

# 基于 ICAO 框架适配的乘务英语口语诊断性评价标准构建

平柳琼 刘珏

上海民航职业技术学院

DOI:10.32629/mef.v8i19.17381

**[摘要]** 乘务英语口语评价面临标准缺失的困境,现有标准抽象模糊,缺乏支撑智能系统诊断功能的细化规则。民航英语领域虽有成熟的ICAO语言能力评估框架,但该框架针对飞行员-管制员通信场景设计,与乘务员-乘客服务场景存在本质差异,无法直接套用。本研究采用“框架适配”策略,通过“场景分析-框架适配-诊断细化-规则设计”四步法,在保留ICAO框架核心维度的基础上重新定义各维度内涵,新增“服务得体性”维度,构建了六维诊断性标准体系。研究将抽象能力描述转化为可量化规则和可编程的算法,提出的“框架适配”策略为ESP领域借鉴成熟框架构建场景化数字评价标准提供了可复制的方法论范式。

**[关键词]** ICAO框架适配; 乘务英语; 诊断性评价标准; 人机协同

**中图分类号:** G633.41 **文献标识码:** A

## Constructing Diagnostic Assessment Standards for Cabin Crew English Speaking Proficiency: An ICAO Framework Adaptation Approach

Liuqiong Ping Yu Liu

Shanghai Civil Aviation College

**[Abstract]** The assessment of cabin crew oral English proficiency faces a dilemma of lacking standards—existing criteria remain abstract and vague, without the refined rules necessary to support the diagnostic functions of intelligent systems. Although the civil aviation English field has a mature ICAO Language Proficiency Rating Scale, this framework was designed for pilot-controller communication scenarios, which differ fundamentally from cabin crew-passenger service contexts and cannot be directly applied. This study adopts a "framework adaptation" strategy, employing a four-step approach of "scenario analysis—framework adaptation—diagnostic refinement—rule design." While preserving the core dimensions of the ICAO framework, this approach redefines the connotations of each dimension and introduces a new "service appropriateness" dimension, thereby constructing a six-dimensional diagnostic criteria system. The research transforms abstract competency descriptions into quantifiable rules and programmable algorithms. The proposed "framework adaptation" strategy provides a replicable methodological paradigm for the ESP field to leverage established frameworks in constructing scenario-specific digital assessment standards.

**[Key words]** ICAO framework adaptation; cabin crew English; diagnostic assessment standards; human-machine collaboration

### 1 引言：问题的提出

民航英语领域存在一个值得关注的现象：针对飞行员的语言能力评价已有成熟的国际标准——国际民航组织（ICAO）语言能力评估框架，但针对乘务员的英语口语评价却长期缺乏统一的评价标准。这一“标准落差”造成了双重困境：一方面，现行乘务英语评价标准普遍抽象模糊，难以支撑智慧评价系统的诊断功能；另一方面，ICAO框架虽然成熟，但其针对飞行员-管制员通信场景的设计定位，与乘务员-乘客服务场景在交际目标、语

言功能上存在本质差异，直接套用会导致评价失效。

如何破解这一困境？本研究认为，关键在于探索如何将成熟框架适配到新场景。ICAO框架历经二十年发展，其六维度结构（发音、结构、词汇、流利度、理解、互动）基本覆盖了口语交际能力的核心要素，其“维度独立、层级清晰、行为导向”的设计理念具有普适价值。

本研究聚焦以下问题：（1）乘务员-乘客服务场景与飞行员-管制员通信场景存在哪些本质差异？（2）如何基于这些差异对

ICAO框架进行适配？(3)如何将适配后的框架转化为支撑AI诊断的细化规则？以乘务英语为例，探索“框架适配”的系统方法，力图为ESP评价标准开发提供可复制的路径。

