外语教师评价素养发展：
理论框架和路径探索

上海交通大学 金艳 2018年4月20日 北京

提要

语言评价素养的理论框架
Conceptual Framework of LAL

语言评价素养研究
Research of LAL

语言评价素养的发展路径
Development of LAL

1

语言评价素养的理论框架

LAL: Conceptual Framework

语言评价素养的研究背景

- 语言测试用途不断拓展、专业化程度不断提高
  (Brindley, 2001; Davies 2008; Spolsky 2008)

- 教育问责制的不断深化，评价的社会性
  (McNamara & Roever, 2006; Chalhoub-Deville, 2016)

- 评价素养对教师职业发展和改进教学的重要性
  (Fulcher 2012; Popham 2009; Rea-Dickins 2008)

- 语言评价素养的认识不断加深，从社会建构主义视角更加全面地理解和定义语言评价素养
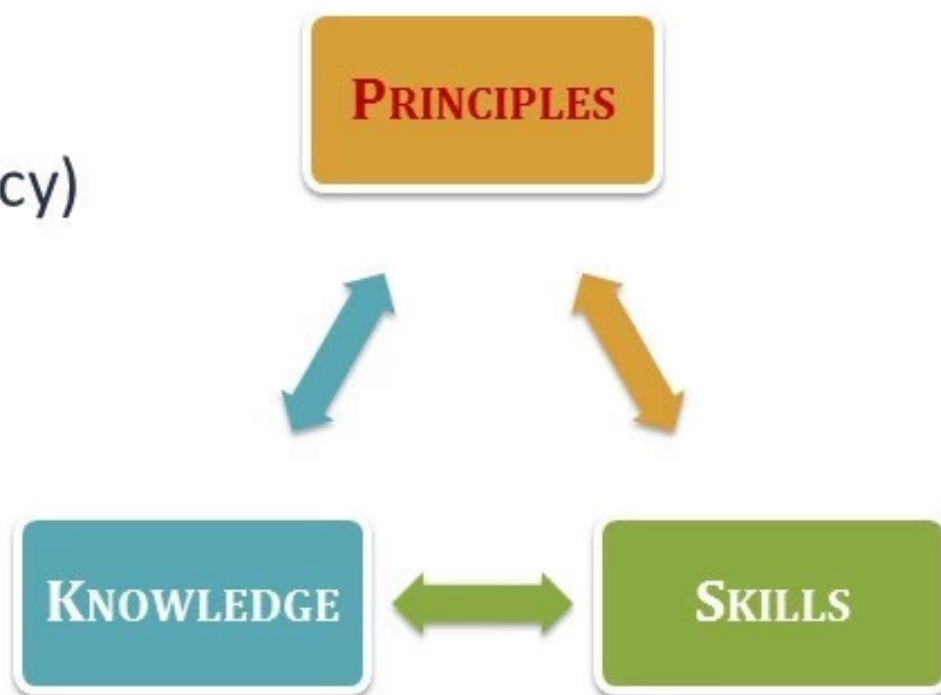 (Inbar-Lourie 2008; Taylor 2009; Xu & Liu 2009)

语言评价素养的概念

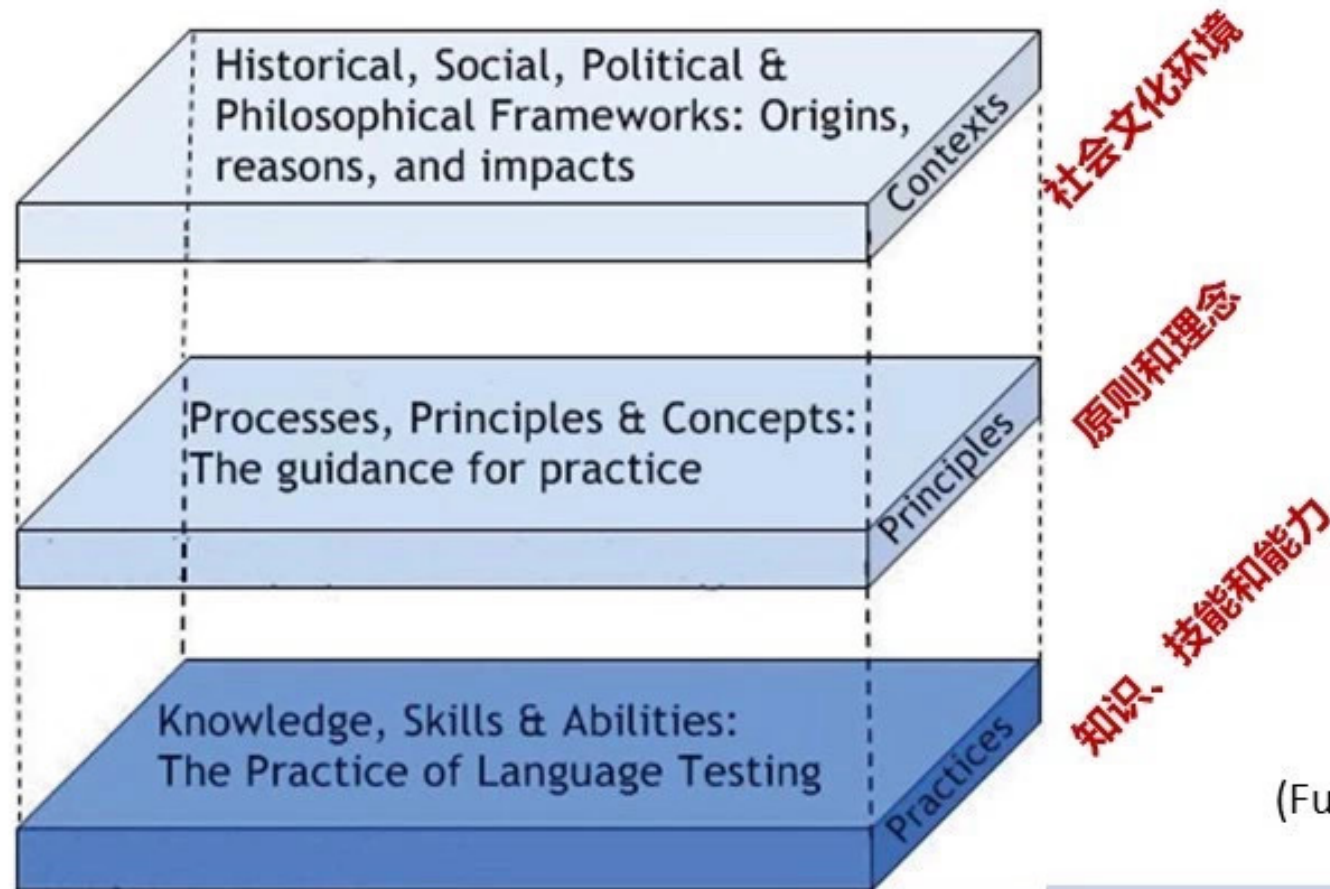依据实践经验界定LAL
(Language Assessment Literacy)

(Davies, 2008)

P: 考试伦理道德及其社会性

K: 语言描述和测量方面知识

S: 考试设计和分析实践能力

PRINCIPLES

KNOWLEDGE

SKILLS

基于实证研究数据分析: An expanded definition

Historical, Social, Political & Philosophical Frameworks: Origins, reasons, and impacts — Contexts — 社会文化环境

Processes, Principles & Concepts: The guidance for practice — Principles — 原则和理念

Knowledge, Skills & Abilities: The Practice of Language Testing — Practices — 知识、技能和能力

(Fulcher, 2012: 126)

FIGURE 1 Language assessment literacy: An expanded definition

实践层面（**practices**）

" The knowledge, skills and abilities required to **design**, **develop**, **maintain** or **evaluate**, large-scale standardized and/or classroom based tests;

开发、实施和评价大规模标准化考试和课堂评价的知识、技能和能力；

(Fulcher, 2012: 125)

