

ICS 01.040.35

CSS L70/84

CIIA

团体标准

T/CIIA xxx—xxxx

英语口语能力等级计算机测评规范

Standard of Spoken English Test on Computer

(征求意见稿)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中国信息协会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	3
5 英语口语能力等级划分	3
5.1 英语口语能力框架	3
5.2 英语口语能力等级计算机测评 (SETOC) 阶段划分	4
5.3 英语口语能力等级计算机测评 (SETOC) 评价框架	4
5.4 英语口语能力等级计算机测评 (SETOC) 分项能力评价	6
6 英语口语能力等级计算机测评 (SETOC) 测试题型	10
6.1 SETOC 考查语音输出能力	10
6.2 SETOC 考查语境交流能力	11
6.3 SETOC 考查口头表达能力	11
7 英语口语能力等级计算机测评 (SETOC) 测评维度	12
7.1 SETOC 语音输出能力的计算机测评维度	12
7.2 SETOC 语境交流能力的计算机测评维度	12
7.3 SETOC 口头表达能力的计算机测评维度	12
8 英语口语能力等级计算机测评 (SETOC) 系统	12
8.1 SETOC 系统定义	12
8.2 SETOC 系统架构	12
8.3 SETOC 考务管理系统	13
8.4 SETOC 命题制卷系统	13
8.5 SETOC 阅卷评分系统	13
8.6 SETOC 考试系统	13
8.7 SETOC 系统运行环境	14
8.8 SETOC 系统性能	15
9 英语口语能力等级计算机测评 (SETOC) 考务	16
9.1 SETOC 考试准备	16
9.2 SETOC 正式考试	16
9.3 SETOC 考后评分	16
9.4 SETOC 成绩发布及复查	16
9.5 SETOC 区块链评测成绩管理系统	17
10 英语口语能力等级计算机测评 (SETOC) 报告	18
10.1 SETOC 学生测评报告	18
10.2 SETOC 班级测评报告	19
10.3 SETOC 学校测评报告	19
10.4 SETOC 地区测评报告	20
11 英语口语能力等级计算机测评 (SETOC) 题库建设	20
11.1 SETOC 通用题库建设	20
11.2 SETOC 自适应题库建设	25

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由中国信息协会提出并归口。

本文件起草单位：苏州驰声信息科技有限公司、驰声教育考试研究院、南昌职业大学、北京市教育学会学科英语分会、北京乐恩宁教育测评技术中心、东北师范大学外语学院、通化师范学院外语学院。

本文件主要起草人：林远东、丁俊华、张继伟、高淑君、薛军、甄文超、刘盼宁、刘丽艳、王少杰、张海龙。

英语口语能力等级计算机测评规范

1 范围

本文件规定了英语口语能力等级考试实施过程中应参照的相关维度、级别体系以及规范标准。

本文件适用于大、中、小学英语口语能力的计算机测评，可供大、中、小学英语口语教学、学习及其它参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对于的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GA/T 1048.1-2013 标准汉译英要求

GB/T 6865-2009 语种熟练程度和外语考试等级代码

GB/T 15528-2013 语音通信用传声器和耳机测量方法

GB/T 25000.62-2014 软件工程 软件产品质量要求与评价(SQuaRE) 易用性测试报告行业通用格式(CIF)

GB/T 29836.3-2013 系统与软件易用性 第3部分：测评方法

GB/T 36350-2018 信息技术 学习、教育和培训 数字化学习资源语义描述

GB/T 36352-2018 信息技术 学习、教育和培训 教育云服务：框架

3 术语和定义

下列术语和定义适用于本文件。

3.1

英语口语能力等级 Level of Spoken English

基于英语口语计算机测评，对不同层次英语学习者口语能力进行区分和说明的评价体系；以《中国英语能力等级量表》为依据，参照《义务教育英语课程标准（2011版）》、《普通高中英语课程标准（2017版）》、《高职高专教育英语课程教学基本要求》、《大学英语课程教学要求》制定。

