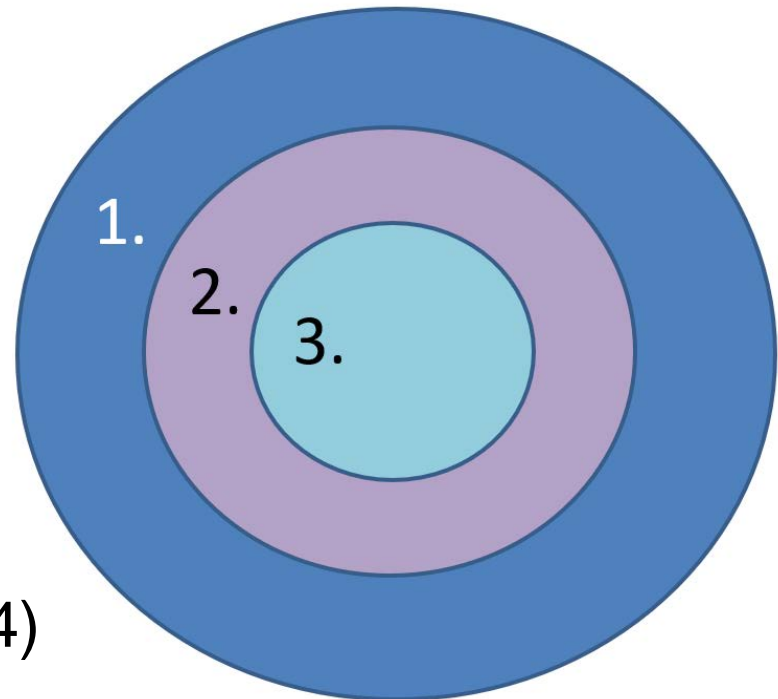31st May 2019 - 2nd June 2019
University College Dublin

https://www.ucd.ie/alc/ealta2019/

Transitions in language assessment:
Topics en vogue, in the doghouse, inert despite lip service & back from the dead
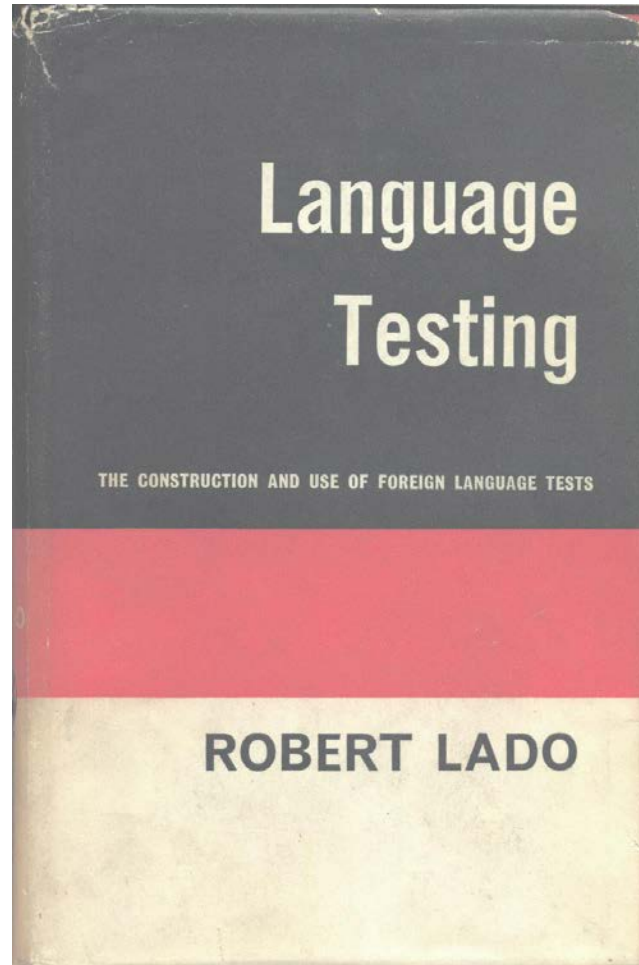
If the conference is EALTA, why not call the theme "Transitions in language testing & assessment?"
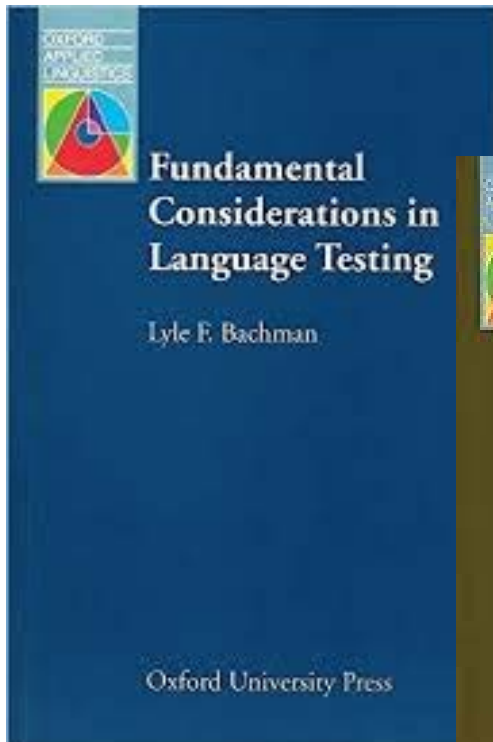
1. Assessment
2. Measurement
3. Test