原则层面（**principles**）

" **Familiarity** with test processes, and **awareness** of principles and concepts that guide and underpin practice, including ethics and codes of practice;
熟悉考试过程，了解指导考试实践的原则和理念，包括道德规范和行为准则；
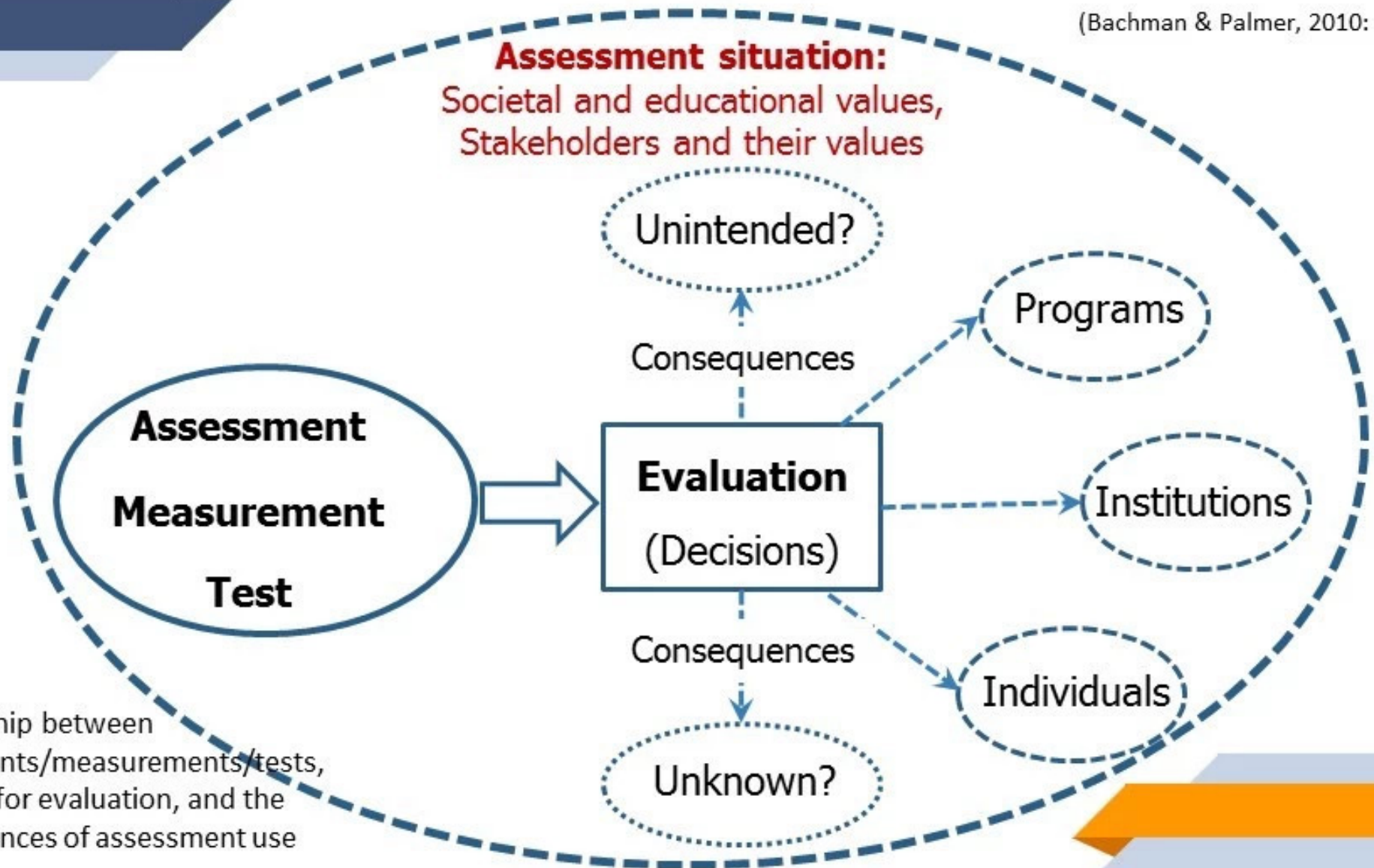
(Fulcher, 2012: 125)

环境层面（**contexts**）

" The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order **understand** why practices have arisen as they have, and to **evaluate** the role and impact of testing on society, institutions, and individuals.

(Fulcher, 2012: 125)

在更广泛的历史、社会、政治和哲学背景下运用评价知识、技能、原则和理念，以更好地理解评价的产生机制及其对社会、机构和个人的影响。

Assessment situation:
Societal and educational values,
Stakeholders and their values

Unintended?

Consequences

Assessment

Measurement

Test

Evaluation
(Decisions)

Programs

Institutions

Consequences

Individuals

Unknown?

Relationship between
assessments/measurements/tests,
their use for evaluation, and the
consequences of assessment use

Assessment Use Argument (AUA)

Claims

1. 考试的后果有益

2. 决策具有价值敏感性且公平公正

3. 分数解释有意义、公平、可靠、相关、充分

4. 考试成绩稳定

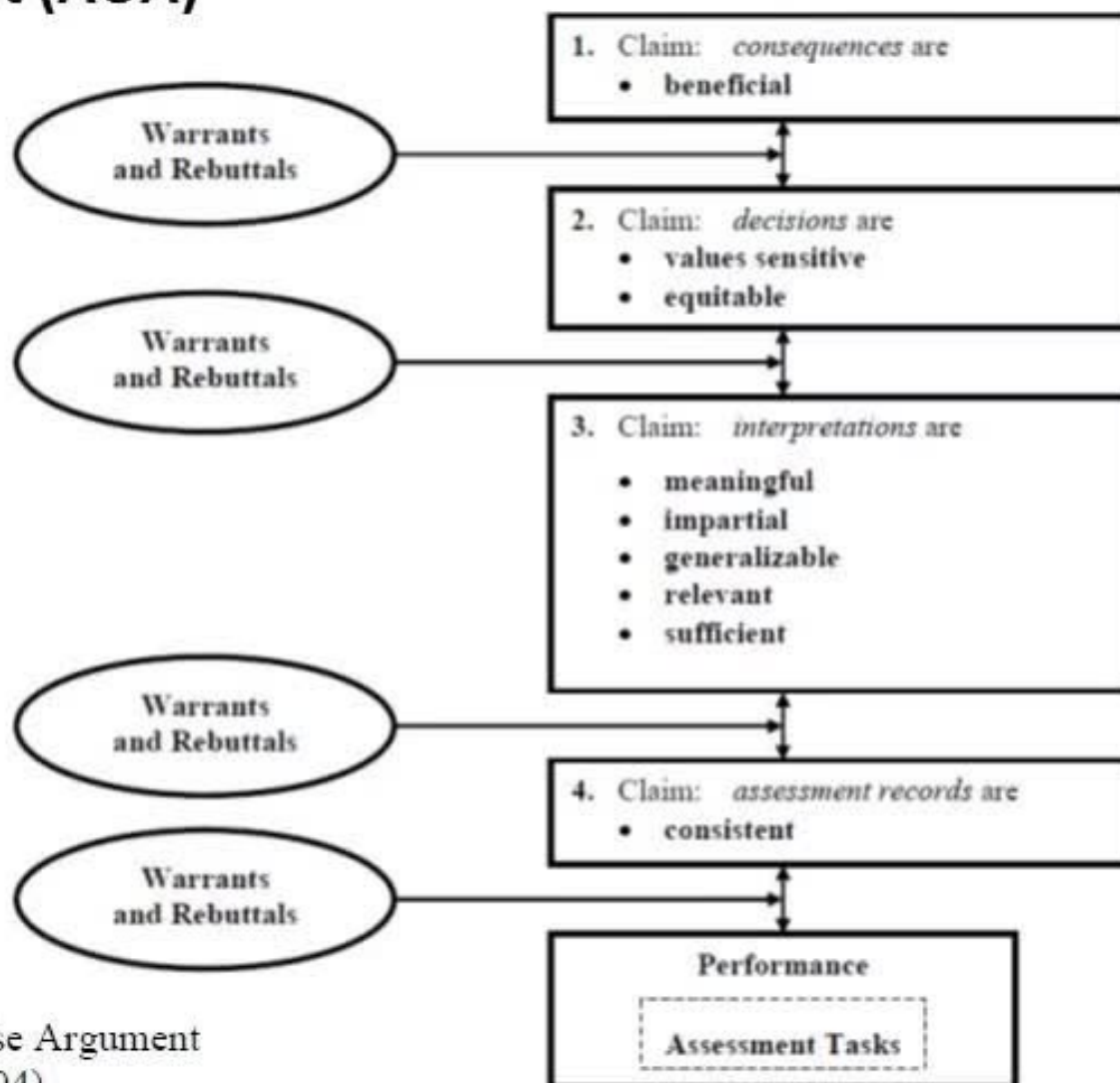Figure 1: Structure of an Assessment Use Argument (Bachman & Palmer, 2010, p. 104)

Qualities of Outcomes of Claims in an AUA

Warrants and Rebuttals

1. Claim: *consequences* are
   - **beneficial**

Warrants and Rebuttals

2. Claim: *decisions* are
   - **values sensitive**
   - **equitable**

3. Claim: *interpretations* are
   - **meaningful**
   - **impartial**
   - **generalizable**
   - **relevant**
   - **sufficient**

Warrants and Rebuttals

4. Claim: *assessment records* are
   - **consistent**

Warrants and Rebuttals

Performance

Assessment Tasks

11

大规模考试的开发和使用

开发设计
- ▶ 考试目的
- ▶ 考试内容
- ▶ 考试题型

实施评价
- ▶ 考试实施
- ▶ 评分计分
- ▶ 分数解释

使用评价
- ▶ 教学决策
- ▶ 其他决策
- ▶ 决策后效

社会环境	• 社会影响 • 教学后效
原则理念	• 道德规范 • 行为准则
考试实践	• 开发、实施 • 评分、报告

语言评价素养观

Postmodern approach:
all knowledge and meaning
is socially constructed

后现代语言评价素养观：
知识和意义都是在社会环
境中形成且是动态变化的

Traditional approach:
prioritizing the large-scale
standardized
psychometrically driven
paradigm

传统语言评价素养观：
强调心理测量理论驱动
下的大规模标准化考试

评价

形成性
AFL, AAL

分级测试
诊断测试
课堂评价
...

