EXAMINING THE DIMENSIONALITY OF L2 READING COMPREHENSION OF TAIWANESE EFL BEGINNERS

Pei-Yu (Marian) Pan
Jeng-Shin Wu
Hsin-Hao Chen
Yao-Ting Sung

LITERATURE REVIEW

- What is reading comprehension?
  - What skills can we measure?
- Importance of identifying the dimensions of reading comprehension:
  - Provide empirical support for test validity
  - Influence the development of theories and models, assessment tools, instruction, and curriculum
- Various classifications have been proposed.

- Gray (1960) proposes three levels of understanding:
  - reading the lines = literal meaning
  - reading between the lines = inferred meaning
  - reading beyond the lines = critical evaluation

- Lennon (1962):
  - word knowledge
  - comprehension of explicitly stated meaning
  - comprehension of implicit/inferential meaning
  - appreciation

- Davis (1968):
  - Recalling word meanings
  - Drawing inferences about the meaning of a word in context
  - Finding answers to questions answered explicitly or in paraphrase
  - Weaving together ideas in the content
  - Drawing inferences from the content
  - Recognizing a writer's purpose, attitude, tone and mood
  - Identifying a writer's technique
  - Following the structure of a passage

- Munby's (1978) taxonomy of microskills:
  - Recognizing the script of a language
  - Deducing the meaning and use of unfamiliar lexical items
  - Understanding explicitly stated information
  - Understanding information when not explicitly stated
  - Understanding conceptual meaning
  - Understanding the communicative value of sentences
  - ……

- Weir (1994) proposed three operations in reading:
  - Skimming
  - Understanding main ideas and important detail
  - Using linguistic contributory skills
    - understanding grammatical notions, syntactic structure, discourse markers, lexical and or grammatical cohesion, and lexis

- Abdullah's (1994) critical reading skills:
  - evaluate deductive inferences
  - evaluate inductive inferences
  - evaluate the soundness of generalization
  - recognize hidden assumptions
  - identify bias in statements
  - recognize author's motives
  - evaluate strength of arguments

- Alderson (2005) - DIALANG:
  - To understand/identify the main idea(s), main information in or main purpose of text(s)
  - To find specific details or specific information
  - To make inferences on the basis of the text by going beyond the literal meaning of the text or by inferring the approximate meaning of unfamiliar words

- Those lists are theoretically persuasive, but lack sufficient evidence.
  - powerful frameworks for test construction

- Can reading comprehension be divided into discrete skills?
  - Unitary: highly overlapped skills → can be represented by one underlying factor
  - Multi-divisible

UNITARY VIEW AND EVIDENCE

- Rost (1993):
  - L1 (Germany) reading comprehension ability of 220 second graders
  - factor analysis: a general competence was found accounting for 85% of the variance for L1 reading comprehension
- van Steensel, Oostdam, and van Gelderen (2013):
  - SALT-reading
  - 200 low-achieving seventh graders (L1)
  - CFA: one underlying skill
- Alderson (2005):
  - the reading test of DIALANG
  - 718 participants from different European nationalities
  - Various factor analyses: one factor emerged and accounted for between 68% and 74% of the variance in reading

MULTI-DIVISIBLE VIEW AND EVIDENCE

- Jang & Roussos (2007)
  - the reading subtest of TOEFL (1997) – July and August testlets
  - about 3000 ESL students
  - DIMTEST:
    - July testlet: vocabulary, anaphora, main idea, synthesis, negation, and extrapolation
    - August testlet: vocabulary, explicit info, inferencing, and synthesis
- Song (2008)
  - the Web-based English as a Second Language Placement Exam (WB-ESLPE) for ESL college students
  - SEM→ 2 subskills
    1. understand the main ideas, supporting information, and specific details (literal)
    2. make inferences (inferential)
- Kong&Li (2009)
  - the reading subtest of TEM4 (Test for English Majors – Level 4)
  - 20,000 college students (English majors)
  - EFA, CFA, and SEM → 2 factors
    1. literal comprehension
    2. all the others (complex)