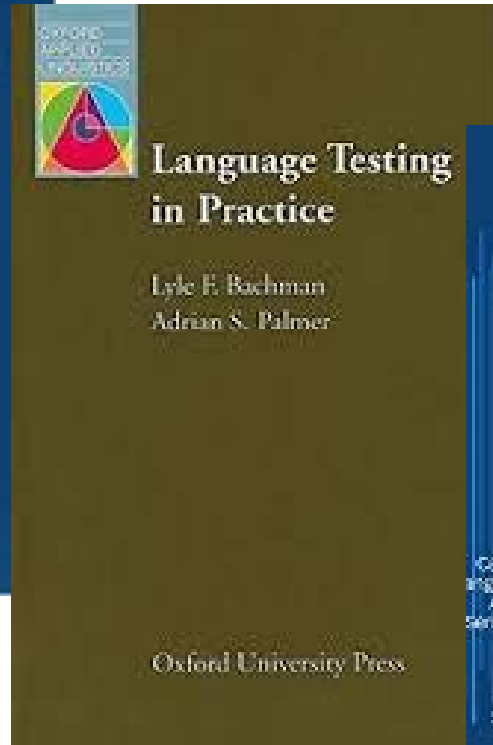Bachman (2004)

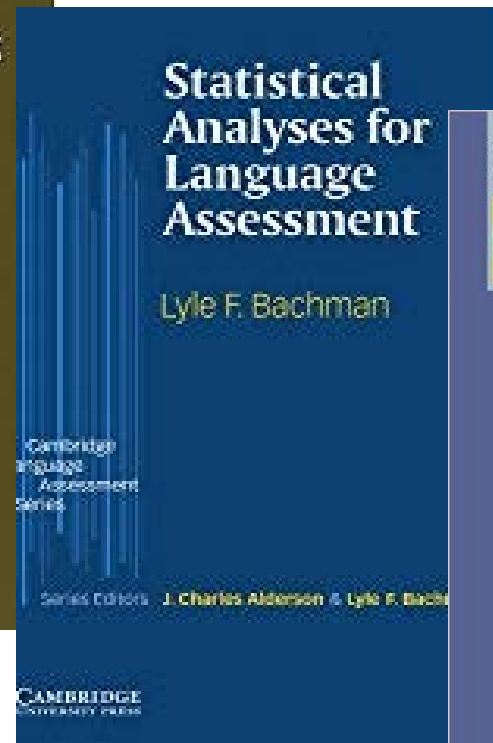# Birth of LT as a field (Spolsky, 1995)

1961

Transition: Evidenced in Bachman's book titles

Fundamental Considerations in Language Testing
Lyle F. Bachman
Oxford University Press

1990

Language Testing in Practice
Lyle F. Bachman
Adrian S. Palmer
Oxford University Press

1996

Statistical Analyses for Language Assessment
Lyle F. Bachman
Cambridge Language Assessment Series
Series Editors: J. Charles Alderson & Lyle F. Bachman
CAMBRIDGE UNIVERSITY PRESS

2004

Language Assessment in Practice
Lyle F Bachman
Adrian S Palmer
OXFORD

2010

Web browser hits (May 18, 2019)

Search terms	Google	Yahoo (powered by Bing)	Baidu
"Language testing"	3.2 million	2.04 million	8.38 million
"Language assessment"	3.58 million	2.24 million	6.86 million
"Language testing and assessment"	168,000	? "and" treated as a Boolean operator	148,000

Major journals in our field

▸ Which was founded first (presented here in alphabetical order)?

- *Assessing Writing*
- *Language Assessment Quarterly*
- *Language Testing*

Years: 1984, 1994, 2004

Just when you start to think that "language testing & assessment" is a collocation…
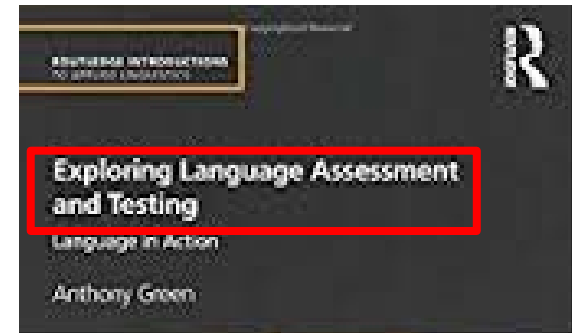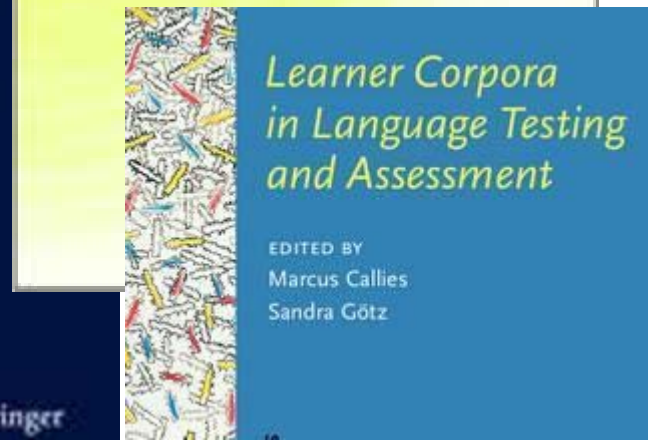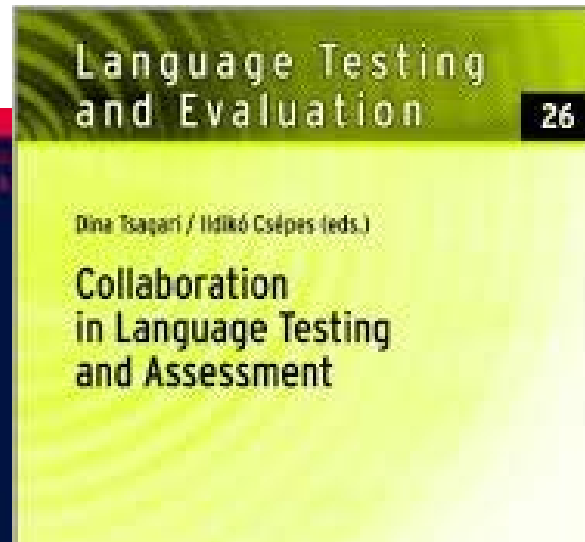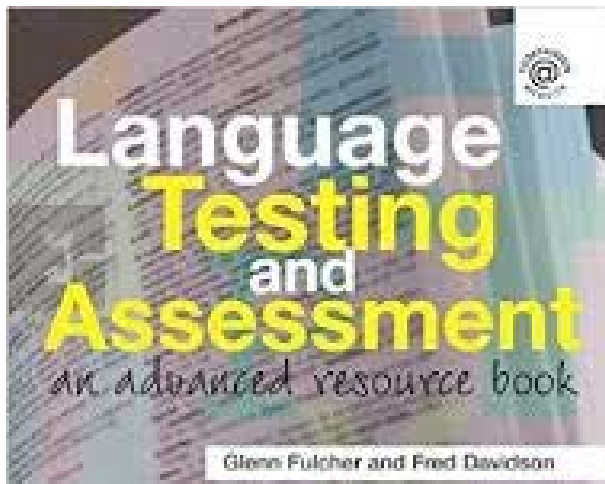
2014