终结性
AOL

期中期末
外部考试
...

课堂评价研究维度

A. 评价目的：改进教学和学习

B. 评价主体：教师和学生

C. 评价手段：多样化（定性、定量、混合）

D. 评价效果：多维度（观点、态度、动机、策略、成绩等）

课堂评价方法（示例）

考试和观察 Tests, Observations	课堂讨论 Class discussions	同伴反馈、自评 Peer feedback Self-assessment
自然交流 Naturalistic talk-in-interaction	项目 Projects Portfolios	即时评价 Moment-to-moment evaluations

课堂评价研究综述

☐ 研究方法：课堂观察、问卷调查、文本/话语分析等；混合方法研究较多；

☐ 国内关注整个测评体系，国外关注具体测评方式

☐ 有效性体现在多方面，例如：

　　☐ 提高自主学习能力和自我效能感

　　☐ 提高英语教学的有效性

　　☐ 提高团队合作效果，改进课堂表现

　　☐ 提升思辨和反思意识 …　　　（孙杭，2018）

Learning-Oriented Assessment (LOA)

学习取向型评价

- highlights learning goals, performance evaluation and feedback

  突出学习目标、评价语言表现并提供反馈

- learning an L2 is a highly individual cognitive process

  学习是一个高度个性化的认知过程

- also a highly intricate socio-cognitive and sociocultural process

  也是一个复杂的社会认知和社会文化过程

- ... other factors (e.g., learner engagement) also play an important role in the assessment process of L2 learning

  学习者参与度等其他因素与课程、教学等因素对评价同样重要

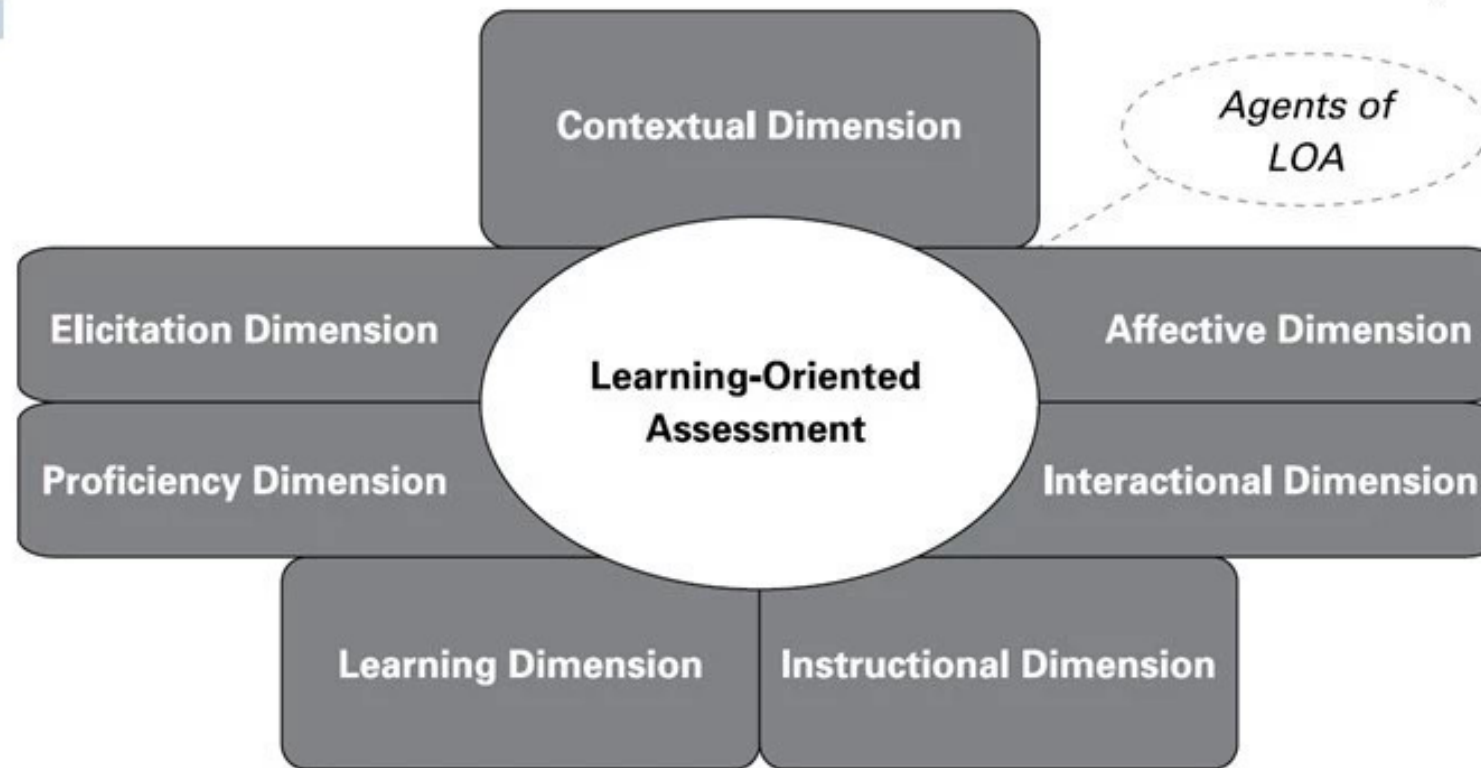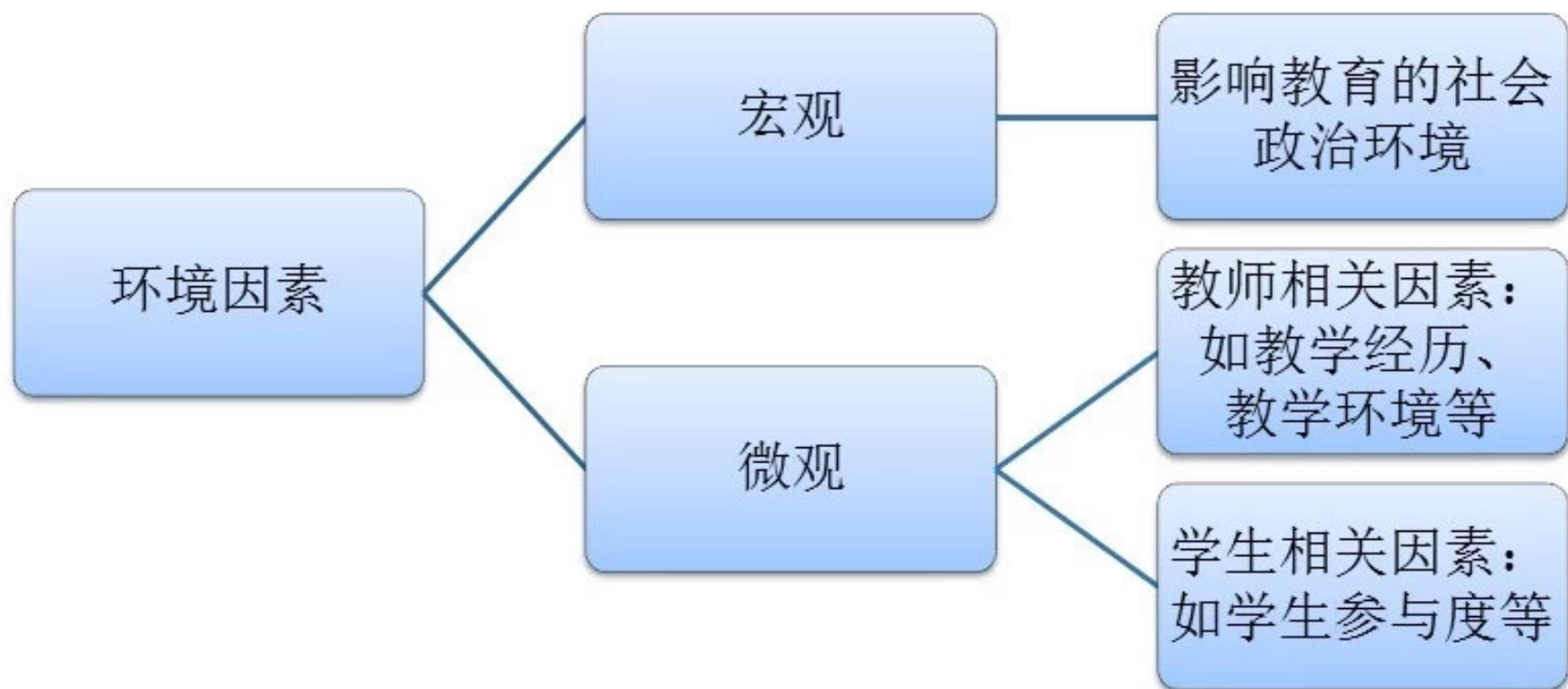# LOA中互相关联的七大关键维度