CONFIRMATORY FACTOR ANALYSIS

	x factors
RMSEA (Root Mean Square Error Of Approximation)	< 0.05
CFI	> 0.90 or 0.95
TLI	> 0.90 or 0.95
WRMR (Weighted Root Mean Square Residual)	> 1

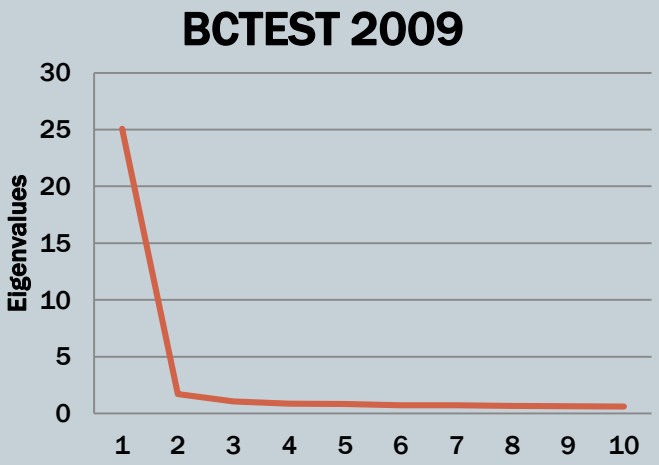
Chi-square test for difference testing	1 vs 2
Value	0.026
Degree of freedom	1
P-value	0.8722

EXPLORATORY FACTOR ANALYSIS

- One of the most common methods to investigate dimensionality
  - No presumptions; exploratory and linear factor analysis
- Compare eigenvalues (>1); the % of the accounted variance

	1	2	3	4	5
Eigenvalues	25.081	1.715	1.057	0.867	0.834

- Scree plot



- Parallel analysis:
  - combines exploratory factor analysis and simulation studies (Horn, 1966)
  - Eigenvalues > simulated eigenvalues

	1	2	3	4	5
Eigenvalues	21.972	1.435	1.034	0.880	0.816
Simulated Eigenvalues	1.369	1.342	1.311	1.296	1.281

NONLINEAR FACTOR ANALYSIS

- Problem of linear factor analysis: overestimate the number of factors
  - item difficulty is sometimes mistaken for a latent variable (Carroll, 1945; McDonald & Ahlawat, 1974)
- NOHARM, normal ogive harmonic analysis robust method (Fraser & McDonald, 2003)

	1 factor	2 factors	4 factors
sum of squares of residuals (SSR)	0.0093	0.0040	0.0026
root mean square of residuals (RMSR)	0.0033	0.0022	0.0017
Tanaka index	0.9975	0.9989	0.9993

NONPARAMETRIC METHOD

- Use conditional covariance to analyze
- DIMTEST (Stout, 1987; Stout, Froelich & Gao,2001):
  - H0: essential unidimensionality vs
    H1: essential multidimensionality

	RESULTS
T	0.4696
P-value	0.3193

Result: do not reject H0 (unidimensional)

	RESULTS
T	2.1946
P-value	0.0141

Result: reject H0 (unidimensional) → multidimensional

- DETECT (Zhang & Stout, 1999a, 1999b):
  - the data must conform to the approximate simple structure, meaning that one item only measures one dimension (more accurate results)

- Maximum DETECT value (Kim, 1994)
  >1, large multidimensionality
  0.4~1, moderate to large multidimensionality
  <0.4, weak multidimensionality
  <0.2, unidimensionality

- DIMPACK v1.0
  - DIMTEST & DETECT
  - Limitation: 7000 samples

TWO ISSUES IN EXISTING STUDIES ON L2 READING DIMENSIONALITY

- mostly applied explorary and confirmatory factor analysis, be it L1 or L2 (e.g., Kong & Li, 2009; Meneghetti, Carretti, & De Beni, 2006; Rost, 1993; Song, 2008; van Steensel, Oostdam, & van Gelderen, 2013; Zwick, 1987)
  - Few lanugage test studies implemented other statistical techniques, such as DIMTEST, DETECT, or NOHARM (e.g., Jang& Roussos, 2007; Kim & Jang, 2009; Schedl, Thomas, & Way, 1996)