## 2 理论基础与框架选择

2.1 框架适配的理论依据。ESP评价理论为框架适配提供了理论依据。Douglas (2000) 强调职业语言评价的“真实性”与“特定性”原则，指出语言能力评价必须根植于目标使用场景<sup>[1]</sup>。Elder & McNamara (2016) 在物理治疗师工作场所沟通评估中提出“本土化标准” (indigenous criteria) 概念，指出职业特定的语言能力无法用通用标准有效评估<sup>[2]</sup>。这些研究揭示了一个核心命题：评价标准必须与目标场景相匹配。

然而，“场景匹配”并不意味着每个职业场景都需要从零构建标准。Knoch (2015) 研究护理英语时发现，职业特定能力虽然无法用通用标准直接评估，但可以在通用语言能力维度基础上增加职业特定维度<sup>[3]</sup>。这一思路印证了框架适配的可行性。

2.2 ICAO框架的结构与价值。国际民航组织 (ICAO) 语言能力评估框架是民航英语领域最具权威性的评价标准。2003年，针对因语言沟通问题导致的航空事故，ICAO发布了语言能力要求，要求所有参与国际飞行的飞行员和管制员必须达到ICAO四级 (运行级) 以上水平。2010年发布的《ICAO语言能力要求实施手册》(Doc 9835) 确立了六维度评估框架，包括发音、结构、词汇、流利度、理解和互动六个维度<sup>[4]</sup>。ICAO框架的学术价值在于：维度结构的完整性、等级描述的行为导向、已被全球190多个成员国采用。这些特征使其成为框架适配的理想基础。

2.3 框架适配的必要性。尽管ICAO框架具有重要借鉴价值，但该框架是针对飞行员-管制员通信场景设计的，与乘务员-乘客服务场景存在本质差异。两类场景的差异体现在多个层面：交际对象上，飞行员面对专业人员 (管制员)，乘务员面对普通公众 (乘客)；核心目标上，前者强调安全指令传达，后者兼顾服务提供与安全保障；语言风格上，前者追求简洁、精准、标准化，后者要求礼貌、得体、灵活应变。这些差异决定了ICAO框架不能简单套用于乘务英语评价，需要在维度重心、评价标准、新维度增设等方面进行针对性调整。

## 3 研究设计

本研究采用“场景分析—框架适配—诊断细化—规则设计”四步法，将成熟评价框架转化为支撑智能评价的诊断性标准。

3.1 数据收集。场景分析数据采用三角验证法收集。文献分析涵盖ICAO Doc 9835、《民航乘务员国家职业技能标准》、6部航空英语教材及3家航空公司乘务培训手册，提取语言能力要求描述并编码归类。课堂观察于2025年2月至6月在某民航职业技术学院进行，累计48课时，采用结构化观察记录表记录典型交际案例217个，归纳任务类型、语言功能和常见错误。专家访谈对象为6位一线教师 (平均教龄8.6年)，采用半结构化访谈提纲，围绕“评价痛点”、“能力要求”、“常见问题”三个主题展开，每人约45分钟，访谈录音转写后进行主题分析。

专家咨询邀请12位专家参与两轮德尔菲验证，专家构成兼顾学术研究与行业实践：乘务英语教师6名，航空公司乘务培训师4位，应用语言测评专家2位。第一轮采用开放式问卷，收集对维度设置、特征描述、权重分配的意见；第二轮采用结构化量表，量化评估各维度的必要性和AI实现的可行性。

3.2 效度验证方法。本研究运用三种方法对标准的科学性予以验证。内容效度比 (Content Validity Ratio, CVR) 用于计算每个维度被专家判定为“必要”的比例，依据Lawshe (1975) 临界值表，当 $n=18$ 时， $CVR>0.42$ 判定为合格。Kendall's W系数用于评估专家对维度权重排序的一致性， $W>0.7$ 表明具有高度一致性。清晰度评估采用5点量表 (1=极度不清晰，5=极度清晰)，分别对诊断特征、反馈示例以及AI实现路径的清晰程度进行评价。