Turner & Purpura (2015: 261)

As LOA involves many *interrelated dimensions*, focusing on the individual dimensions and their relationships, across different agents, may lead to an understanding of the whole.

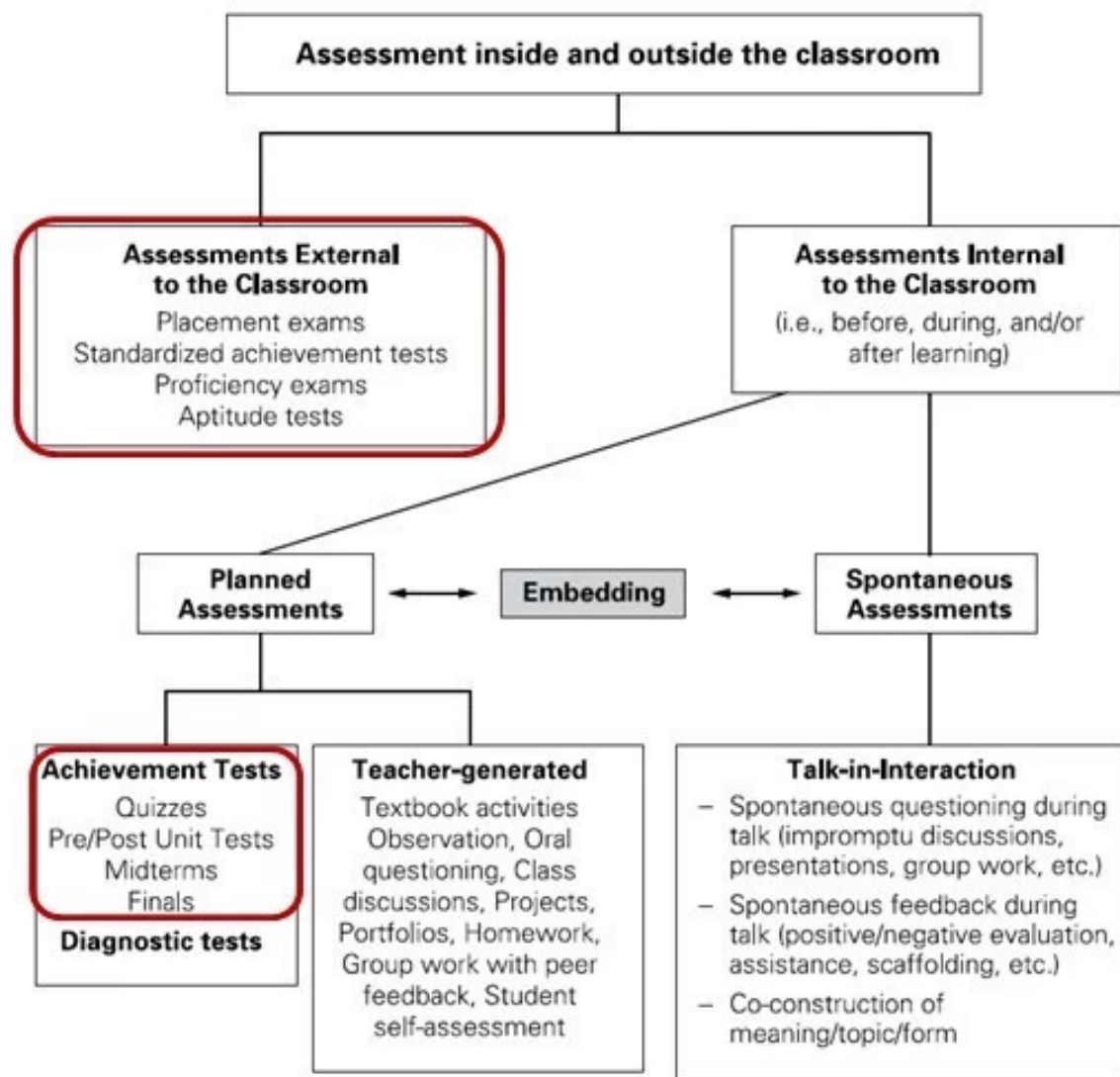**Fig. 1:** Working framework of LOA.

20

1 Contextual Dimension

```
                                    ┌─────────────────┐
                          ┌────────│      宏观         │────── 影响教育的社会
                          │         └─────────────────┘        政治环境
  ┌──────────┐           │
  │  环境因素  │───────────┤                                    教师相关因素：
  └──────────┘           │         ┌─────────────────┐          如教学经历、
                          └────────│      微观         │────────  教学环境等
                                    └─────────────────┘
                                                          ────── 学生相关因素：
                                                                  如学生参与度等
```

2 Elicitation

Assessment inside and outside the classroom

Assessments External to the Classroom
Placement exams
Standardized achievement tests
Proficiency exams
Aptitude tests

Assessments Internal to the Classroom
(i.e., before, during, and/or after learning)

Planned Assessments ↔ **Embedding** ↔ **Spontaneous Assessments**

Achievement Tests
Quizzes
Pre/Post Unit Tests
Midterms
Finals
Diagnostic tests

Teacher-generated
Textbook activities
Observation, Oral questioning, Class discussions, Projects, Portfolios, Homework, Group work with peer feedback, Student self-assessment

Talk-in-Interaction
– Spontaneous questioning during talk (impromptu discussions, presentations, group work, etc.)
– Spontaneous feedback during talk (positive/negative evaluation, assistance, scaffolding, etc.)
– Co-construction of meaning/topic/form