Language testing (LT) as a field

- In the early days, heavily driven by statistical analyses of large-scale or standardised tests
  - Heavily influenced by psychometrics; highly technical

- 1979, Boston, US: *Colloquium on the validation of oral proficiency tests*
  - Became what is now the Language Testing Research Colloquium (LTRC)

Selected presentations, 1979 colloquium

- Clark, J. *Convergent-divergent test validation* at *Educational Testing Service*
- Groot, P. *On the need for construct validation of oral proficiency tests*
- Hinofotis, F., Bailey, K, & Stern, S. *The development of a performance test of oral proficiency for foreign teaching assistants*.
- Lowe, P. *Structure of the oral interview and content validity*.
- Engelskirchen, Cottrell, & Oller. *The reliability and validity of the Ilyin Oral Interview.*
- Palmer, A. *Measurements of reliability and concurrent validity of two picture description tests of oral communication.*
- Shohamy, E. *The concurrent validity of the oral interview with the cloze procedure in Hebrew as a second language.*
- Stevenson, D. *Beyond faith and face validity: The multitrait-multimethod matrix and the convergent and discriminant validity of oral proficiency tests.*

Transitions in our field: From stats driven to…

- more pluralistic → content, paradigms, contexts

LTRC2015
TORONTO, MARCH 16-20

From language testing to language assessment:
Connecting teaching, learning and assessment

LTRC 2016–Palermo, Sicily

Language Constructs, Contexts, and Content in

Classroom and Large-Scale Assessments

LTRC2017
39TH LANGUAGE TESTING RESEARCH COLLOQUIUM

Universidad de
los Andes
Colombia

LANGUAGE ASSESSMENT
LITERACY ACROSS
STAKEHOLDER BOUNDARIES

LTPC
2-6 July 2018
Auckland
New Zealand

**Language Assessment in Times of
Movement, Transition, and Change**

LTRC 2019
41ST LANGUAGE TESTING RESEARCH COLLOQUIUM
4-8 March 2019 ❖ Atlanta, Georgia USA

LANGUAGE TESTING AND SOCIAL JUSTICE

2020
9-13 June 2020 Hammamet, Tunisia

**Assessment in multilingual contexts:
Models, practices, policies & challenges**

Transitions in language assessment:
Topics en vogue, in the doghouse, ~~inert~~ *static* ~~despite lip service~~ & back from the dead

Topics en vogue

State-of-the-Art Review *Language Teaching, 2001 & 2002*

Language testing and assessment (Part 2)
J Charles Alderson and Jayanti Banerjee Lancaster University, UK

LTRC2017
39TH LANGUAGE TESTING RESEARCH COLLOQUIUM
JULY 17–21, 2017 · BOGOTÁ, COLOMBIA
LANGUAGE ASSESSMENT
LITERACY ACROSS
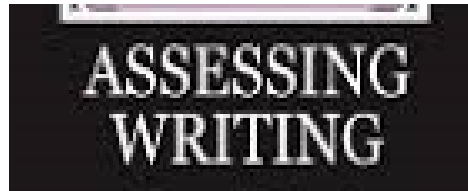STAKEHOLDER BOUNDARIES

New Special Interest Group, International Language Testing Association (ILTA):
Language Assessment Literacy

Alderson & Banerjee:
"Literacy" → computer literacy
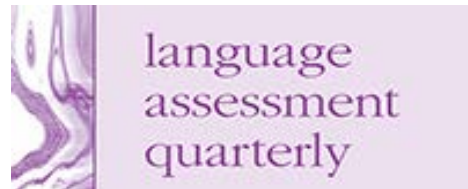→ assessment of literacy

Topics en vogue

Special Issues:

Automated assessment of writing (2013)

Automated scoring & feedback systems (2010)

Conceptualizing & operationalizing speaking assessment for a new century (2018)

Alderson & Banerjee (2001, 2002):

Automated scoring: **E-rater**® (ETS), **Phonepass** (Ordinate)

Transition → Community views on automated assessment (Isaacs, 2018)

Initially – Fierce debate

- E.g., Test review & rebuttal, Phonepass/Versant
  (Chun, 2006, 2008; Downey et al., 2008)