3.2

人机对话 Man-machine Interaction

人机对话是借助计算机及网络技术对考试进行实施、管理的一种测试形式；它可以根据考试设计的需求，有针对性地进行命题、组卷，并完成试题呈现、接受答案、计分、数据分析以及结果解释等一系列环节。在考试操作环节，考生可以读/听计算机屏幕上试题，通过鼠标手动选择或通过麦克风录音朗读正确答案的形式完成测试。

3.3

英语口语计算机测评 Assessment of Spoken English on Computer

基于人工智能语音测评技术和大数据语料，采用在线测试手段，通过全程“人机对话”的形式对学生的英语口语能力进行测评。

3.4

英语口语计算机测评维度 Assessment Dimension of Spoken English on Computer

指用于对考生英语口语能力进行描述和评价的相关“属性、范围、系数”。在本文件范围内，测评维度主要指：语音朗读、语境应用和口头表达。计算机对学习者的口语能力的打分和评价将主要从这三个维度考虑。

3.5

语音输出能力 Phonetic Reading Ability

指学习者英语词汇发音的准确度，以及词汇在句子和语篇层面的朗读准确度、流利度，包括词汇的重音、重读、弱读等发音和朗读技巧的运用能力。

3.6

语境交流能力 Situational Interaction Ability

在不同的口语交流语境下，学习者选择和应用词汇和语义结构并完成交流任务的能力。

3.7

口头表达能力 Oral Expression Ability

指学习者组织语言，根据某一设定的话题以及所给的关键信息词，完成表达任务的能力，该话题与日常生活、语言文化和社会现象等主题有关。

3.8

英语口语计算机测评系统 Assessment System of Spoken English on Computer

指包含线上英语口语考试平台、英语智能语音打分系统、英语口语测试题库、用户数据统计和分析系统等在内的综合考试实施和管理系统。

3.9

英语口语计算机测评标准化考场 Standard Assessment Room of Spoken English on Computer

按照相应的计算机软硬件指标和设施标准配置的用于英语口语测评的考场。

3.10

英语口语计算机测评报告 Assessment Report of Spoken English on Computer

指学习者完成线上英语口语能力考试后，获得的和该次测试相关的语音朗读、语境交流和口头表达能力数据统计和分析结果；同时也指班级、学校、地区等集体单位获得的该次测评所有学习者的口语能力数据统计和分析结果。

3.11

英语口语计算机测评题库 Assessment Database of Spoken English on Computer

根据英语口语计算机测评的需要所建设的、根据相关规则能自动生成测评试卷的题库，存储于数据库。

4 缩略语

下列缩略语适用于本文件。

SETOC：英语口语能力等级计算机测评（Spoken English Test on Computer）

5 英语口语能力等级划分

5.1 英语口语能力框架

5.1.1 语音输出能力 Phonetic Reading Ability

表1 语音输出能力描述

词汇发音 (Word)	准确度 (Accuracy)
	发音技巧 (Pronunciation Skill)
句子朗读 (Sentence)	准确度 (Accuracy)
	发音技巧 (Pronunciation Skill)
	朗读技巧 (Reading Skill)
语篇朗读 (Passage)	准确度 (Accuracy)
	发音技巧 (Pronunciation Skill)
	朗读技巧 (Reading Skill)

5.1.2 语境交流能力 Situational Interaction Ability

表2 语境交流能力描述

词汇使用 (Word)	关键词汇准确度 (Accuracy of Key Words)
	辅助词汇准确度 (Accuracy of Supporting Words)
结构搭配 (Structure)	关键结构准确度 (Accuracy of Key Structure)
	语法准确度 (Grammatical Correctness)
主题契合 (Content)	信息完整度 (Information Completeness)
流利度 (Coherence)	内容及语流连贯度 (Coherence of Content and Language Stream)

5.1.3 口头表达能力 Oral Expression Ability

表3 口头表达能力描述

词汇使用 (Word)	关键词汇准确度 (Accuracy of Key Words)
	辅助词汇准确度 (Accuracy of Supporting Words)
结构搭配 (Structure)	关键结构准确度 (Accuracy of Key Structure)

	语法准确度 (Grammatical Correctness)
主题契合 (Content)	语义完整度 (Content Completeness)
	逻辑合理度 (Sequence Reasonableness)
流利度 (Coherence)	内容及语流连贯度 (Coherence of Content and Language Stream)