Fig. 2: Ways to elicit language for assessment inside and outside the classroom. How can they serve LOA?
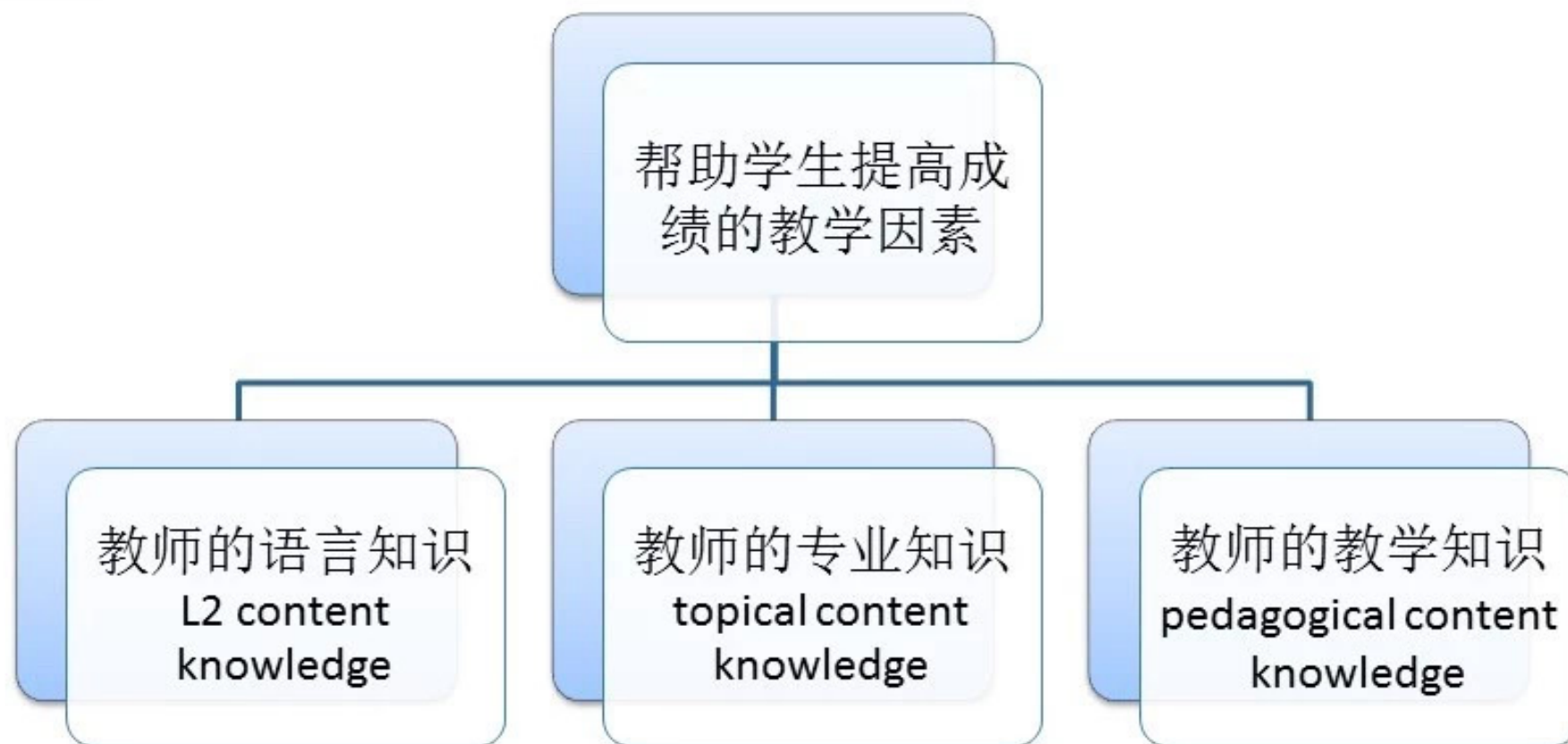
3 Proficiency

关于语言水平及其发展的理论为LOA提供指导：

- 评价什么？
  what should be assessed

- 评价结果的解释
  how evidence from performance is interpreted

- 如何提供反馈和帮助
  what should be targeted by feedback and assistance

4 Learning Dimension

学习理论 LOA and theories of learning and cognition	反馈作用 Learning, LOA, the role of feedback and assistance	自我调整 Learning, LOA and the role of self-regulation

5 Instructional Dimension

帮助学生提高成绩的教学因素

教师的语言知识
L2 content
knowledge

教师的专业知识
topical content
knowledge

教师的教学知识
pedagogical content
knowledge

6 Interactional Dimension

Exchange patterns that provide
a positive evaluation
of a learner's learning or performance (*Good!*) related to a learning goal
正面反馈

or **a negative evaluation**
(*well ...*) followed by scaffolded assistance from teachers or peers in repairing some aspect of communication or learning with hints (*did that happen today or yesterday?*)
负面反馈 + 帮助提示

or by **a more elaborated learning intervention**.

更细致干预

7 情感因素（the affective dimension）

- learners' emotions
- beliefs about learning and competence
- personality characteristics (e.g., extroversion)
- attitudes towards learning and performance (e.g., persistence)
- motivation

情感
信念
个性
态度
动机

An LOA approach: 评价、学习、教学紧密联结

" ... move beyond dichotomous depictions of CBA as formative/summative or AOL/AFL, toward an approach that characterizes **assessment, learning, and instruction, while different, as intrinsically intertwined**.

Turner & Purpura (2015)

评价环境	• 教学环境因素 • 教师和学生因素
评价理念	• 学习取向型
评价实践	• 开发、实施 • 反馈、改进

2

语言评价素养研究

Research of LAL

(林敦来、武尊民 2014)

```
语言评价素养研究
├── 与语言工作紧密相关的人员
│   ├── 语言评价知识库理论
│   ├── 语言评价培训需求
│   ├── 语言评价培训材料
│   ├── 语言评价课程
│   └── 语言评价培训手段
└── 更广大的利益相关者
    └── 语言评价熟悉程度/评价素养连续体
```

语言评价素养研究

理论框架和体系构建
- 概念界定（狭义/广义）
  - 林敦来、高淼 2011
  - 唐雄英 2013
  - 许悦婷 2013
  - 林敦来、武尊民 2014
- 对LAL的评价
  - Xu & Brown 2017

现状分析和需求调查
- LAL课程/培训内容和方法
  - Jin 2010
  - R. Lam 2015 (香港)
  - Jin & Jie 2017
- 利益相关者LAL需求和实践
  - 许悦婷、刘永灿 2008
  - Xu & Liu 2009
  - Fan & Jin 2013
  - Xu & Brown 2017
  - Jin, Zhu, & Wang, 2017
  - Zhang & Yan 2018
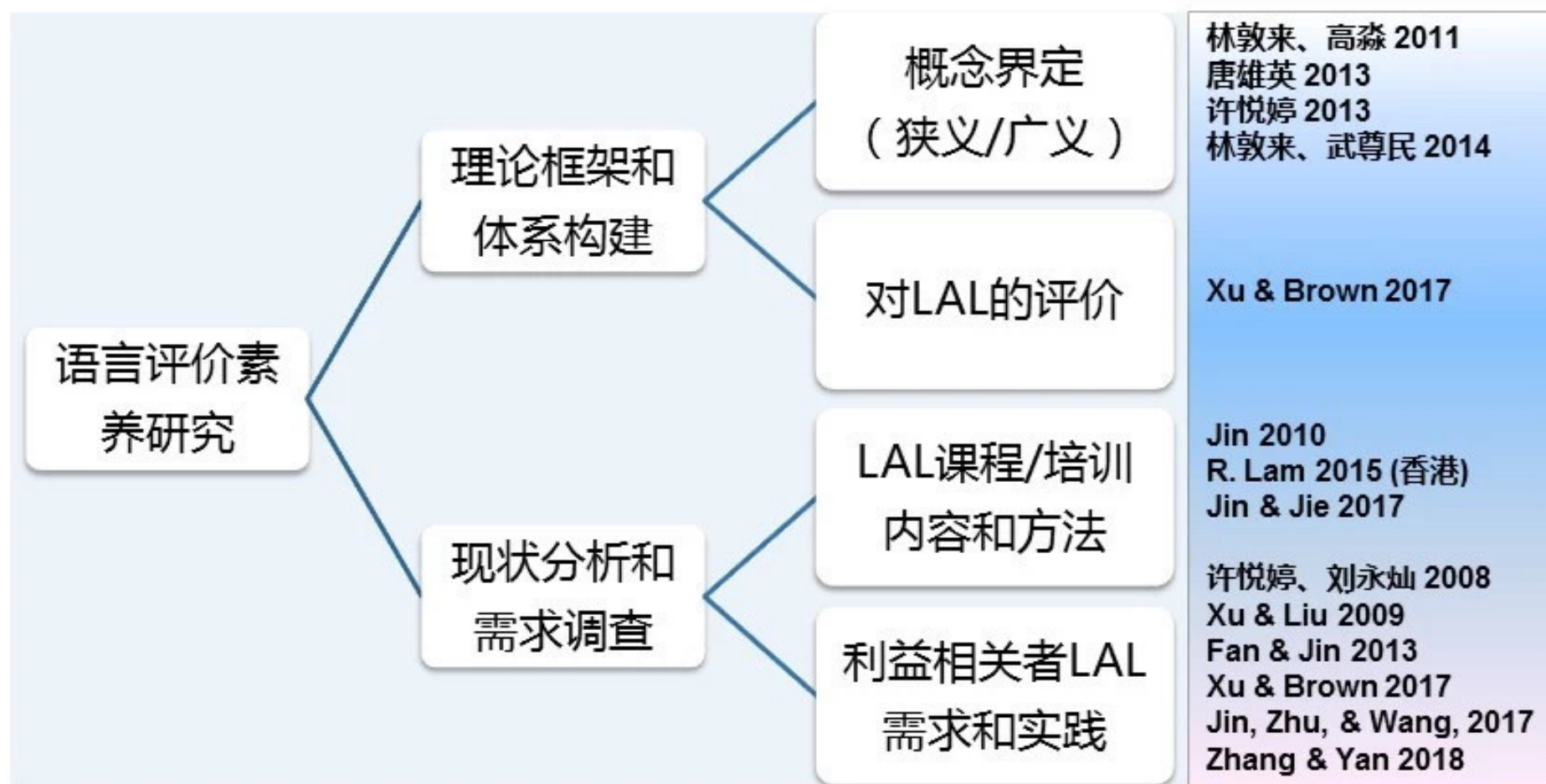
评价素养研究（案例）

LAL RESEARCH

作文自动评分(AES)中的评价素养

A need for improving AES-related assessment literacy among teachers and students in the transition from human rating to machine scoring.