- tests being analyzed (e.g., TOEFL)→ more proficient learners
  - lack observations on learners with low proficiency
- Weir and Porter (1994): skill divisibility might be a function of the proficiency level
  - Proficient readers → unidimensional
  - Less proficient readers → possibly multidimensional
- Alderson (2000): skills are more identifiable for beginning, weak, dyslexic or low-level second-language readers before their skills are matured and become integrated during the reading process
- May find multidimensionality of reading comprehension with less proficient readers (Alderson, 2000; Weir & Porter, 1994)
  - Taiwan EFL students (junior high school students): ALTE level 1, CEFR A2, and ACTFL intermediate

RESEARCH METHOD

BCTEST

- Basic Competence Test for Junior High School Students (BCTEST)
- a standardized achievement exam for 5 subjects, including English, Chinese, social studies, natural science, and math
- all junior high school students upon graduation in Taiwan

RESEARCH METHOD

- BCTEST 2009, 2010, and 2011
  - Conducted twice an year (May and July)
  - Combined the reading comprehension items from both tests

	May	July	Sum
2009	21	23	44
2010	24	25	49
2011	21	21	42

(27-28)

Below is an ad of Carefree Helpline.

Mom hates. Dad forgets. Brothers and sisters never care.
Friends lie. People laugh. Even your dog turns away.
