



清华大学

Department of Foreign Languages, Tsinghua University

# 试探基于写作任务的多元评价 和反馈智能系统的构建

清华大学外文系 张文霞

2018年3月





针对目前外语写作教学对现有写作自动评阅系统特别是内容评判方面的需求，本论坛交流发言以“促学评价”理论为基础，尝试提出一个基于具体写作任务的多元评价和反馈智能系统的构建方案，强调对文章内容的反馈和修改，加强写作课堂教学对意义表达和思维训练的重视。



# 提纲

- 一、背景
- 二、理论：以评促学
- 三、问题：自动和人工评阅及反馈
- 四、系统：架构、功能和特点
- 五、总结讨论



Tsinghua University

# 一、背景





# 《大学英语教学指南》（2016试）

## 评价与测试

- 大学生英语能力测试应包括形成性测试与终结性测试，应加强形成性反馈，处理好共同基础测试与校内测试、综合语言能力测试与单项语言技能测试、基础英语测试与专门用途英语测试等各方面的关系，实现“对学习结果的终结性测试”与“促进学生学习的形成性测试”的有机结合。
- 学校大学英语教学管理部门应根据本校的实际情况，对学生提出切合实际的英语能力要求，开发科学、系统、个性化的大学英语课程评价体系和大学生英语能力测试体系，充分发挥评价和测试对大学英语教学的导向、激励、诊断、改进、鉴定、咨询、决策等多重功能，更好地促进大学英语课程的建设 and 大学生英语能力的提高。



Tsinghua University

# 大学英语测试

- 特点：综合体系（科学、系统、个性）
- 形式：形成性+终结性
- 核心：校本、人才培养

## 问题思考：

- 课程测试中如何落实？（测试与教学的有机结合、测试手段/技术的掌握与运用）



# 大学外语教育

- 培养目标：1) 综合应用能力；2) 跨文化语言交流能力；3) 思辨能力
- 指导思想：分类指导、因材施教
- 教学要求：基础、提高、发展

## 问题思考：

- 课程测试中如何落实？（课堂/教学测试如何体现教学/培养目标？）



Tsinghua University

## 二、评价理论





# 促学评价

- 三位一体：AfL, AoL, AsL (Crusan, 2010; Earl, 2003; Lee, 2010; William, 2009), LOLA (Hamp-Lyons, 2015)
- 建构主义；基于过程；反思；多元；共享；人机结合（智能？）



Tsinghua University

# CBA Design

- Where to go?

Specify your goals:

- Where to start?

What Ss know:

- How to get there?

Possible methods/tools:





Tsinghua University

# 基于课堂的评价研究模型

(Hill and McNamara, 2011)

- Definition of CBA (Hill & McNamara, 2011, p 396)
- Any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes.

**Table 7. A framework for research on CBA processes**

**1. What do language teachers do?**

1.1. Planning Assessment	Is there planning for assessment? How detailed is planning? What is its intended relationship to instruction? How does it relate to external standards and frameworks?
1.2. Framing Assessment	Is assessment made explicit to learners? How is this done?
1.3. Conducting Assessment	What opportunities does the classroom provide for assessment? Does assessment tend to focus on the class, group/pairs of students or individuals?
1.4. Using Assessment Data	How is assessment-related information used? Teaching Learning (feedback) Person-referenced Task-referenced <ul style="list-style-type: none"><li>• Confirmatory</li><li>• Explanatory</li><li>• Corrective</li></ul> Reporting Management Socialization

**2. What do teachers look for?**

*What information about valued enterprises, qualities and standards is available?*

2.1. In Advance	in written/verbal instructions and/or assessment rubrics?
2.2. In Feedback	in written and/or verbal feedback?
2.3. In Reporting	in reporting deliberations and/or in written reports?

**3. What theory or 'standards' do they use?**

3.1 Teacher Theories & Beliefs	What does the data reveal about teachers' beliefs about <ul style="list-style-type: none"><li>• the subject or content area</li><li>• second language learning and teaching, and</li><li>• the nature of assessment?</li></ul>
--------------------------------	--

**4. Do learners share the same understandings?**

4.1. Learner Theories & Beliefs	What does the data reveal about learners' beliefs about <ul style="list-style-type: none"><li>• second language learning, and</li><li>• the nature of assessment?</li></ul>
---------------------------------	---