Their <u>knowledge</u> about, <u>perceptions</u> and <u>evaluations</u> of AES will affect how teaching and learning are to be affected and how AES is to be accepted.

(Jin, Zhu, Wang, 2017)

12. 同一篇作文由不同的自动评分系统进行
   评分后会得到不同的评分结果。

   ○ 不同意

   ○ 同意

   ○ 不了解

全体	教师	学生
172/888=19.4%	88/376=23.4%	84/512=16.4%

Do workshops really work?
语言评价培训有效性

A quasi-experiment design to infer the effect of training:
- An experiment and a control group;
- Before and after the training.

A series of **questionnaires** to track and evaluate the progress on components of LAL provided in the program

(Jin & Jie, 2017)

Cronbach α for the survey forms (experimental and control groups)

	Number of items	Average interitem covariance	Scale reliability coefficient
E-Q1	30	.1496443	0.8304
E-Q2	10	.1382941	0.8023
E-Q3	12	.1561992	0.8914
E-Q4	25	.1879472	0.6929
E-Q5	20	.1317928	0.9023
E-Q6	13	.1042245	0.8254
E-Q7	30	.0930029	0.7076
C-Q1	30	.1354330	0.7715
C-Q7	30	.1104562	0.7295

41. 考试的公平公正性是指考试实施流程的
标准化，与试题质量无关。

○ 正确

○ 错误

	全体	正确	错误
控制组	N=85	17 (20%)	68 (80%)
实验组	N=52	3 (5.8%)	49 (94.2%)

44. 请反馈您对以下语言测试知识和技能的
需要程度。1. 完全不需要 ⟶ 5. 迫切需要

	1	2	3	4	5
A. 大规模考试	3.85	5.77	13.46	28.85	48.08
B. 课堂评价	0	3.85	25.00	28.85	42.31
C. 校本考试	0	1.92	21.15	25.00	51.92
D. 分级测试	1.92	3.85	26.92	26.92	40.38
E. 诊断测试	0	0	15.38	32.69	51.92

The place of language testing and assessment in the professional preparation of foreign language teachers in China

Yan Jin
Shanghai Jiao Tong University, P.R. China
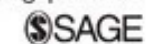
LAL课程调查

83	多变量回归分析（multiple regression analysis）	1.60
84	主成分分析（principal components analysis）	1.64
85	因子分析（factor analysis）	1.66
86	结构方程模型（structural equation modeling）	1.23
87	单参数试题响应理论（one-parameter IRT）	1.46
88	双参数试题响应理论（two-parameter IRT）	1.40
89	三参数试题响应理论（three-parameter IRT）	1.38
90	Rasch分析（Rasch）	1.51
91	多层面Rasch分析（MFRM）	1.24

语言测试课程教学内容的时间分配

1：没有
2：很少
3：一般
4：较多
5：很多

"Research into assessment literacy is in its infancy."

" 当前，国内对教师评价能力的
探讨比较零散，对教师评价能
力的理解还不够全面，能够结
合国内教学实践和教学情境进
行教师评价能力的探讨则更少

（唐雄英，2013：51）

" 在我国，教师评估素质多年来是一项研究空白，到近几年才有零星的成果。……对我国语言教师特别是外语教师的评估素质的实证研究目前仍是空白

（许悦婷，2013：46-47）

3

语言评价素养的
发展路径
Development of LAL

Standards for Teacher Competence in Educational Assessment of Students

Seven competency domains in which teachers should be skilled:

1. choosing assessment methods appropriate to instructional decisions;

2. developing assessment methods appropriate to instructional decisions;

3. administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods;

(AFT, NCME, & NEA, 1990; 参见Boyles, 2005)

Standards for Teacher Competence in Educational Assessment of Students

4. using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement;

5. developing valid pupil grading procedures;

6. communicating assessment results to various stakeholders;

7. recognizing unethical, illegal, and inappropriate assessment methods and uses of assessment information.
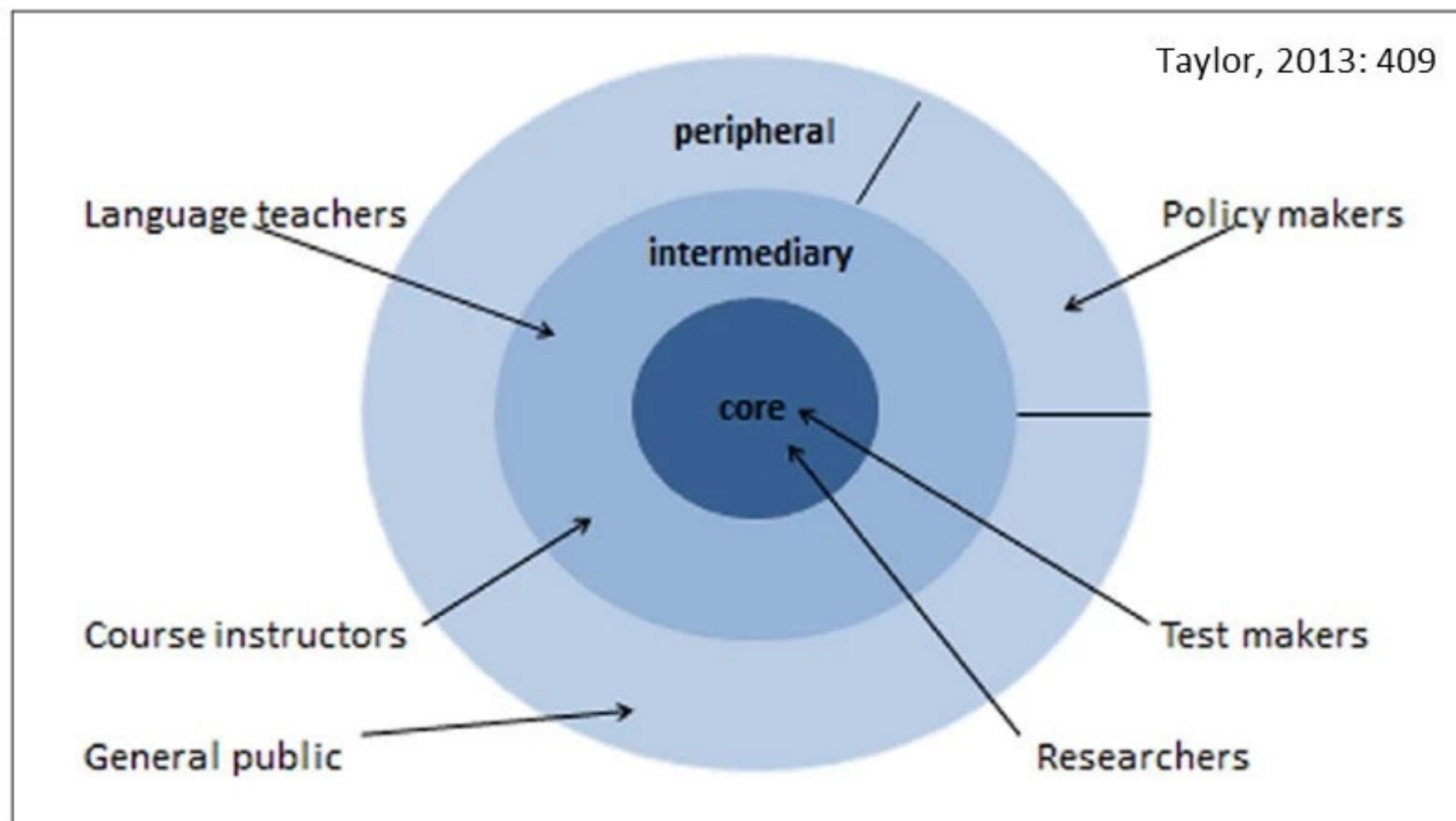
教师对学生的教育评价能力标准

1）**选择合适的评价工具**为教学决策服务；

2）**设计合适的评价工具**为教学决策服务；

3）利用外在或自己设计的评价工具**实施评价，评分并解释结果**；

4）**运用评价结果**判断学生、设计教学、开发课程以及提高学校业绩等；

5）开发有效的**评价学生手段**；

6）向学生、家长，以及其他非专业人士和教育工作者**汇报评价结果**；

7）判断在评价中**是否违背伦理、不合法或不恰**当地使用了评价手段或信息。

> **"** While Fulcher's (2012) expanded working definition of language assessment literacy helpfully draws together a broad range of elements that could be key components of the construct, **it does not address the issue of the degree or depth** that might be necessary for any of the elements depending upon the nature and extent of the stakeholder's involvement in assessment.