You feel lost. Nobody helps.
You talk. Nobody listens.

On Helpline: You talk. We listen.
We listen to all you would like to share.
It costs nothing to ask for help.
Call 080-911911 anytime, from Monday to Sunday.
Carefree Helpline cares for you, day and night.

☎ *To talk face to face, please call 080-122133.*

🕮
ad 廣告

27. Which is NOT said about Carefree Helpline?
(A) They listen to people in trouble.
(B) People can talk face to face with them.
(C) They collect money for people in need.
(D) People can call them anytime, any day.

28. What does your dog turns away mean in the reading?
(A) Your dog bites other people.
(B) Your dog does not care about you.
(C) Your dog is kicked and runs away.
(D) Your dog cannot find its way home.

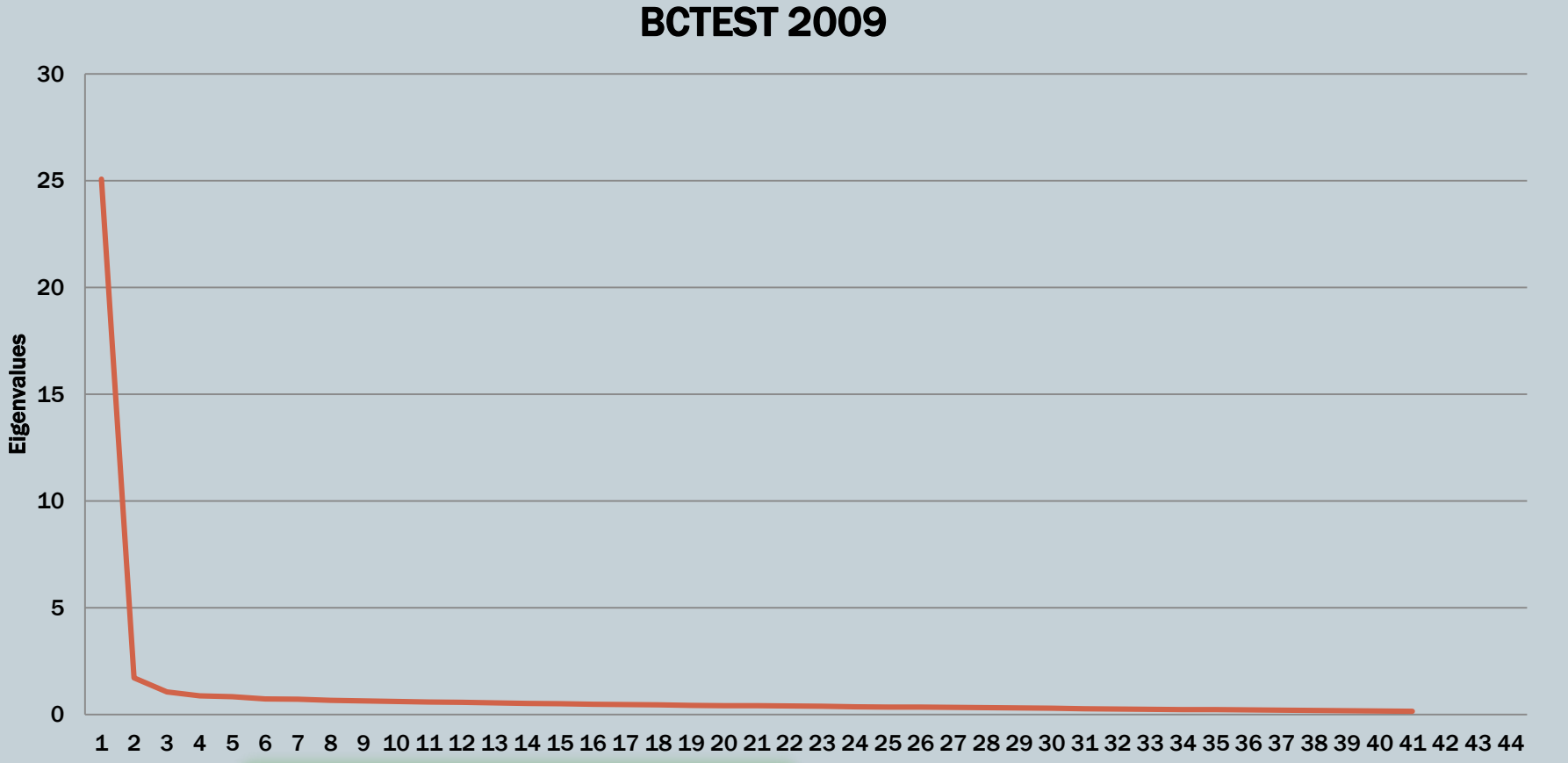
- Literal comprehension:
  - Extraction: retrieve required information from the text
  - Integration: locate relevant pieces of information and integrate them to understand the main idea of the text or to obtain the answer
- Inferential comprehension:
  - Local inference: locate relevant information (usually 2 or 3 sentences) and infer its embedded meaning or message
  - Global inference: : incorporate relevant information throughout the text (sometimes in conjunction with background knowledge) and infer its embedded meaning and message

