

An Argument-Based Validation for the KIMEP English Entrance Test (KEET)

Marina Gerassimenko,
University of Leicester

Kazakhstan Institute of Management,
Economics and Strategic Research (KIMEP)



Outline

- ❑ Background
- ❑ Validity as an argument
 - Toulmin's structure of an argument
 - Kane (1992, 2006) and Chapelle et al. (2008)
- ❑ The KEET's validity argument
- ❑ Summary of the KEET's validity argument
- ❑ Lessons learned

Background

KIMEP: Kazakhstan Institute of Management, Economics and Strategic Research, Almaty, Kazakhstan, Central Asia

An English-medium university

Degrees in economics, management, finance

87% undergraduate students; 13% graduate; 3.6% international



The KEET:

paper-based

100 items

multiple-choice

three sections – Grammar, Vocabulary in Use, Reading

Comprehension

parallel forms

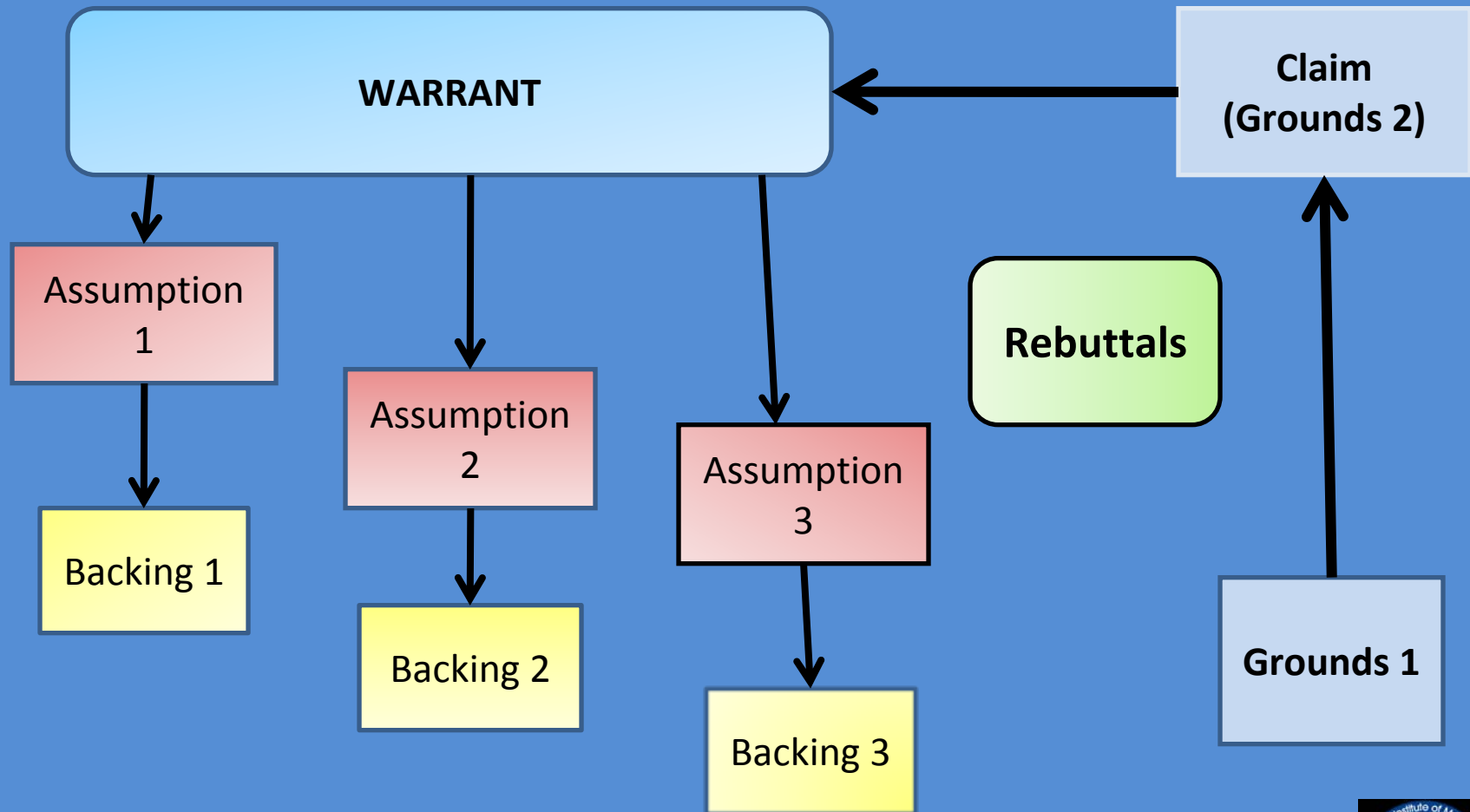


A validity argument

“A validity argument should present and integrate evidence and rationales from which validity conclusion can be drawn pertaining to particular score-based inferences and uses of a test.”

Chapelle (1999:263)

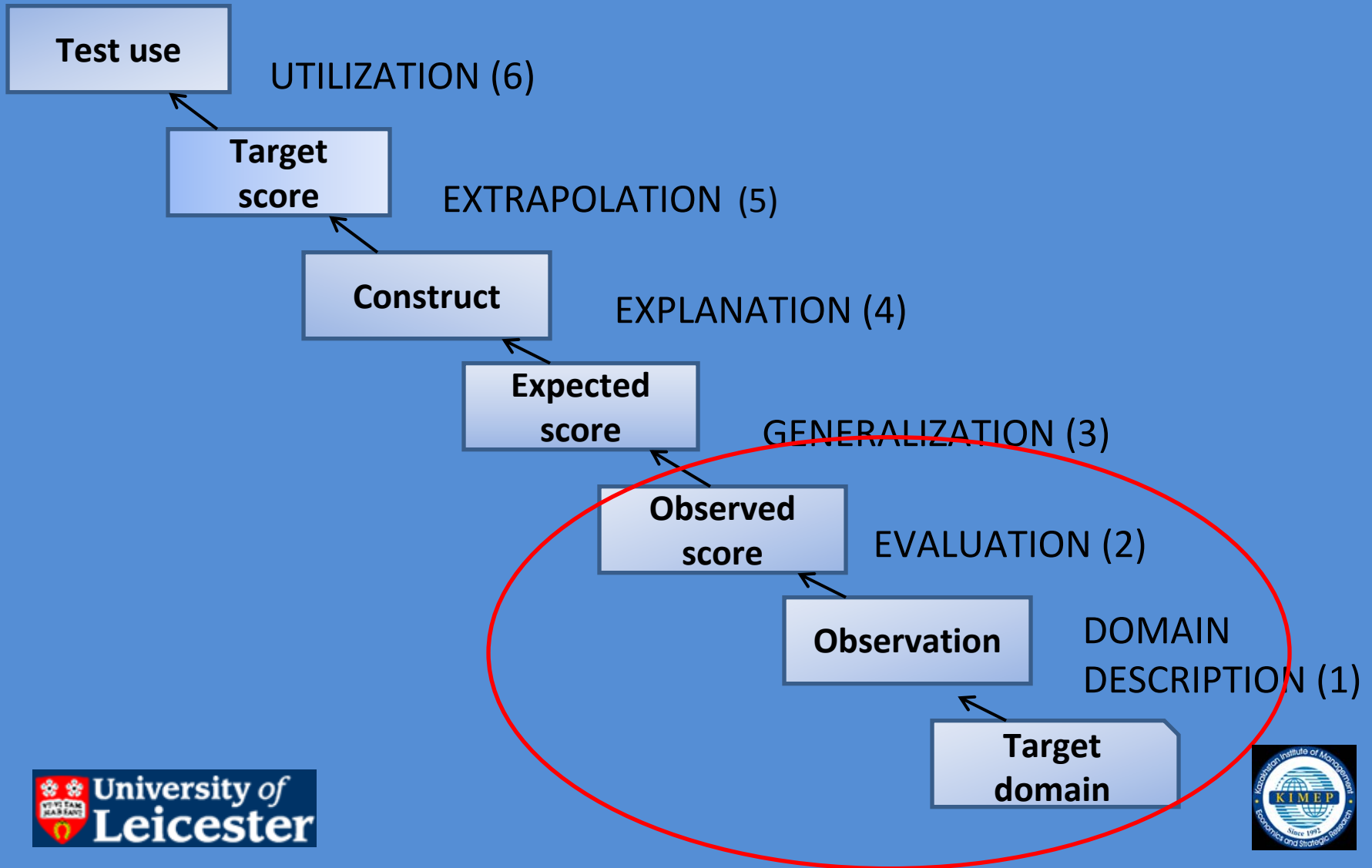
An Argument Structure (Toulmin, 1958, 2003) :