- Now more pragmatic understanding that automated assessment here to stay

- Given technological constraints, to be reliable, automated speaking tasks more reminiscent of grammar translation & audiolingual activities than communicative

Topics back from the dead

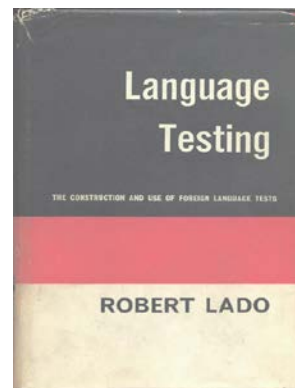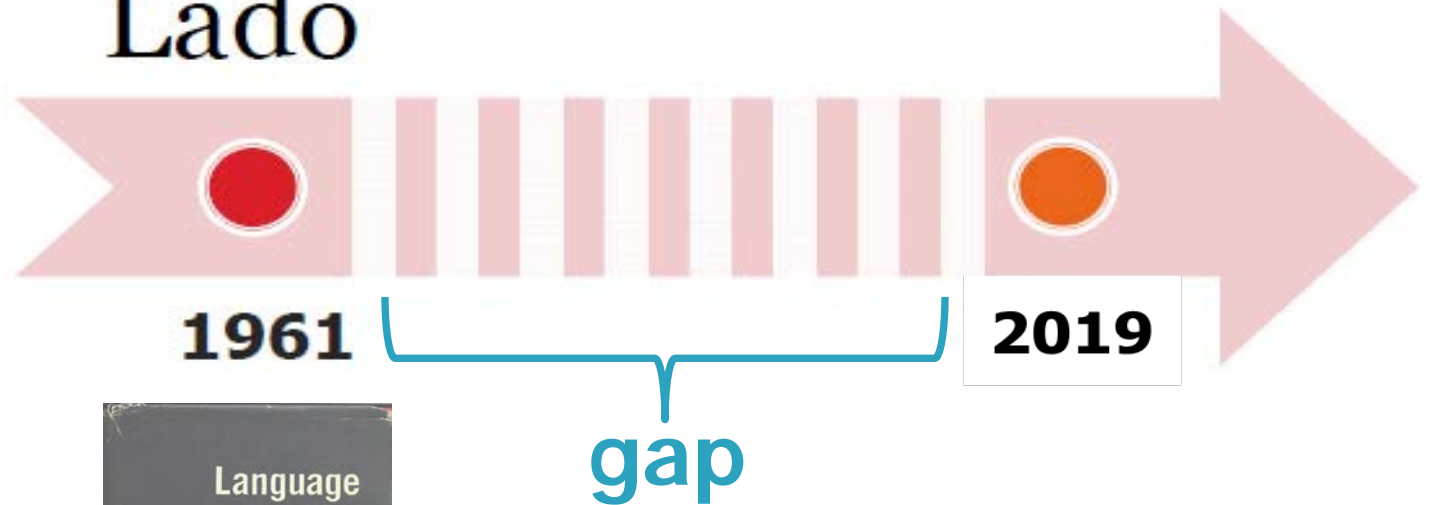## L2 pronunciation assessment

Stuck in a time warp

Lado

1961 — gap — 2019

Language Testing
THE CONSTRUCTION AND USE OF FOREIGN LANGUAGE TESTS
ROBERT LADO

Publications on **transitions** (Isaacs, 2018 a & b)

▸ Fully automated speaking assessment: **Changes** to proficiency testing and the role of pronunciation. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 570–584). Routledge. [Download]

▸ **Shifting sands** in second language pronunciation teaching and assessment research and practice. *Language Assessment Quarterly, 15,* 273–293. [Download]

Publications downlodable:
http://www.taliaisaacs.com

19

Transitions in language assessment:

Topics en vogue, <u>in the doghouse</u>, inert despite lip service & back from the dead

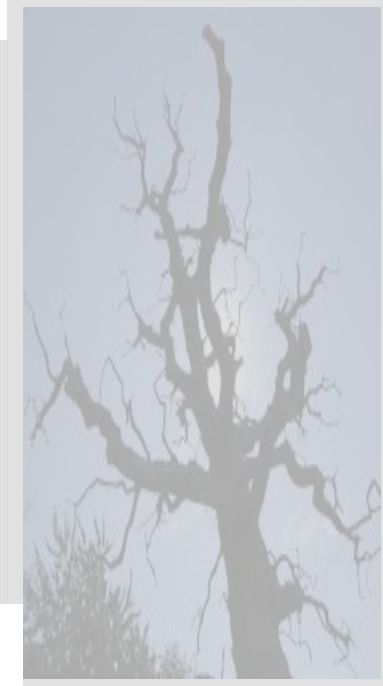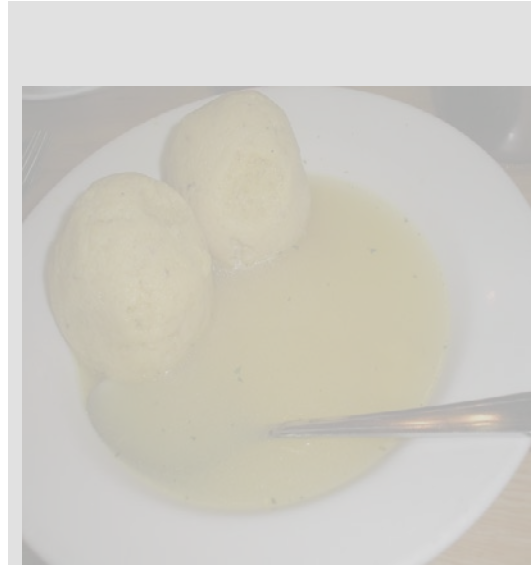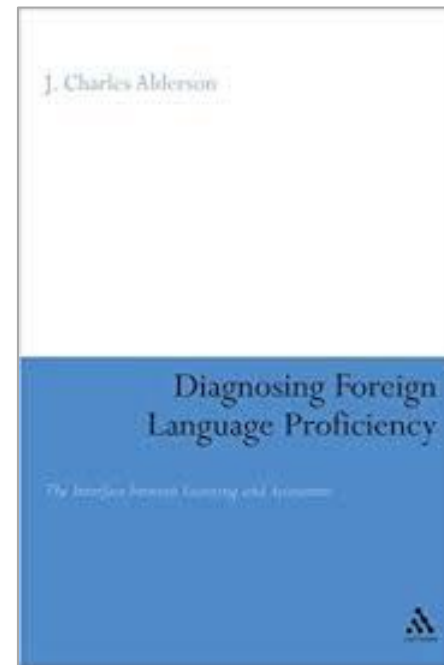