Taylor, 2013: 410-411

Taylor, 2013: 409

Figure 1. Levels of AL/LAL differentiated according to stakeholder constituency.

(b)

Knowledge of theory

Technical skills

Scores and decision making

Principles and concepts

Personal beliefs/attitudes

4
3
2
1
0

Local practices

Language pedagogy
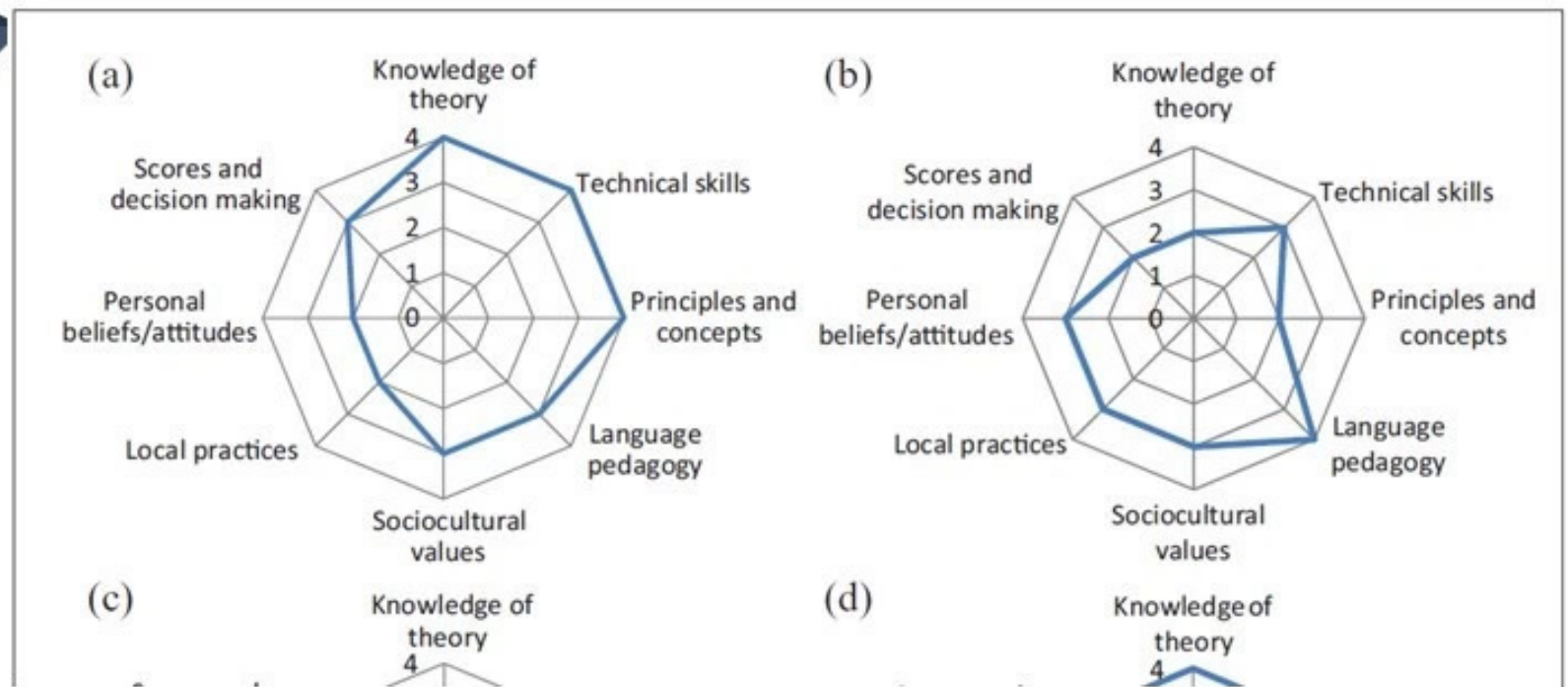
Sociocultural values

语言教师素养图

语言教学（4）
专业技能（3）
社会文化（3）
本土实践（3）
信念态度（3）
理论知识（2）
原则理念（2）
评分决策（2）
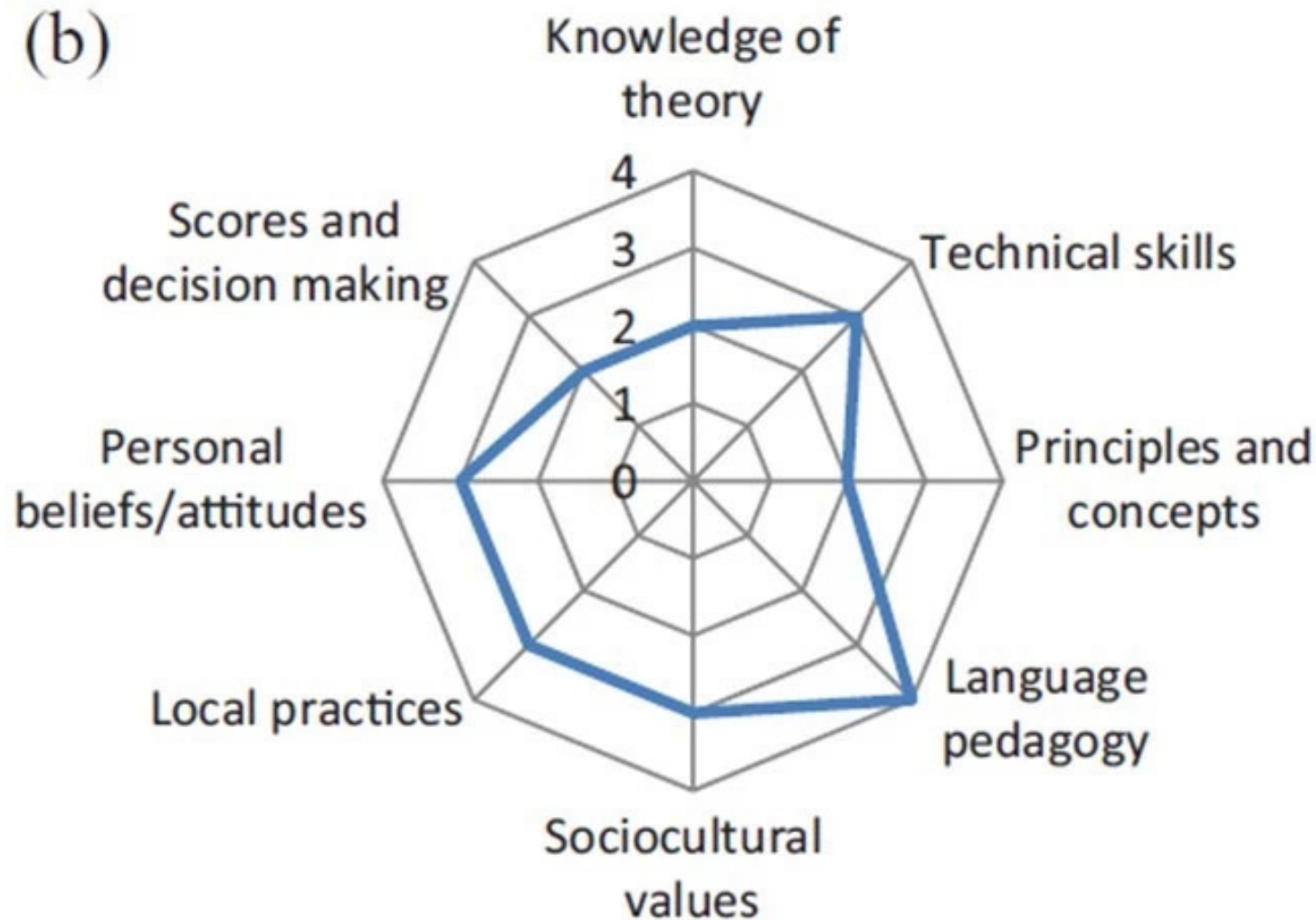
Figure 2. Differential AL/LAL profiles for four constituencies.
(a) Profile for test writers.
(b) Profile for classroom teachers.
(c) Profile for university administrators.
(d) Profile for professional language testers.

this approach helps move us away from a notion of AL/LAL as a polarized dichotomy (i.e. one is either literate or illiterate in this area) towards a continuum which identifies and describes progressive stages of literacy along the way, for example nominal, functional, procedural/conceptual and multidimensional. This characterization merits further thought and exploration to see if it might provide us with a 'literacy ladder' for assessment; that is, a way of conceptualizing AL/LAL progression which could assist in designing modular courses in language testing and even in certificating levels of achievement as part of a broader professional development programme for certain stakeholder groups if appropriate to do so.