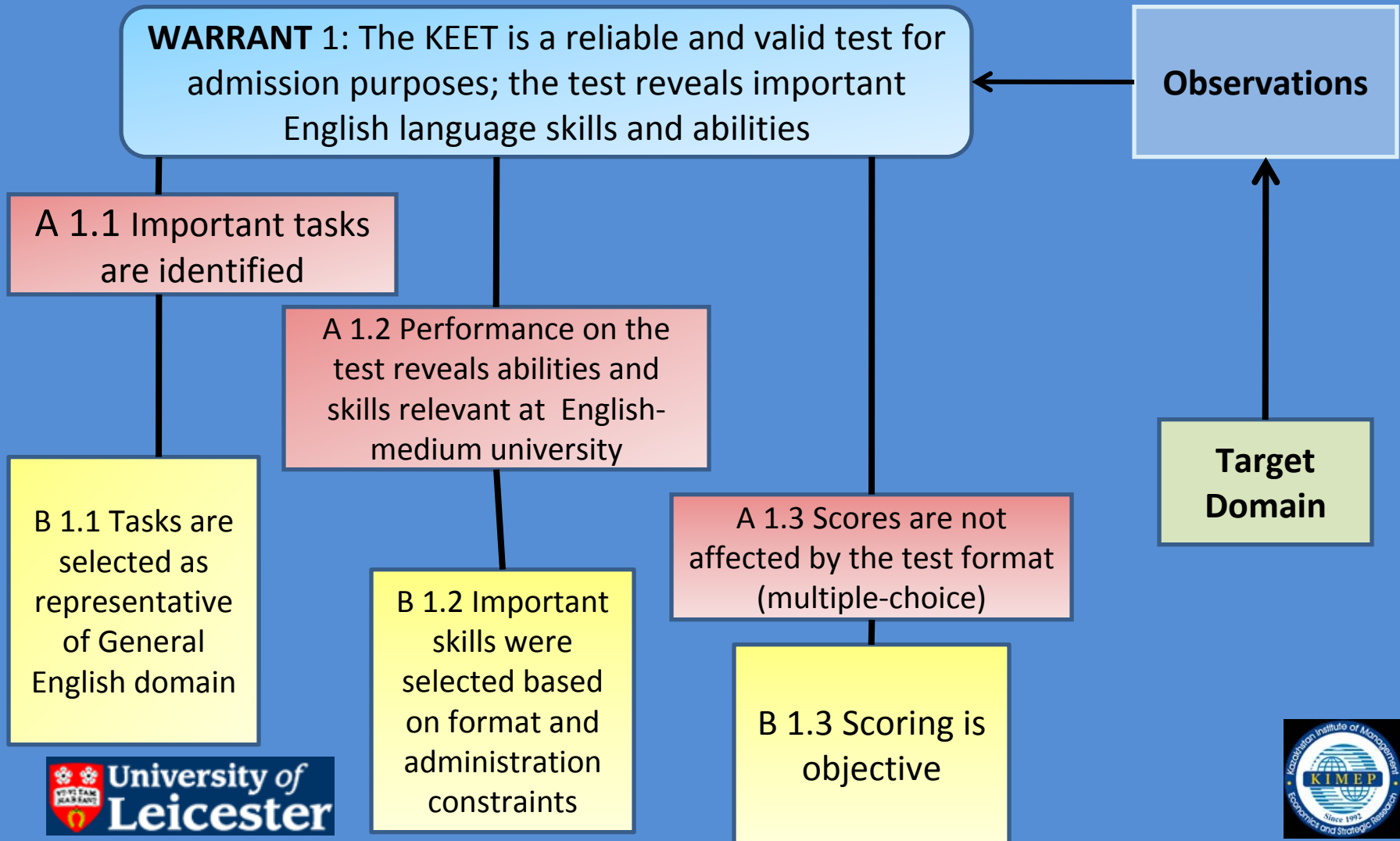
Kane (1992, 2006): Two Stages of Arguments