Photo attributions (CC BYlicense): Don LaVantage; George Estreich; Nikchick; Helen L Cave

My first LTRC → 2005 in Ottawa, Canada

▶ Charles Alderson – Plenary

- The challenge of (diagnostic) testing: Do we know what we are measuring?

J. Charles Alderson

Diagnosing Foreign
Language Proficiency

The Interface between Learning and Assessment

2005   21

Original Paper

The Inclusion of Ethnic Minority Patients and the Role of Language in Telehealth Trials for Type 2 Diabetes: A Systematic Review

Talia Isaacs[1*], PhD (iD) ; Daniel Hunt[2*], PhD (iD) ; Danielle Ward[3], MPH (iD) ; Leila Rooshenas[4], PhD (iD) ; Louisa Edwards[4], PhD (iD)

Underrepresentation of black and ethnic minority patients in clinical trials must be addressed

BAME patients may be underrepresented in trials for diabetes treatment, study finds

Language could affect minority community participation in diabetes research trials

Open access article: https://www.jmir.org/2016/9/e256/

LTRC conference abstract: *How is language proficiency operationalized in randomized controlled trials (RCTs) in healthcare settings?*

Reviewer 2: This is a highly novel and interesting topic which demonstrates a clear cross-over between [...] of medical health researchers and langu[...] paper would be a valuable addi[...] nce program.

Reviewer 1: I don't th[...] interesting to most conference particip[...] s is the underrepresentation of minor[...] ical research. Not clearly a study for a [...] guage testing audience. It would be more appropriate for a conference on medical research design.

REJECT

E-mail: Hello & bias against interdisciplinary - LTRC "Language testers really need to know about these issues & their input is crucial to addressing this area, which is high-stakes for many patients. The topic also fits closely with the conference theme: "any aspect of the interface among language constructs, contexts & content (or disciplinary subject matter, themes, topics, etc.) in language assessment…""

Reviewer 1: I don't think this would be interesting to most conference participants. The focus is the underrepresentation of minorities in medical research. Not clearly a study for a language testing audience. It would be more appropriate for a conference on medical research design.

24

LTRC conference chair:

- <u>Theme</u> (<u>context + constructs</u>) not <u>explicitly expressed in title</u> or <u>abstract</u>

- Do not think this was interdisciplinary "bias"

- **My reply:** Relevance to conference theme strongly implied

<u>TITLE</u>: ***How is <u>the</u> language proficiency <u>construct</u> operationalized in randomized controlled trials (RCTs) in healthcare settings?***

<u>In the doghouse</u>

My reply: "Reviewer 1 was not commenting on the conference theme. He/she seemed to feel that the topic was not interesting to most conference goers & "not clearly a study for a language testing audience"---- hence my view on interdisciplinary bias. I personally think this reflects a narrow mentality of what fits and doesn't fit into the language testing box, although Reviewer 2 appears to have contradicted this with a more positive response."

In the doghouse

LTRC conference abstract: *How is language proficiency operationalized in randomized controlled trials (RCTs) in healthcare settings?*

▸ Selection .... → organisers of invited AAAL-LTRC joint s........... had encouraged submission:
  - did not .... age for specific purposes]
    topic

REJECT

[In the doghouse](https://www.flickr.com/photos/nycgeo/6450561201)

- **Writing an abstract**

How is language proficiency operationalized in randomized controlled trials (RCTs) in healthcare settings?

Randomize... ...health intervention research. However, patient recruitme... ...their recruitment targets, leading ... ...ent of certain sociode... ...Particularly pressing ... ...conditions that dispropo... ...atments due to their greater ... ...cacy of those clinical interventions (Bartlett et al., ...)

One challenge to minority participation in RCTs relates to patients' language pro... ...demonstrated greater potential for miscommunication when there is a mismatch between the patient's and health professional's primary language than when they share the same mother tongue (Segalowitz & Kehayia, 2011). Ensuring adequate language ability is

We will work through examples of more successful abstracts

The Inclusion of Ethnic Minority Patients and the Role of Language in Telehealth Trials for Type 2 Diabetes: A Systematic Review

Talia Isaacs[1*], PhD (iD) ; Daniel Hunt[2*], PhD (iD) ; Danielle Ward[3], MPH (iD) ; Leila Rooshenas[4], PhD (iD) ; Louisa Edwards[4], PhD (iD)

Underrepresentation of black and ethnic minority patients in clinical trials must be addressed

BAME patients may be underrepresented in trials for diabetes treatment, study finds

Language could affect minority community participation in diabetes research trials

Open access article: https://www.jmir.org/2016/9/e256/

Why an interest in randomized controlled trials (RCTs)?