Tsinghua University

## 四、问题





Tsinghua University

# 自动评阅 (AWE)

- 评价：教学环境vs测试环境
- 反馈：语言形式vs内容表达





Tsinghua University

# 人工（教师、学生）评价反馈

- 评价：1) 标准：测试标准**vs**教学（任务）标准；2) 效率：数量**vs**质量
- 反馈：时效、内容、接收度





Tsinghua University

## 四、系统：构架、功能和特点



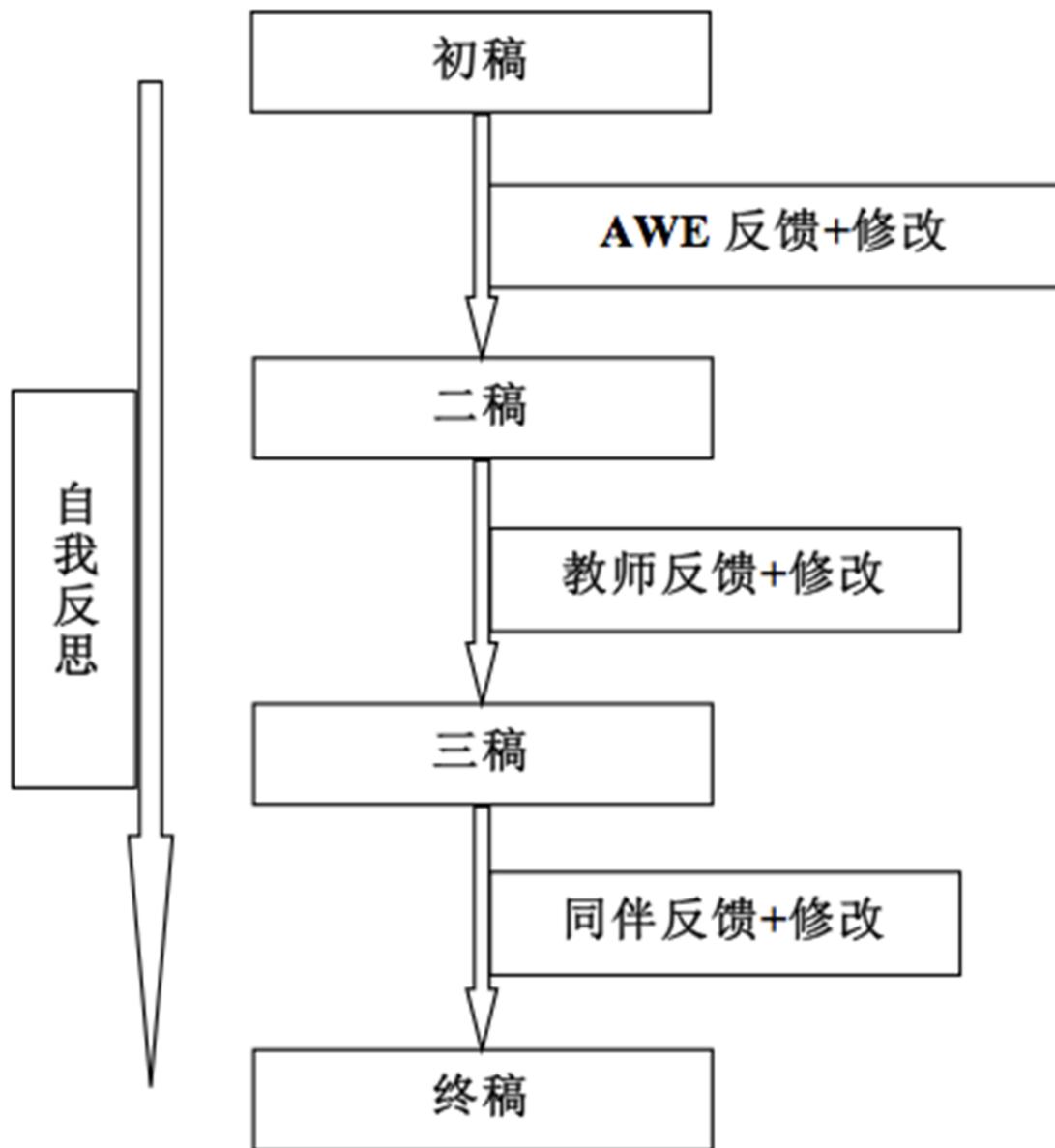




表2 “快乐英语写作” 过程化施教评价反馈方案

时间		课程安排		标准	教研团队	学生
单周	课中	课堂讲解		课程和具体写作任务评价标准	任务描述, 标准传达	了解任务, 明确标准
	课后	形成性反馈评价	机器反馈	机器内置标准	查看提交、反馈使用情况	研读反馈, 反思学习
			教师反馈	课程和具体写作任务评价标准	实施“评价标准”反馈	研读反馈, 了解标准 修改文本, 反思学习
双周	课中	课堂评议		课程和具体写作任务评价标准	教师总结, 互评展示 小组互评培训, 范文分析	深化标准, 反思学习
	课后	形成性反馈评价	同学互评	课程和具体写作任务评价标准	检查互评情况, 矫正教学	实施标准, 反思学习
			反思自评	课程和具体写作任务评价标准	了解学生学习, 矫正教学	内化标准, 反思学习
		终结性反馈评价	课程和具体写作任务评价标准	过程、终稿质量结合打分	接受反馈	



# 多元评价和反馈模型

	阶段	主体	互动形式	反馈形式	维度	权重 (%)	
形	机评	机器软件	机器	书面	语法层面	10	
成	教师评	教师/助教	师生	书面\口头	多层面	50	
性	互评	学生	生生	书面\口头	多层面	20	
评	总评	教师学生	全班 小组 个人	书面\口头	多层面		10
价	自评	学生自己	反思	书面	多层面	20	
终结性评价	评分						



表 1 基于写作任务的作文自动评阅系统

组成部分	内容		标准	实施者
反馈	分数	针对作文的语言 (总分和分项分)	系统内置标准	AWE 系统
		针对作文的内容 (总分和分项分)	具体写作任务评价标准	人工 (老师/同伴/作者)
	文字反馈	针对作文的语言	系统内置标准	AWE 系统/人工 (同伴/作者)
		针对作文的内容	具体写作任务评价标准	人工 (老师/同伴/作者)
	口头反馈	针对作文的内容	具体写作任务评价标准	人工 (老师/同伴/作者)
诊断报告	针对作文的语言		系统内置标准	AWE 系统
	针对作文的内容		具体写作任务评价标准	AWE 系统
学习/教 学建议	针对作文的语言		系统内置标准	AWE 系统