- An interpretative argument specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances
- The validity argument provides an evaluation of the interpretative argument

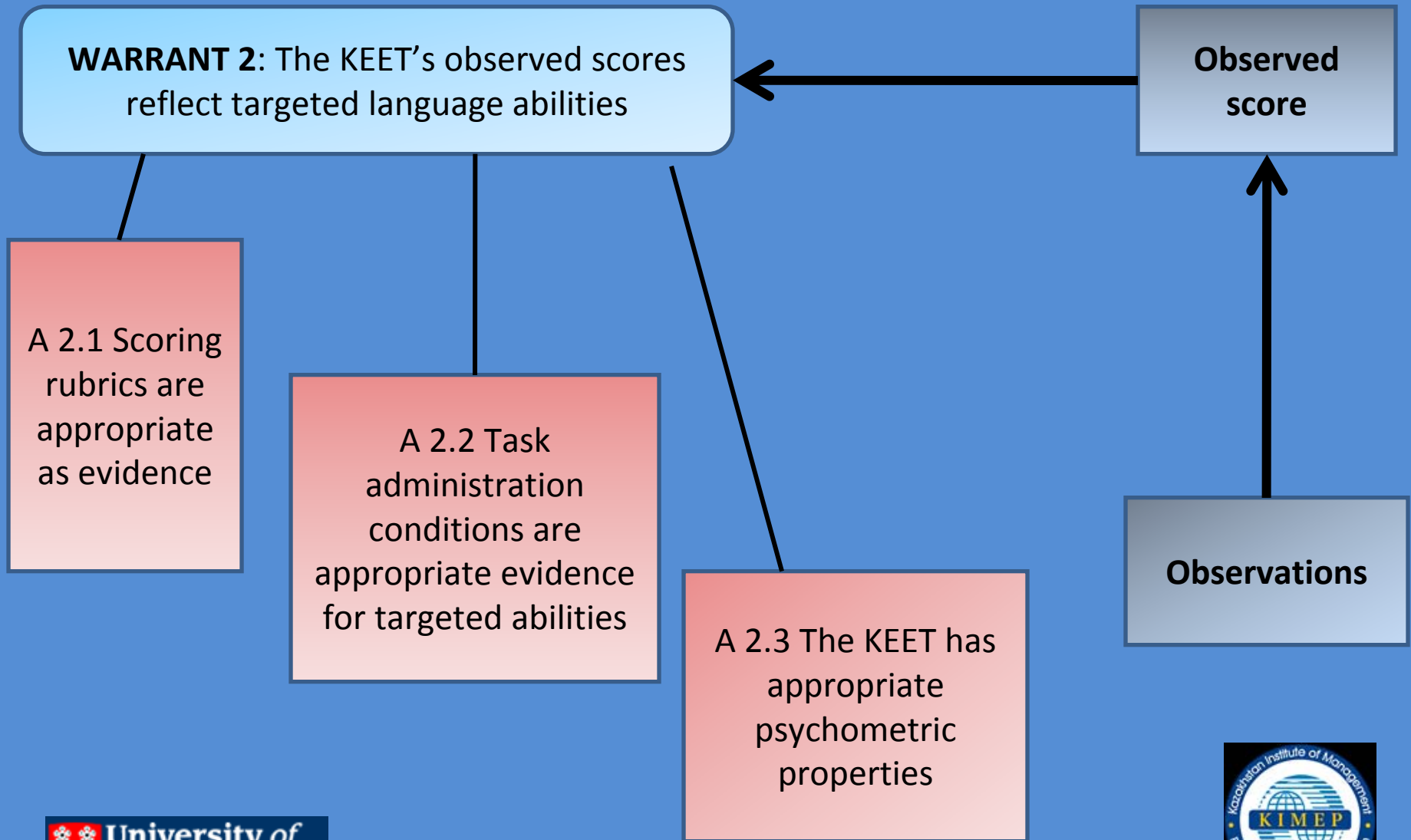
Chapelle et al. (2008)