Skill	Sub-skill	2009	2010	2011
Literal comprehension	Extraction (local)	19	21	14
	Integration (global)	15	10	12
Inferential comprehension	Local inference	7	11	8
	Global inference	3	7	8

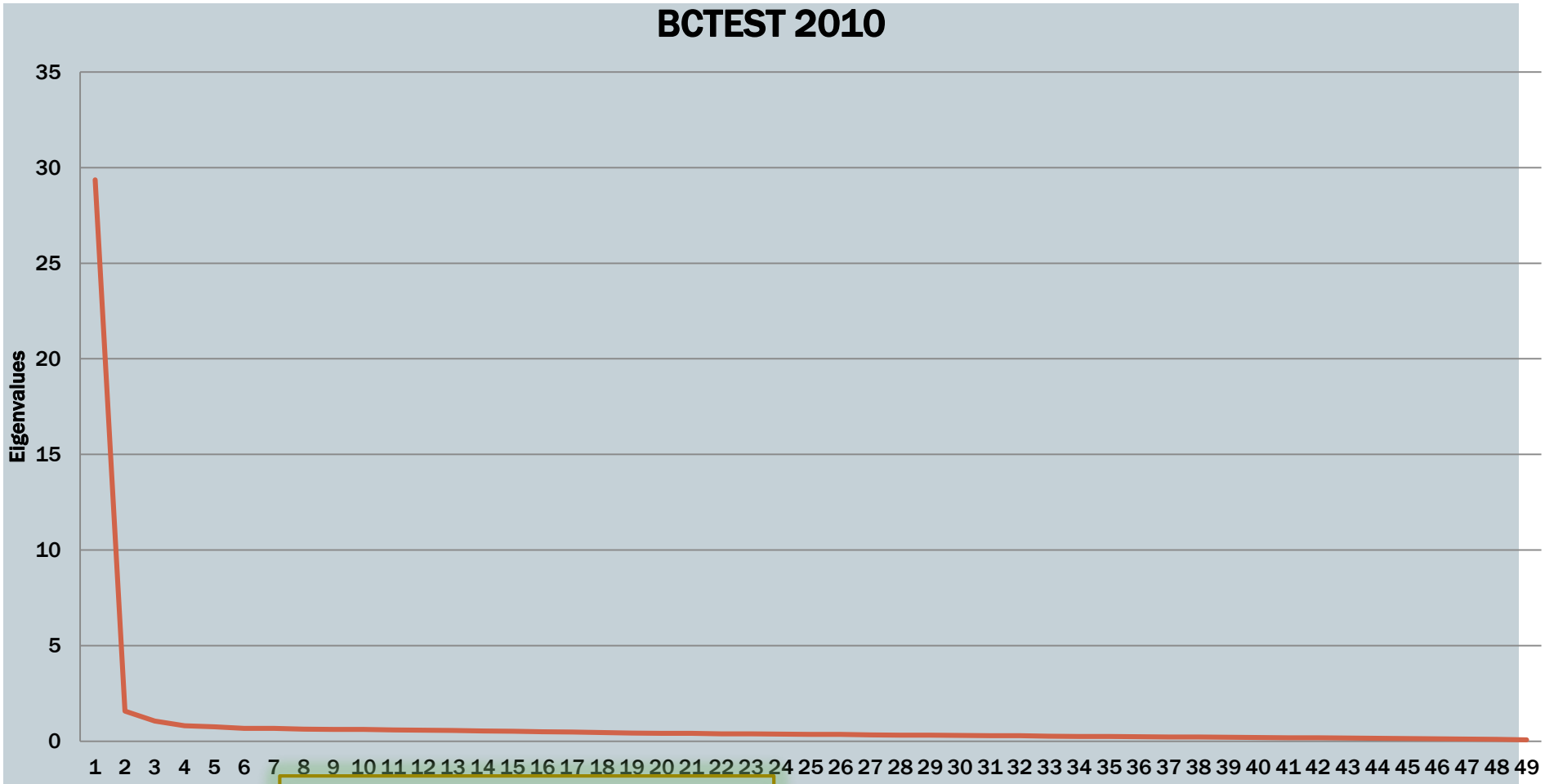
- Each year: random 7,000 participants
  - Due to the limitation of DIMPACK (7000 only)
  - Total: 21,000 participants