5.2 英语口语能力等级计算机测评 (SETOC) 阶段划分

表4 英语口语能力等级阶段划分

等级		对应阶段
SETOC 一级	1D 入门 (Beginner)	小学三年级上学期(含一、二年级)
	1C 基础 (Basic)	小学三年级下学期
	1B 熟练 (Proficient)	小学四年级上学期
	1A 精通 (Expert)	小学四年级下学期
SETOC 二级	2D 入门 (Beginner)	小学五年级上学期
	2C 基础 (Basic)	小学五年级下学期
	2B 熟练 (Proficient)	小学六年级上学期
	2A 精通 (Expert)	小学六年级下学期
SETOC 三级	3C 基础 (Basic)	初中一年级
	3B 熟练 (Proficient)	初中二年级
	3A 精通 (Expert)	初中三年级
SETOC 四级	4C 基础 (Basic)	高中一年级/中职
	4B 熟练 (Proficient)	高中二年级/高职高专
	4A 精通 (Expert)	高中三年级/高职高专
SETOC 五级	5C 基础 (Basic)	大学专科
	5B 基础 (Basic)	大学本科
	5A 熟练 (Proficient)	英语专业 (本科)
SETOC 六级	6B 基础 (Basic)	
SETOC 七级	6A 熟练 (Proficient)	英语专业 (研究生)
	7B 熟练 (Proficient)	
SETOC 七级	7A 精通 (Expert)	
	8B 熟练 (Proficient)	
SETOC 八级	8A 精通 (Expert)	高端外语人才
	9B 熟练 (Proficient)	
SETOC 九级	9A 精通 (Expert)	

5.3 英语口语能力等级计算机测评 (SETOC) 评价框架

表5 英语口语能力等级计算机测评 (SETOC) 能力描述及评价框架

等级 (Level)	能力项目 (Competence Item)	评价参考 (Description)
---------------	---------------------------	-----------------------

SETOC 一级	<p>语音输出：能够准确完成适合本阶段学习的字母和词汇的拼读、听读、辨读。（语料语速：80-100词；文本长度：5—30词。）</p> <p>语境交流：能够完成日常问候、与学校及家庭生活有关的常见场景应答。</p> <p>口头表达：能够完成和个人年龄、爱好、家庭成员等的简短介绍。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>过程：不连贯/连贯</p>
SETOC 二级	<p>语音输出：能够准确完成适合本阶段学习的词汇、短语和句子的拼读、听读、辨读。（语料语速：90-120词；文本长度：20—40词。）</p> <p>语境交流：能够完成日常问候、与日常学习生活有关的常见场景问答，并发表个人见解。</p> <p>口头表达：能够完成个人信息介绍、家庭成员和朋友介绍，且信息详尽；能够对一次具体的学校或家庭活动进行过程性描述。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>过程：不连贯/连贯</p>
SETOC 三级	<p>语音输出：能够准确完成适合本阶段学习的词汇、短语、句子和语篇的拼读、听读和朗读。（语料语速：110-130词；文本长度：30—50词。）</p> <p>语境交流：能够完成熟悉的学校、家庭和社会场景交流，能发表个人见解。</p> <p>口头表达：能够完成个人及家庭成员的详细介绍，能够就朋友间关系及喜好等信息进行描述，能够对亲身经历的一次具体活动或熟悉的事件/话题展开阐述。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>
SETOC 四级	<p>语音输出：能够准确完成适合本阶段学习的词汇、句子和较长语篇的朗读和听读。（语料语速：120-140词；文本长度：40—70词。）</p> <p>语境交流：能够完成常见生活场景的问答，能够参与并完成熟悉的话题讨论，发表个人见解并交换意见。</p> <p>口头表达：能够完成人物介绍或熟悉的事物描述，能够就熟悉的经历进行描述或就熟悉的事件/话题进行阐述，并给出评价或建议。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>
SETOC 五级	<p>语音输出：能够准确完成适合本阶段学习的句子和语篇的朗读和听读。（语料语速：140-170词；文本长度：60—100词。）</p> <p>语境交流：能就日常话题提供信息，表达观点和意见，参与讨论；能在口头表达中进行适当的自我修正，能根据话题进行情景对话。</p> <p>口头表达：能够就有关日常生活、兴趣爱好、风俗习惯、科学文化等方面的话题发表评价和建议。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>
SETOC 六级	<p>语音输出：能够准确完成适合本阶段学习的句子和语篇的朗读和听读。（语料语速：150-190词；文本长度：90—120词。）</p> <p>语境交流：能够参与复杂话题的讨论并发表个人见解，语言丰富、流畅，思路清晰。</p> <p>口头表达：能够就有关日常生活、兴趣爱好、风俗习惯、科学文化等方面的话题发表评价和建议，所表述内容逻辑紧密。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>