## 4 基于ICAO框架适配的六维标准体系

4.1 适配过程与结果。基于场景差异分析，本研究对ICAO六维度框架进行了系统适配。适配过程遵循“保留核心结构、重新定义内涵、增设特色维度”的原则，逐维度进行分析和调整。表1呈现了适配过程及调整说明。

表1 ICAO框架适配过程与调整说明

ICAO原维度	适配决策	适配后维度	调整理由
Pronunciation	保留并调整内涵	发音清晰度	从“标准口音”转向“可懂度”导向，容忍不影响理解的口音特征
Vocabulary	保留并调整内涵	词汇适切性	词汇范围从航空术语转向服务用语，增加礼貌标记词评价
Structure	保留并降低权重	语法准确性	服务场景对语法容错度较高，强调“可理解性优先于正确性”
Fluency	保留并细分	表达流畅度	区分“标准话术流畅度”与“灵活应对流畅度”两类要求
Comprehension	保留并扩展	理解与回应	扩展为“理解+回应”，增加隐含需求识别
Interactions	重新定义	服务得体性	从“通话程序遵守”转向“服务态度与应对策略”，体现服务性语言特征

最显著的适配体现在“互动”维度。ICAO框架中的“互动” (Interactions) 强调标准通话程序遵守、话轮转换规范和澄清确认策略，这些要素对管制通信至关重要，但与乘务服务场景的核心需求存在根本错位。乘务场景的“互动”本质是服务提供过程中的人际交往，其核心能力是礼貌表达和得体应对。因此，本研究将该维度重新命名为“服务得体性”。

4.2 六维标准框架总览。发音清晰度权重25%，核心诊断内容包括关键词清晰度、错误音素及重音语调等，其AI可行性得分高达4.8。这充分体现了服务场景对语言可理解性的核心要求。词汇适切性和表达流畅度的重要性次之，权重均为20%，AI可行性得分分别为4.5和4.0。语法准确性权重较低 (15%)，主要考察基础时态与句子完整性，体现了“意义传达优先于形式正确”的原则。理解与回应维度权重为12%，AI可行性为3.6。值得强调的是，尽管服务得体性权重为8% (AI可行性3.2)，但作为独立维度的设立，其本身即体现了对礼貌热度和应对策略等服务性语

言特征的重视,弥补了国际民用航空组织(ICAO)标准框架在服务性语言评估方面的缺失。

4.3从维度到可操作规则制定。框架适配完成后,需要将各维度转化为支撑精准诊断的细化规则。本节以发音清晰度和服务得体性为例,阐述诊断细化的设计逻辑。选择这两个维度的原因在于:发音清晰度是AI可行性最高的维度(4.8分),代表可高度自动化的评价内容;服务得体性是AI可行性最低的维度(3.2分),代表需要人机协同的评价内容。两者的对照有助于揭示诊断细化的完整逻辑。

发音清晰度分解为三个诊断子项:关键服务词汇清晰度(10分)针对safety、emergency等30个高频词建立诊断基准,通过语音识别API的音素级分析实现自动诊断;常见错误音素(10分)基于对比分析假说识别中国学习者的典型偏误,如th→s、v→w等;重音与语调(5分)诊断词重音错位和句重音缺失。

服务得体性是对ICAO“互动”维度的重新定义,涉及礼貌热情度和应对策略两个子项。该维度的AI实现可行性较低(3.2分),原因在于其评价对象涉及情感理解和价值判断,难以通过规则化算法完全捕捉。因此,本研究建议该维度采用“人工评判为主、AI辅助为辅”的模式。

4.4人机协同模式设计。基于六个维度的AI实现可行性评估结果,本研究提出分层人机协同评价模式。AI可行性评分在4.4-4.8区间的维度(发音清晰度、词汇适切性、语法准确性)采用“AI自动评判”模式;评分在3.6-4.0区间的维度(表达流畅度、理解与回应)采用“AI评分+人工抽检”模式;评分低于3.5的维度(服务得体性)采用“人工评判+AI辅助”模式。