▸ Considered **gold standard** of health intervention research (Meyer et al., 2019)

- Test **effectiveness & safety** of new medical treatments

▸ **BUT** difficulty recruiting enough patients → **underpowered** studies (Donovan et al., 2014)

- cost & resource implications

▸ Recruited sample should be **representative** of the target population (Bartlett et al., 2005)

- **External validity**
  - **Equitable** access to health services; **inclusion**

Focus of systematic review (1/2)

Type 2 diabetes

▸ Ethnic minorities vulnerable

▸ In the UK compared to Caucasians, incidence
- up to 6x higher → South Asian descent
- up to 3x higher → African/Afro-Caribbean descent

▸ poorer outcomes (Negandhi et al., 2013)

BUT underrepresented in RCTs
- stand to benefit the most from participating

Teleheath RCTs

▸ Telehealth = remote healthcare delivery, telecommunications technology (Ackerman et al., 2010)

- universally accessible?
  - heightened language demands → some telehealth interventions
    - E.g., phone more difficult than face-to-face

RCTs & the role of language

Need to find out

- whether & **how** trial recruiters **determine** that patients have the requisite language **proficiency** to
  - provide informed consent
    - ethical imperative, applies to ALL RCTs
  - engage with the intervention
- **Link** between **language screening** & **ethnic composition** of the **recruited sample**
  - could studies that refer to language requirements **deter hard-to-reach patients** from participating?

Research objectives: Examine
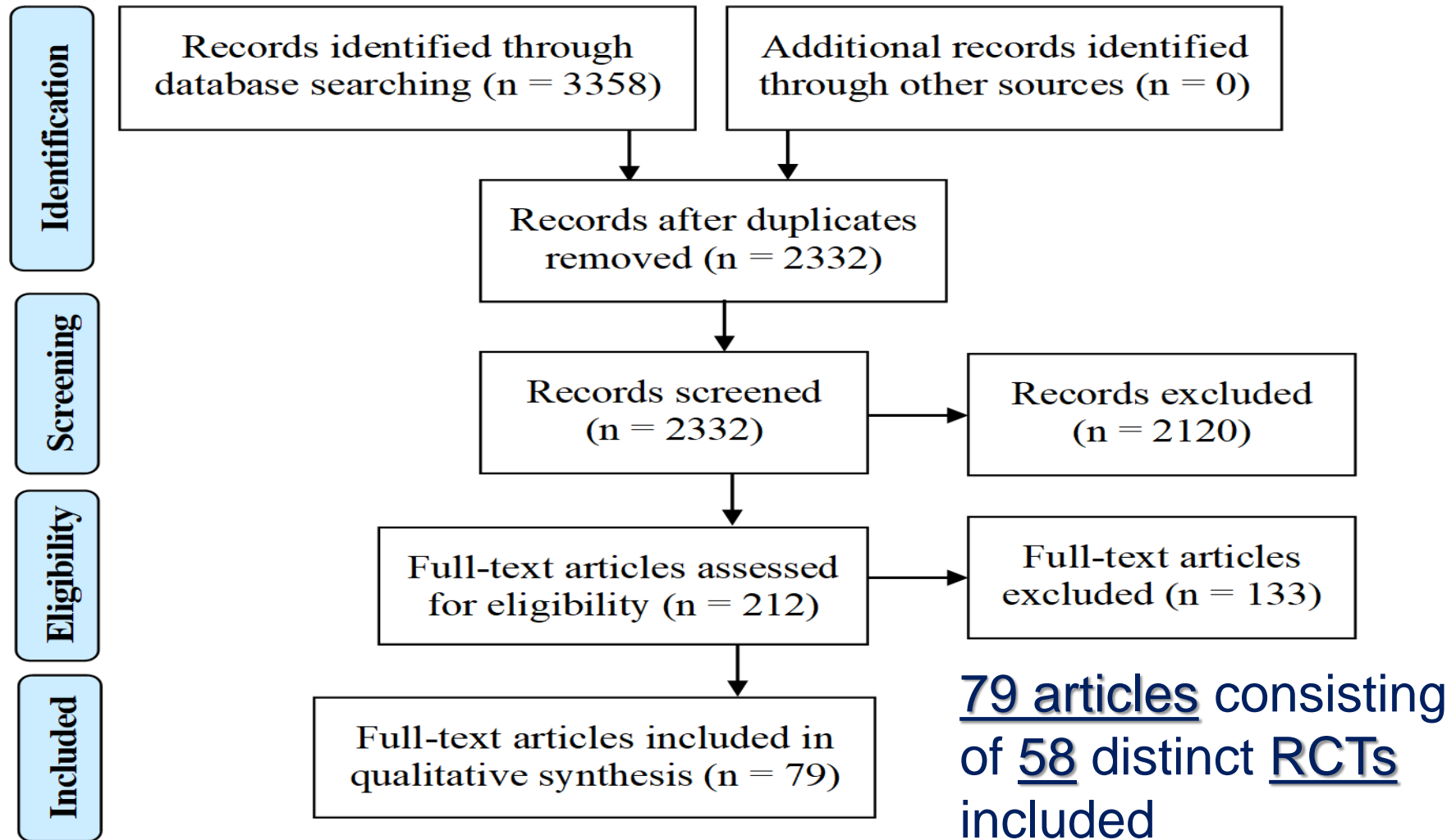
▸ Research reporting & prevalence of ethnic minorities recruited to telehealth diabetes RCTs
  - trial characteristics associated with successful recruitment

▸ Role of language & its prevalence as a screening criterion
  - operationalization of language proficiency in making patient inclusion/exclusion decisions