SETOC 七级	<p>语音输出：能够准确完成适合本阶段学习的句子和语篇的朗读和听读。（语料语速：160—200词；文本长度：100—150词）</p> <p>语境交流：能够参与复杂话题讨论，可以完成多轮交流，完整发表个人见解，语言丰富、流畅，思路清晰。</p> <p>口头表达：能够对有关历史文化、自然科学和社会经济等话题等方面的话题发表评价和建议，所表述内容逻辑紧密。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>
SETOC 八级	<p>语音输出：能够准确完成适合本阶段学习的常用语篇朗读和听读。（语料语速：170—220词；文本长度：110—160词。）</p> <p>语境交流：能够参与或主导复杂话题讨论，可以完成多轮交流，完整发表个人见解，并总结归纳，语言丰富、流畅，思路清晰。</p> <p>口头表达：能够对有关历史文化、自然科学和社会经济等话题等方面的话题发表评价和建议，所表述内容逻辑紧密。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>
SETOC 九级	<p>语音输出：能够准确完成适合本阶段学习复杂语篇朗读和听读。（语料语速：170—220词；文本长度：110—160词。）</p> <p>语境交流：能够根据各类主题需要，主导话题或研讨，分析并总结，语言丰富、流畅，思路清晰。</p> <p>口头表达：能够对有关学术交流、历史文化、自然科学和社会经济等话题等方面的话题发表评价和建议，所表述内容逻辑紧密。</p>	<p>发音：错误/个别错误/准确</p> <p>朗读：不流畅/有停顿/流畅</p> <p>语调：生硬/一般/自然</p> <p>应用：不恰当/恰当</p> <p>内容：不完整/有缺失/完整</p> <p>过程：不连贯/连贯</p>

5.4 英语口语能力等级计算机测评（SETOC）分项能力评价

5.4.1 语音输出能力-Phonetic Reading Ability

5.4.1.1 准确度和清晰度——Correctness and Clearness

表6 英语口语能力等级计算机测评语音输出能力描述（准确度和清晰度）

能力等级	准确度和清晰度能力描述
SETOC 一级	能够完成字母和常用词汇的拼读；以及日常生活中常见场景短语和句子的朗读或听读；发音准确，辨识度高。（句子：简单句）
SETOC 二级	能够完成常用词汇的拼读，以及日常生活中常见场景短语和句子的朗读或听读，发音准确，辨识度高。（句子：简单句+并列句）
SETOC 三级	能够完成常见词汇以及日常生活中常见场景短语、句子、语篇的拼读、朗读或听读，词汇发音准确，辨识度高。（句子：并列句+复杂句；语篇30—50词）
SETOC 四级	能够日常生活中常见场景句子和较长语篇的朗读或听读，语篇中词汇发音准确，重读准确，辨识度高。（句子：并列句+复杂句；语篇40—70词）
SETOC 五级	能够完成日常生活中常见场景复杂语篇的朗读或听读，语篇中词汇发音准确，重读准确，辨识度高。（句子：多种句式；语篇60—100词）
SETOC 六级	能够完成多种话题复杂语篇的朗读，以及常见话题语篇的听读，语篇中词汇发音准确，重读准确，辨识度高。（句子：多种句式；语篇90—120词）
SETOC 七级	能够完成多种话题复杂语篇的朗读和听读，语篇中词汇发音准确，重读准确，辨识度高。（句子：多种句式，包含复杂词汇；语篇100—150词）
SETOC 八级	能够完成多种话题复杂语篇的朗读和听读，语篇中词汇发音准确，重读准确，辨识度高。（句子：多种句式，包含复杂词汇；语篇110—160词）