- Conduct EFA, NOHARM, DIMTEST, and DETECT

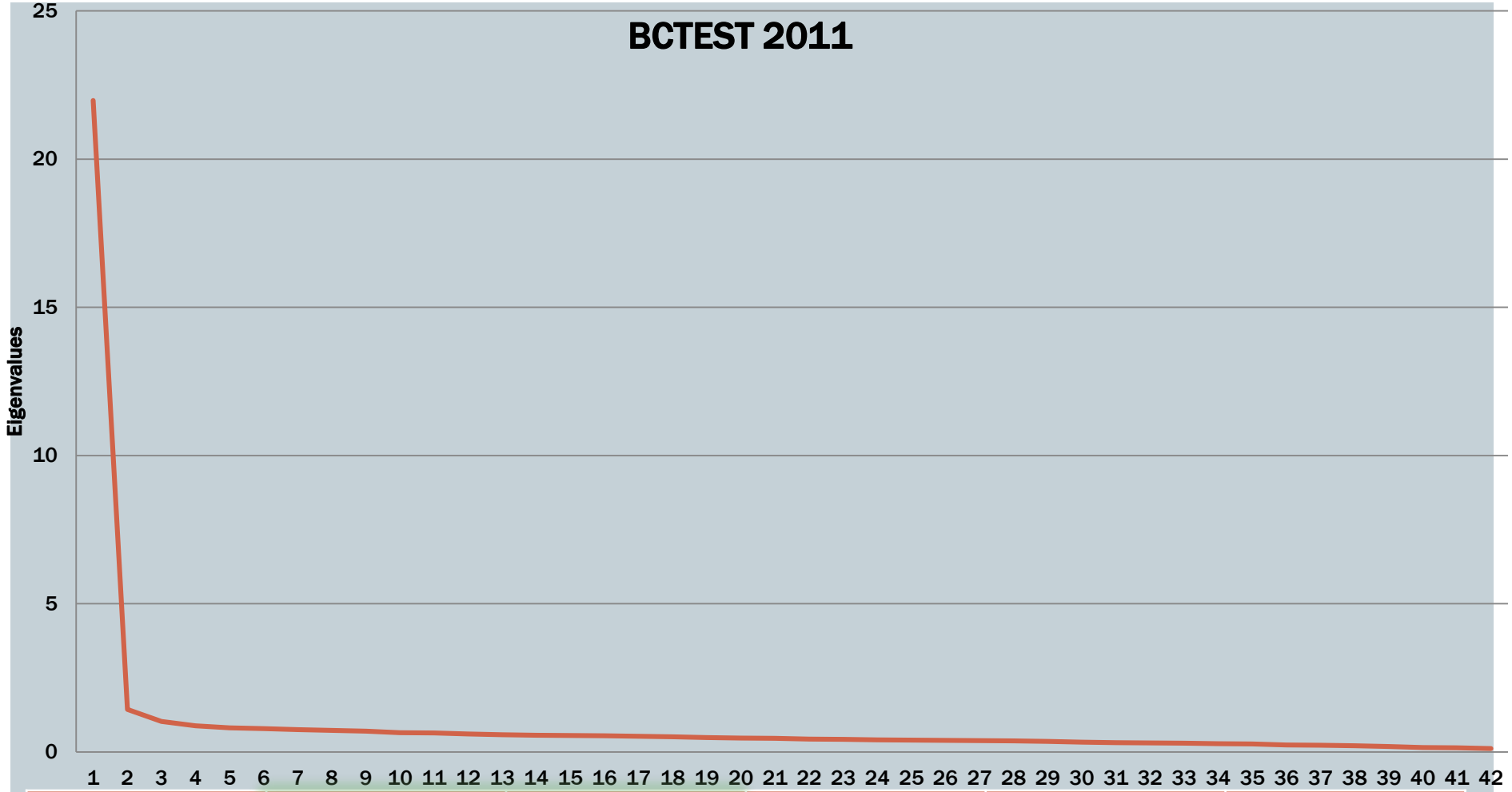
RESULT



	1	2	3	4	5
Eigenvalues	25.081	1.715	1.057	0.867	0.834
Simulated Eigenvalues	1.513	1.452	1.416	1.393	1.332
Accounted variance	0.570	+ 0.033 = 0.603			



	1	2	3	4	5
Eigenvalues	29.359	1.574	1.048	0.810	0.748
Simulated Eigenvalues	1.539	1.493	1.466	1.433	1.381
Accounted variance	0.599	+ 0.032 = 0.631			



	1	2	3	4	5
Eigenvalues	21.972	1.435	1.034	0.880	0.816
Simulated Eigenvalues	1.369	1.342	1.311	1.296	1.281
Accounted variance	0.523	+ 0.034 = 0.557			

BCTEST 2009

Chi-square test for difference testing	1 vs 2 (lit and inf)	1 vs 2 (loc and glob)	1 vs 4
Value	0.026	*warning message	*warning message
Degree of freedom	1		
P-value	0.8722		

	1 factor
RMSEA	0.024
CFI	0.991
TLI	0.991
WRMR	1.399

BCTEST 2010

Chi-square test for difference testing	1 vs 2 (lit and inf)	1 vs 2 (loc and glob)	1 vs 4
Value	*warning message	6.502	*warning message
Degree of freedom		1	
P-value		0.0108	

	1 factor
RMSEA	0.018
CFI	0.995
TLI	0.995
WRMR	1.189

BCTEST 2011

Chi-square test for difference testing	1 vs 2 (lit and inf)	1 vs 2 (loc and glob)	1 vs 4
Value	1.661	*warning message	*warning message
Degree of freedom	1		
P-value	0.1975		

	1 factor
RMSEA	0.019
CFI	0.994
TLI	0.994
WRMR	1.166

NOHARM

BCTEST 2009	1-factor	2-factor	4-factor
Sum of squares of residuals	0.0175064	0.0126149	0.0114454
Root mean square of residuals	0.0043018	0.0036517	0.0034783
Tanaka index	0.9951499	0.9965051	0.9968291