表 2　语言评价素养连续体(改编自 Pill & Harding 2013) (林敦来、武尊民2014：717)

编码	程度	具体描述
0	素养完全缺失 (illiteracy)	完全忽视语言评价的概念和方法
1	极少的素养 (nominal literacy)	知道某个特定的术语属于评价领域，但是理解有偏差
2	功能性的素养 (functional literacy)	对基本术语和概念有良好的理解
3	程序性和概念性的素养(procedural and conceptual literacy)	理解评价领域的核心概念，并能将之运用于实践
4	多维的素养 (multidimensional literacy)	理解超出了普通概念，而包含评价中的哲学的、历史的和社会的维度内容。

对语言评价素养发展的启示

- 并非"非有即无"的概念，而是循序渐进的连续体；
- 可以分层次、分等级，形成一个有坡度的LAL阶梯；
- 是职业发展的一部分，可分模块培训，甚至可发等级证书。

语言评价素养培训内容和方式

☐ 培训要平衡"technical know-how, practical skills, theoretical knowledge and understanding of principles" (Taylor 2009)

☐ 培训可以通过工作坊、在线学习、自学等进行；

☐ 工作坊尽管费时、昂贵，仍是最常用的方法，是课程的有效补充 (Malone 2008)

☐ 通过实践发展语言评价素养，做中学，研究中提高

☐ 加强语言评价素养的评价工具研究，探索培训有效性

校本考试的设计和实施

ESP测试的理论与实践

大规模语言测试的改革

命题的原则、方法和实践

课堂教学中的诊断测试

2016
SJTU
语言测试与评估暑期培训班

考试数据的统计分析

评分标准的设计和论证

内容：

校本考试

课堂评价

大规模考试

理论与实践结合

方法：
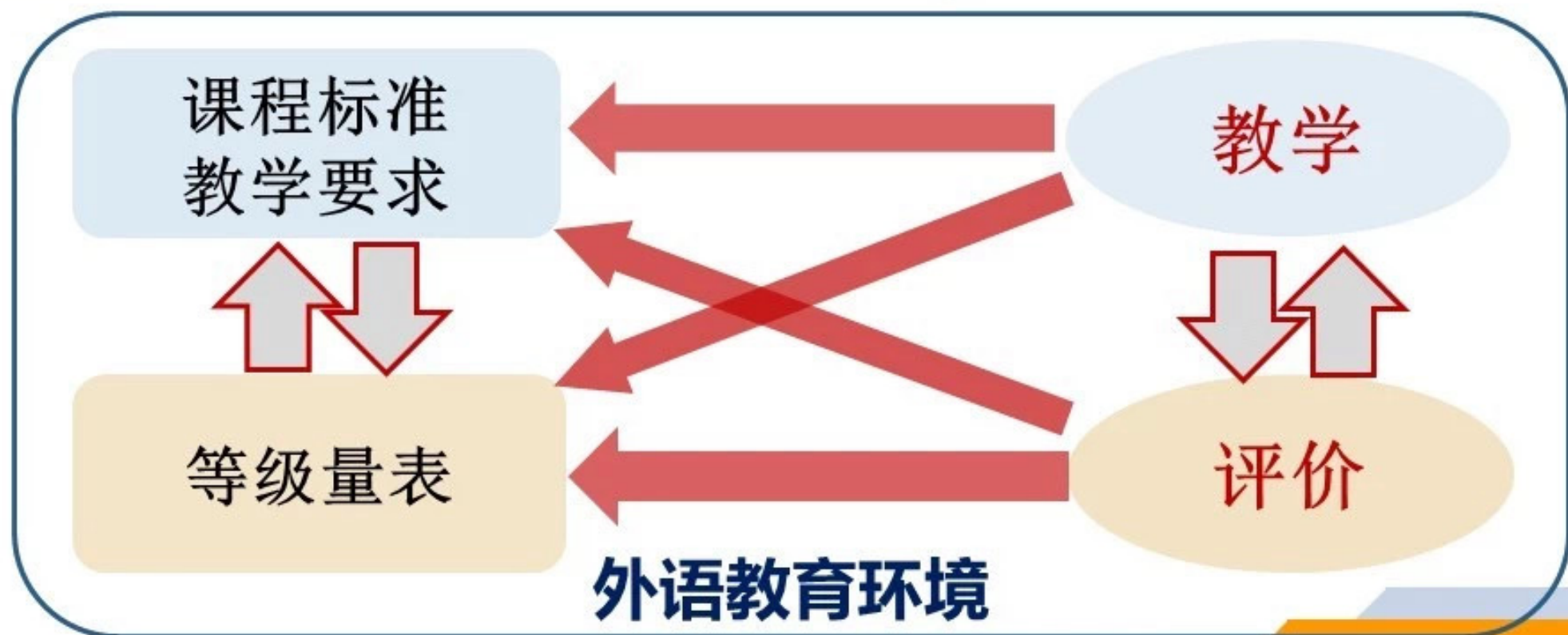
教师与学员互动

学员与学员讨论

中国外语教育特点

- 逐步从"考试文化"向"评价文化"转变
- 评价既是外语教育改革的内容，也是改革的重要推动力
- 评价研究和评价素养发展应紧密结合外语教育改革的目标
- 以中国英语能力等级量表研究为例

量表、课程与教学和评价

课程标准
教学要求

教学

等级量表

评价

外语教育环境

理解、使用和研究量表

理解量表	使用量表	研究量表
☐ 理论基础	☐ 教学	☐ 量表开发
☐ 框架结构	☐ 学习	☐ 对接研究
☐ 等级设计	☐ 评价	☐ 效度论证
☐ 描述语		☐ 使用后效

"我国的评价素养研究和发展不仅要与国际接轨，更要突出我国教育环境特点，要有自己的特色。我们的教学和评价环境、规模、管理体系等与国外都不同，我们的教学和评价方法、教师评价素质与国外也有差异，在评价素养研究中须充分考虑环境因素的影响。

Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (ed.). National Assessment Summit Papers. Ames, IA: Iowa State University, 11-15.

Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (eds). Experimenting with Uncertainty: Essays in Honour of Alan Davies . Cambridge: Cambridge University Press, 126-136.

Chalhoub-Deville, M. (2106). Validity theory: Reform policies, accountability testing, and consequences. Language Testing, 33(4) 453–472.

Davies, A. (2008). Textbook trends in teaching language testing. Language Testing 25 (3): 327-347.

Fulcher, G. (2012). Assessment literacy for the language classroom. Language Assessment Quarterly 9(2): 113-132.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. Language Testing 25(3): 385-402.

Inbar-Lourie, O. (2013a). Language assessment literacy. In C. A. Chapelle (ed.). The Encyclopedia of Applied Linguistics. Wiley-Blackwell, 2923-2931 .

Inbar-Lourie, O. (2013b). Guest editorial to the special issue on language assessment literacy. Language Testing 30(3): 301-307.

McNamara, T. & C. Roever. (2006). Language Testing: The Social Dimension. Oxford, England: Blackwell.

Popham, W J. (2009). Assessment literacy for teachers: Faddish or fundamental? Theory into Practice, 48(1): 4-11.

Rea-Dickens, P. (2008). Classroom-based language assessment. In E. Shohamy & N. H. Hornberger (eds.). Encyclopedia of language and education. Vol. 7: Language testing and assessment, 2nd ed., 257–271. New York, NY: Springer Science + Business.

Spolsky B. (2008). Introduction: Language Testing at 25: maturity and responsibility? Language Testing, 25(3): 297-305.

Taylor L. (2009). Developing assessment literacy. Annual Review of Applied Linguistics, 29: 21-36.

THANKS!

Any questions?

yjin@sjtu.edu.cn