1st systematic review on telehealth diabetes RCTs to focus on language

Methods: Study selection criteria

▸ peer-reviewed articles in English

▸ recruiting adult patients with type 2 diabetes
- Western countries; English official & majority language
  - Australia, Canada, Ireland, New Zealand, UK, US

▸ not ethnically-targeted

▸ intervention: any telehealth medium

▸ Excluded RCTs also targeting gestational or type 1 diabetes

Methods: Prisma diagram & search strategy

Records identified through database searching (n = 3358)

Additional records identified through other sources (n = 0)

Records after duplicates removed (n = 2332)

Records screened (n = 2332)

Records excluded (n = 2120)

Full-text articles assessed for eligibility (n = 212)

Full-text articles excluded (n = 133)

Full-text articles included in qualitative synthesis (n = 79)

79 articles consisting of 58 distinct RCTs included

Identification · Screening · Eligibility · Included

Selection of eligible studies. Searched MEDLINE, PsycINFO, EMBASE, CINAHN, CENTRAL & references of related reviews

Download article: https://www.jmir.org/2016/9/e256/   36

Methods: Data extraction

2 independent reviewers recorded study methodological details, participant demographics & intervention characteristics

English language/literacy part of inclusion/exclusion criteria?	Yes/No/Multilingual (if English is one of two or more languages permitted)
Reason(s) English language required	Informed consent/ethics; Fitness of/ability to participate in the intervention; Lack of resources; Other (specify); None given
Language proficiency definition/operationalization	Note: Provide the description then categorize according to the language skill emphasized (speaking, listening, reading, writing), whether phone communication is specified, etc.
Compensation provided for participation?	Yes (elaborate)/No
Translation/interpretation included/ bilingual interventionists	Yes (elaborate)/No

Methods: Risk of bias assessment

2 independent reviewers assessed methodological quality of included studies
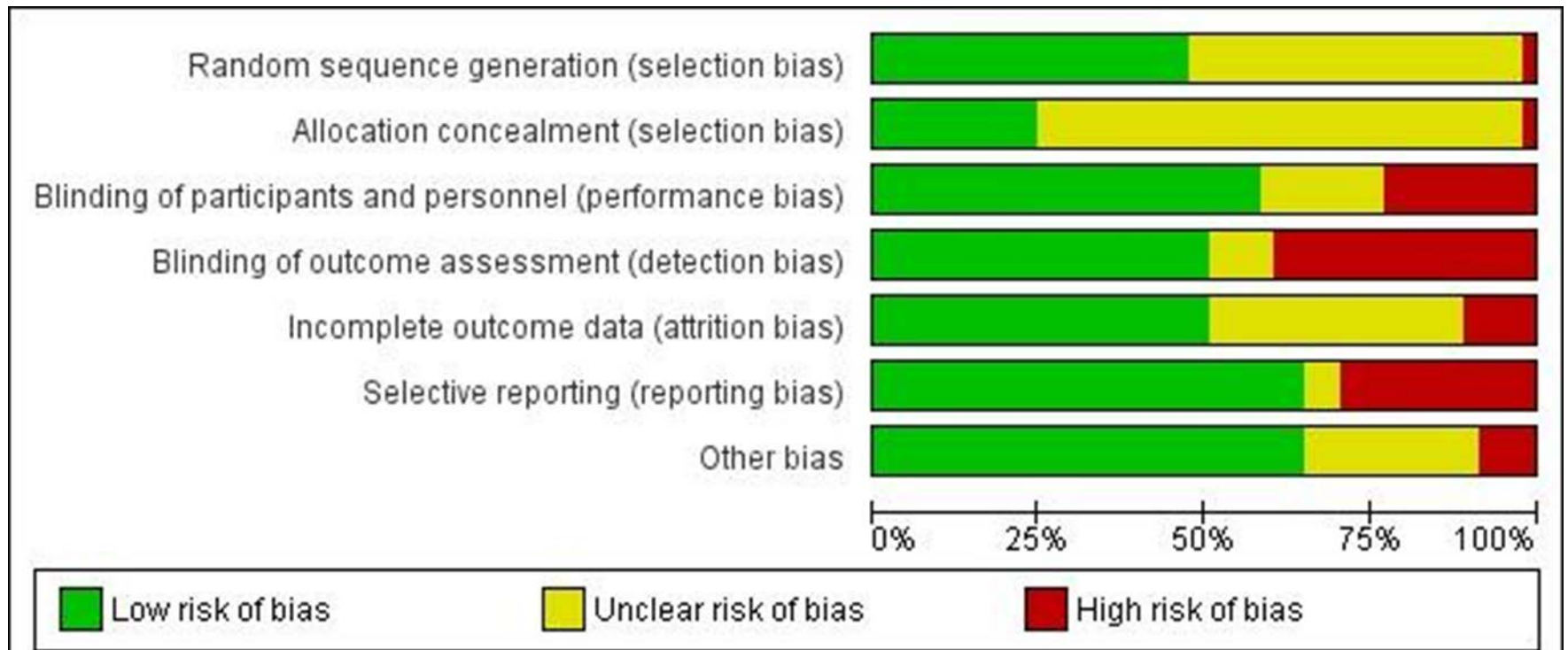
Results – Ethnicity

- ▶ < 2/3 of included RCTs (38/58) reported on ethnic composition of recruited sample
  - Median proportion of ethnic participation: **23.5%**, range: 0%–97.7%

- ▶ < 1/4 (14/58) recruited a sizeable proportion of ethnic minorities (**≥30%** of total recruited sample)
  - All US-based
  - Urban areas; mostly deprived, underserved, or uninsured patients

High (14/28) vs. low (14/28) ethnic minority recruiting RCTs

▸ **High** (≥30 minority recruitment)