BCTEST 2010	1-factor	2-factor	4-factor
Sum of squares of residuals	0.0230873	0.0217678	0.0195405
Root mean square of residuals	0.0044308	0.0043023	0.0040763
Tanaka index	0.9942921	0.994524	0.9950843

BCTEST 2011	1-factor	2-factor	4-factor
Sum of squares of residuals	0.0164699	0.0152079	0.0122780
Root mean square of residuals	0.0043736	0.0042027	0.0037763
Tanaka index	0.9961087	0.9964069	0.9970991

Result: unidimensional

DIMTEST

BCTEST 2009	T	P-value
Trial 1	0.7935	0.2138
Trial 2	0.4621	0.3220
Trial 3	0.8687	0.1925

BCTEST 2010	T	P-value
Trial 1	0.4696	0.3193
Trial 2	1.3067	0.0957
Trial 3	0.5062	0.3063

BCTEST 2011	T	P-value
Trial 1	-0.9864	0.8380
Trial 2	1.4442	0.0743
Trial 3	1.0958	0.1366

Result: unidimensional

DETECT

	Maximum DETECT value
BCTEST 2009	0.1075
BCTEST 2010	0.0803
BCTEST 2011	0.1131

Maximum DETECT value (Kim, 1994)
>1, large multidimensionality
0.4~1, moderate to large multidimensionality
<0.4, weak multidimensionality
<0.2, unidimensionality

Result: unidimensional

SUM UP

- EFA (+ parallel analysis): the first factor accounted most of the variance (.52-.60)
- CFA: one factor (except for the bctest 2010:local and global)
- NOHARM
  - SSR, RMSR, and Tanaka → 4 factors (but the differences are actually very small) → essentially 1 factor
- DIMTEST
  - P-value > .05 → don't reject HO unidimensional
- DETECT
  - Maximum DETECT values < .2 → unidimensional

DISCUSSION

POSSIBLE CONSTRAINTS OF THE ITEMS

- MC items – students are limited to those options even when they may come up with their own unique interpretation which is equally legitimate

- "the very act of assessing and testing will inevitably affect the reading process, and the fact that a learner has answered a question posed by a tester incorrectly does not necessarily mean that he or she has not understood the text in other ways or to his or her own satisfaction." (Alderson, 2005, p. 120)

Sophia: The pizzas here are very good. Do you want some?
Takako: Yeah, sure. Look! They have **artichokes** for the pizza *La Primavera*. What is an artichoke?
Sophia: Well, it is a big flower. It has a **heart** in it. People take the heart and use **it** in salad or pizza. You can buy them in supermarkets. There is one near the train station. We may go there later. Here in **Italy**, people make pizzas with artichoke hearts.
Takako: Cool! I want the pizza *La Primavera* then!
Sophia: Great. Look! Your favorite chocolate ice cream comes with it. Isn't it wonderful?
Takako: I can't wait!

Dictionary: artichoke 朝鮮薊(一種蔬菜); heart 菜心; Italy義大利

- According to the reading, where are Takako and Sophia?

Answer: a restaurant

Other plausible answers: a place in a train station which sells pizza / in a train station

- Local vs. Global
  - In contrast to TOEFL or WB-ESLPE, TEM4, items are short and easy.
    - One or two paragraphs maximum →the distinction between local and global skills did not differ much

DIMTEST	BCTEST 2009	BCTEST 2010	BCTEST 2011
Local vs. Global (T and p-value)	-1.7032 (0.9557)	-1.5662 (0.9414)	0.4071 (0.3419)

FINAL REMARKS

- Results are not meant to be generalized to other contexts (BCTEST → EFL in Taiwan).
- BCTEST: standardized assessment (IRT)
- Currently, developing a reading comprehension test, covering from elementary to senior high school in Taiwan
  - Only removed the items which had low discriminative power (2 or 3 items only)
  - Conducted some initial analyses on dimensionality (gr 7 and 8)
  - Still unidimensional
- Psychological vs. psychometric dimensionality (Henning, 1992)
  - Psychometrics can be confounded by the sample and the items being implemented.