SETOC 九级	能够完成多种话题复杂语篇的朗读和听读，语篇中词汇发音准确，重读准确，辨识度高。（句子：多种句式，包含专业词汇；语篇150-200词）
----------	--

5.4.1.2 韵律和语调——Rhythm and Tone

表7 英语口语能力等级计算机测评语音输出能力描述（韵律和语调）

能力等级	韵律和语调能力描述
SETOC 一级	能够辨别和区分常用词汇中的音节的重读、弱读，句子语调运用恰当。
SETOC 二级	能够辨别和区分常用词汇中的音节的重读、弱读，能够使用连读、省略等技巧完成日常生活中常见场景句子和语篇朗读，语调恰当，能注意句子中词汇的重读和弱读。
SETOC 三级	能够辨别和区分词汇中音节的重读、弱读；能够使用连读、省略等技巧完成日常生活中常见场景句子和语篇朗读，语调恰当，能够根据语境实现词汇的重读和弱读。
SETOC 四级	能够使用连读、省略等技巧完成句子和较长语篇朗读，语调恰当，能够根据语境实现词汇的重读和弱读，并通过停顿等技巧的运用提升语篇朗读的韵律性。
SETOC 五级	能够使用多种朗读技巧完成句子和复杂语篇朗读，语调恰当，能够根据语境实现词汇的重读和弱读，并通过停顿等技巧的运用提升语篇朗读的韵律性。
SETOC 六级	能够使用多种朗读技巧完成复杂语篇朗读，语调恰当，能够根据语境实现词汇的重读和弱读，主题突出，并通过停顿等技巧的运用提升语篇朗读的韵律性。
SETOC 七级	能够使用多种朗读技巧完成多种题材复杂语篇朗读，语调恰当，能够根据语境实现词汇的重读和弱读，主题突出，并通过停顿等技巧的运用提升语篇朗读的韵律性。
SETOC 八级	能够使用多种朗读技巧完成多种题材复杂语篇朗读，语调自然，能够根据语境调整词汇的重读、弱读等，强调或突出主题，并通过停顿等技巧的运用提升语篇朗读的韵律性。
SETOC 九级	能够使用多种朗读技巧完成多种题材复杂语篇或专业文档的朗读，语调自然，能够根据语境调整词汇的重读、弱读等，强调或突出主题，并通过停顿等技巧的运用提升语篇朗读的韵律性。

5.4.2 语境交流能力 Situational Interaction Ability

表8 英语口语能力等级计算机测评语境交流能力描述

能力等级	能力描述
SETOC 一级	能够选择恰当的词汇、短语或句子完成与日常生活相关简单场景的应答任务，如个人信息介绍、辨识物品、谈论天气/喜好等日常生活常见话题。
SETOC 二级	能够选择恰当的词汇、短语或句子发起或完成与日常生活相关简单场景的交互任务，如个人信息介绍、辨识物品、谈论天气/喜好等日常生活常见话题、日常文体活动、日常出行等。
SETOC 三级	能够选择恰当的词汇、短语或句子发起或完成与日常生活相关较复杂场景的交互任务，如人物描述、旅行或聚会、学习和工作计划、问题沟通与交流、各类文体活动等。
SETOC 四级	能够选择恰当的词汇、短语或句子发起或完成与日常生活相关复杂场景的交互任务，如就某一话题发表见解，谈论未来发展规划，医院、餐馆等常见社会场所的谈话，赞扬和鼓励他人，以及描述事务的特征等。
SETOC 五级	能够参与并完成与日常生活、历史文化和社会经济等相关复杂场景的交互任务，就某一话题发表见解、建议，提出方案或规划，能够根据场景需要，对相关环节进行描述。
SETOC 六级	能够参与并完成与日常生活、历史文化和社会经济等相关复杂场景的多环节交互任务，就某一话题发表见解、建议，提出方案或规划，能够根据场景需要，对相关环节进行描述。