## 5 效度验证结果

5.1内容效度与专家共识。12位专家对六个维度的必要性及诊断细则质量进行评估。结果显示,所有维度CVR均显著超过临界值0.42( $p<0.001$ ),其中“发音清晰度”和“服务得体性”获全体专家认可(CVR=1.00),反映出专家对基础语音能力与行业特色维度的高度共识。Kendall's W系数为0.78( $\chi^2=70.2$ ,  $df=5$ ,  $p<0.001$ ),表明专家在权重排序上具有良好一致性。

5.2诊断细则评估与专家反馈。诊断细则的清晰度评估显示,六个维度的平均得分均在4.0以上,综合评价为“清晰”。AI实现可行性评估呈现明显梯度,为人机协同模式设计提供了实证依据。

专家定性反馈肯定了框架适配策略的价值。在适配过程方面,专家认为“将ICAO框架适配到乘务场景的思路具有方法论意义,可推广至其他ESP领域”;在维度设计方面,专家评价“服务得体性作为独立维度的设立填补了现有标准的空白”;在诊断细化方面,专家认为“诊断特征具体、错误类型清晰,可直接转化为AI算法规则”。

## 6 讨论与结论

6.1理论贡献与实践启示。本研究的理论贡献体现在两个层面。其一,提出并验证了“框架适配”策略,为ESP评价标准开发提供了可复制的方法论,即以成熟框架为基础,通过场景差异分析进行针对性调整。其二,揭示了智慧评价效能提升的根本路径在于评价范式从“总结性打分”向“诊断性评价”的转变。

本研究对六个维度AI实现可行性的实证评估(从4.8到3.2的梯度分布),为人机协同评价模式提供了具体依据。这一发现超越了关于AI评价的两种极端立场,倡导审慎的技术现实主义:承认AI在可编程化维度的效率优势,也正视其在复杂情境判断上的局限,通过科学的人机分工实现优势互补。

6.2局限与展望。本研究的局限在于:场景分析基于单一院校数据,样本代表性有待扩展;完成了内容效度验证,但尚未进行大规模实证检验标准的预测效度。

未来研究应在三个方向深化:(1)技术验证,开发原型系统验证AI评分一致性;(2)效度检验,开展教学干预实验验证诊断反馈效果;(3)方法论推广,将“框架适配”策略应用于其他ESP领域,检验其普适性。

综上,本研究以乘务英语为例,探索了通过“框架适配”策略构建诊断性评价标准的系统方法。通过“四步法”将ICAO框架转化为支撑精准诊断的六维标准体系,实现了从成熟框架到场景化标准的有效转化。

### [基金项目]

上海市高等教育学会2024年度规划课题“融入课程思政的混合式教学设计与实践研究——以《乘务英语听说》课程为例”项目编号:(2QYB24211)。

### [参考文献]

- [1]Douglas D.Assessing languages for specific purposes [M].Cambridge: Cambridge University Press,2000.
- [2]Elder C,McNamara T.The hunt for "indigenous criteria" in assessing communication in the physiotherapy workplace [J].Language Testing,2016,33(2):153-174.
- [3]Knoch U,Macqueen S,O'Hagan S,et al.Assessing English for professional purposes:Does occupation-specific language matter?[J].TESOL Quarterly,2015,49(2):369-381.
- [4]International Civil Aviation Organization.Manual on the Implementation of ICAO Language Proficiency Requirements(Doc 9835)[S].2nd ed. Montreal:ICAO,2010.

### 作者简介:

平柳琼(1981--),女,汉族,江苏昆山人,职称:副教授,研究方向:教育技术。

刘珏(1982--),女,汉族,湖南株洲人,职称:副教授,研究方向:民航乘务英语。