	针对作文的内容		具体写作任务评价标准	AWE 系统

## Wisdom Evaluator (WE)

一起学习，智慧写作，共享评阅

[登录](#) [注册](#)

邮箱

密码

记住密码

[登录](#) [忘记密码](#)

[下载 WE APP](#)



Tsinghua University

# 主要特点

- 针对性：直击痛点、针对性强
- 科学性：任务导向、标准明确
- 开放性：教学相长、共享评阅
- 高效性：人机结合、省时高效





Tsinghua University

# 总结讨论

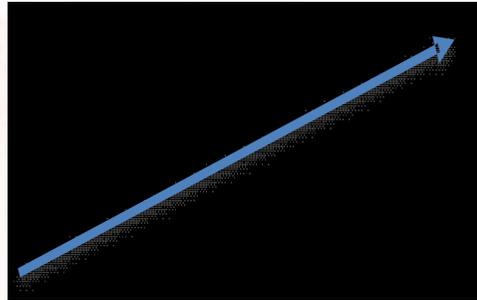
- 教师：理念、模式、方法、导引
- 学生：共识、动机、培训、策略
- 系统（平台）：环境、硬件、学习、智能



Tsinghua University

# 基于标准的‘以评促学’写作教学

Where I am going



Where I am

“以评促学”示意图



# Selected References

- Black, P., & William, D. (1998). Inside the Black Box: Raising Standards through Classroom Assessment, *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Hill, K. and McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3) 395–420.
- Hyland, K. & Hyland, F. (2006). Feedback on second language students' writing. *Lang. Teach.* 39, 83–101.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 145-165.
- 《大学英语课程教学要求》（2007）
- 《大学英语教学指南》（2016试）
- 黄静、张文霞（2014）.多元反馈对大学生英语作文修改的影响，《中国外语》11,1,51-56
- 武永，张文霞. (2016). 作文自动评价系统和教师反馈对大学生英语作文修改的影响研究.《中国外语教育》，9(1), 12-19
- 杨永林（2016）. 体验英语写作资源（TRP 3.0版），高教社
- 张文霞，武永. (2017). 以评促学：构建基于写作任务的作文自动评阅系统.《现代教育技术》，3,99-104.



清華大學

Department of Foreign Languages, Tsinghua University

Thank

You