- 8/14 offered intervention in both English & Spanish
- E.g., bilingual staff, phone interventions
  - 2 used interpretation/translation during recruitment; 4 cultural tailoring for ethnic minorities, 6 tailoring for low literacy

≠

▸ **Low** (≤30% minority recruitment)

- no bilingual provision offered

Results – Ethnicity
Headlines tell part of the story

BAME patients may be underrepresented in trials for diabetes treatment, study finds

Underrepresentation of black and ethnic minority patients in clinical trials must be addressed

▸ Ethnic minorities underrepresented
  • Unclear whether the intervention being tested is safe & effective

▸ Telehealth delivery mode mixed across high minority recruiting studies

41

Results – Language proficiency

▸ **Half** of included RCTs (29/58) **reported patient language proficiency** as an **eligibility criterion**
  - Underreported or language not considered?
    - 4000 word journal articles

▸ **No common procedures/instruments** across RCTs **to determine patients' eligibility** on language grounds
  - No mention of CEFR levels

How language operationalized – 29 RCTs reporting it as screening criterion

- 9 - speak/communicate in English
  - E.g., "must speak English" (Quinn et al., 2011)
- 2 - speak & understand
  - E.g., "Ability to clearly speak/understand English via telephone" (Williams et al., 2012)
- 4 - speak & read
- 2 - read & understand
- 7 - read & write
- 5 - English (or Spanish) primary/main language; native speakers
  - E.g., "excluded patients who... were non-English speaking" (Ralston et al., 2009)

43

Examples of mismatch between telehealth medium & language skills reportedly screened

RCT	Language screening criterion	Telehealth medium
Glasgow et al. (2012)	"ability to **read & write** in English or Spanish"	Interactive website, **telephone** calls (automated & human)
Carter (2011)	Exclusion: "**Illiteracy** or inability to read at an **eighth grade reading** level" "Brief test" to assess literacy (no details given)	Laptop with home monitoring, **videoconference** & online personal health record

Headlines tell part of the story

Language could affect minority community participation in diabetes research trials

- Language mentioned as eligibility criteria in only half of included RCTs
- Operationalization of language proficiency haphazard
  - No standardized instruments or common/procedures for measuring it
 - Little guidance for recruiters on how to assess it

45

Implications

▸ **I**nclusion/exclusion decisions depend on trial recruiters' subjective judgments

- strong potential for bias
  - e.g., perceptible L2 accent that does not impede communication
- including patients who do not understand the nature of the intervention due to language barriers also problematic (unethical)

Future research

▸ Design a simple, practical means to more fairly & consistently screen patients' language proficiency in RCT recruitment

- minimize the possibility of patients being unfairly excluded based on arbitrary human decision-making on language grounds

- One-size-fits all instrument for use across RCTs?
- Possible to identify threshold level?
- Role of health literacy?

Take-home message

▸ Examples from this talk demonstrate that whether or not something is

IN

OUT

Language testing/ assessment box
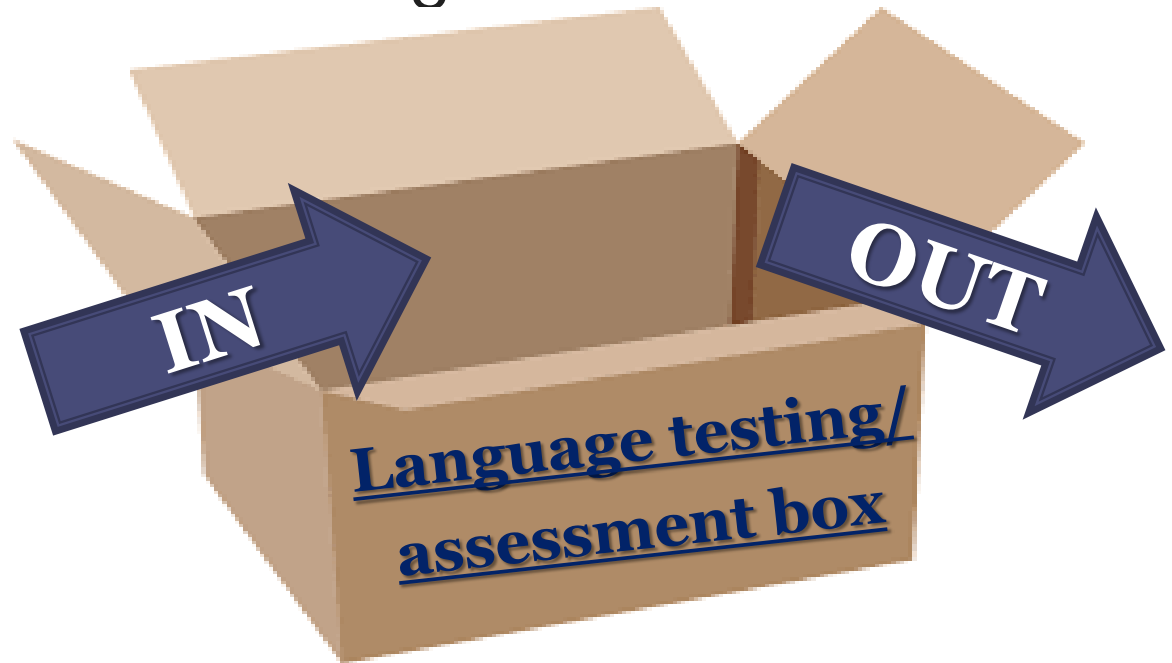
- subject to change over time
- in the eye of the beholder

Thank you!

Go raibh maith agaibh

talia.isaacs@ucl.ac.uk

http://www.taliaisaacs.com