Step 1: The Domain Description Inference



Step 2: The evaluation Inference



B2.1

Rubrics:
**grammar, use
of English,
reading
comprehension**
provide
evidence for
targeted
abilities

B 2.2 Task
administration
conditions are
standardized
by Admission
Office
protocol

B 2.3 Acceptable
for standardized
tests
**coefficients
Alfa:**
form A **.935**;
form B **.922**
Both forms have
**distribution
close to normal**

Backing 3.3: The KEET psychometric data for forms A and B:

Scale	Form A	Form B
N of items	100	100
Sample size	232	241
Mean	58.375	54.095
Variance	279.467	232.584
Median	58.000	55.000
Std. Dev.	16.717	15.251
SEM	4.235	4.273
Mean p	0.584	0.541
Alfa	0.935	0.922

ITEMAN software



Research Questions:

How significant is the difference?

Why do parallel forms have different means, variances and SD?

Do parallel forms have items of similar difficulty level?



How parallel is parallel?

Research methods: (1) parallel test theory as part of Classical Test Theory (Kline, 2005):

- Modes
- Means
- Medians
- Variances

(2) t-test: how significant is the difference?

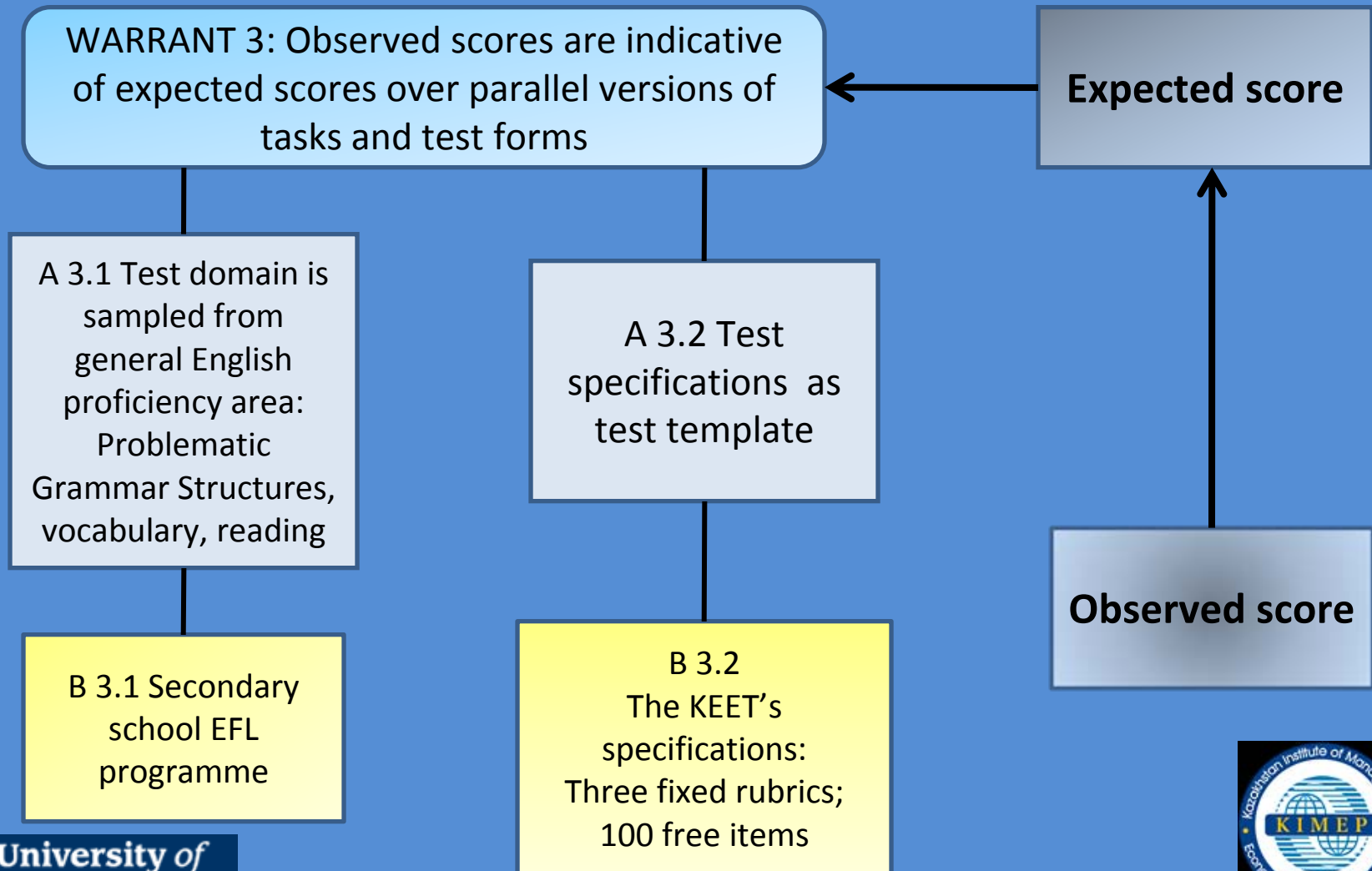
(3) Cohen d-test: what is the effect-size?

Findings:

(2) $T = 2.9108$, df (degrees of freedom) is 471, 2-tail confidence at 99.62% the difference in **means** is **significant**;

(3) Cohen's $d = 0.27$ Cohen (1988) – not significant; Wolf (1986) – **practically significant**

Step 3: The Generalization Inference



Backing for assumption 3.2

Test Specifications

R 3.2

Test Specifications : fixed and free elements

RESEARCH QUESTION:

could free elements in test specifications cause test SD, means vary?

METHODOLOGY:

p-values compared across sections of grammar, vocabulary in context and reading comprehension and across two parallel forms

FINDINGS:

Grammar	Form A	Form B
Mean p-value	.66	.59
P-value range	.11 - .91	.18 - .97

Vocabulary	Form A	Form B
Mean p-value	.49	.57
P-value range	.09 - .87	.17 - .89

Reading comprehension	Form A	Form B
Mean p-values	.52	.34
P-values range	.38 - .76	.12 - .60

The KEET example

Form A

14. This time next week, we.....
the chemistry exam.

- A. had finished
- B. have been finishing
- C. will have finished
- D. will have been finishing

Prop.	Disc.	Point
Correct	Index	Biser.
0.61	0.31	0.24

Form B

14. Don't worry – I writing my
report by Friday next week. I'll
give it to you by then.

- A. finish
- B. will be finished
- C. am finishing
- D. will have finished

Prop.	Disc.	Point
Correct	Index	Biser.
0.72	0.56	0.50

Reading Comprehension

	Form A	Form B
Length	600	700
Topics	General interest	Specific interest
Discourse features	Narrative	Narrative imbedded with descriptions
Vocabulary	General	Specific
Question types	Asking for explicit information	Asking for inferences

Summary of the claims about the KEET

- ❑ 1st claim: the test purpose
- ❑ 2nd claim: the test construct
- ❑ 3d claim: the test specifications
- ❑ 4th claim: the item analysis
- ❑ 5th claim: the use of the test scores

Conclusion: Lessons Learned

- The KEET's construct
- The test specifications
- Analysis of parallel forms
- Item analysis

Thank you for your attention!