SETOC 七级	能够主导并完成与日常生活、历史文化和社会经济等相关复杂场景的多环节交互任务，并根据任务需求把控各环节交流进程以及讨论主题，并恰当发表个人见解、建议，提出方案或规划，能够根据场景需要，对相关环节进行描述。
SETOC 八级	能够主导并完成各类历史文化、社会经济、环境科学等相关复杂场景的多环节交互任务，并根据任务需求把控各环节交流进程以及讨论主题，并恰当发表个人见解、建议，提出方案或规划，能够根据场景需要，对相关环节进行描述。
SETOC 九级	能够主导并完成与历史文化、社会经济、环境科学和学术讨论等相关复杂场景的多环节交互任务，并根据任务需求把控各环节交流进程以及讨论主题，并恰当发表个人见解、建议，提出方案或规划，能够根据场景需要，对相关环节进行描述。

5.4.3 口头表达能力—Oral Expression Competence

5.4.3.1 完整度——Completeness

表9 英语口语能力等级计算机测评口头表达能力描述（完整度）

能力等级	完整度能力描述
SETOC 一级	针对简短话题所表述内容关键信息点完整，前后逻辑符合。
SETOC 二级	针对简短话题所表述内容关键信息点完整，并对细节进行恰当描述，前后逻辑符合。
SETOC 三级	针对简短话题所表述内容关键信息详实，并对细节进行恰当描述，前后逻辑符合。
SETOC 四级	针对较复杂话题所表述内容关键信息点完整，并对细节进行恰当描述，前后逻辑符合。
SETOC 五级	针对较复杂话题所表述内容关键信息详实，并对细节进行恰当描述，前后逻辑符合。
SETOC 六级	针对较复杂话题所表述内容关键信息详实，细节描述完整，前后逻辑符合。
SETOC 七级	针对复杂话题所表述内容关键信息详实，细节描述完整，前后逻辑符合。
SETOC 八级	针对复杂话题所表述内容关键信息详实，细节描述完整，并根据需要适当扩充，前后逻辑符合。
SETOC 九级	针对复杂或专业话题所表述内容关键信息详实，细节描述完整，并根据需要适当扩充，前后逻辑符合。

5.4.3.2 准确度——Correctness

表10 英语口语能力等级计算机测评口头表达能力描述（准确度）

能力等级	准确度能力描述
SETOC 一级	能够使用和所要求表达任务有关的词汇，完成简单主题任务的表达；相关度高。
SETOC 二级	能够使用和所要求表达任务有关的词汇、短语，完成简单主题任务的表达；相关度高，语句符合语言使用规则。
SETOC 三级	能够使用和所要求表达任务有关的词汇、短语和句子，完成简单主题任务的表达；相关度高，语句符合语言使用规则。
SETOC 四级	能够使用和所要求表达任务有关的词汇、短语和句子，完成较长主题任务的表达；相关度高，语句符合语言使用规则，语义重点突出。
SETOC 五级	能够使用和所要求表达任务有关的词汇、短语和句子，完成较复杂主题任务的表达；相关度高，语句符合语言使用规则，语义重点突出。
SETOC 六级	能够使用和所要求表达任务有关的词汇、短语和句子，完成复杂主题任务的表述；相关度高，语句符合语言使用规则，语义重点突出。
SETOC 七级	能够使用和所要求表达任务有关的词汇、短语和句子，完成复杂主题任务的表述；相关度高，语句符合规则，且有适当扩展，语义重点突出。

SETOC 八级	能够使用和所要求表达任务有关的词汇、短语和句子，完成复杂主题任务的表述；相关度高，语句符合规则，句式丰富，语义重点突出且全面。
SETOC 九级	能够使用和所要求表达任务有关的词汇、短语和句子，完成复杂主题或学术专题的表述；相关度高，语句符合规则，句式丰富，语义重点突出且全面。

5.4.3.3 转述/复述能力——Paraphrase/Retell

表11 英语口语能力等级计算机测评口头表达能力描述（转述/复述能力）

能力等级	转述/复述能力描述
SETOC 一级	能够完成词汇和短语的复述, 发音准确, 关键词汇无缺失。
SETOC 二级	能够完成短语和简单句子的转述/复述, 发音准确, 关键词汇无缺失。
SETOC 三级	能够完成简单句子以及简短语篇的转述/复述, 所述内容逻辑正确, 关键信息完整。
SETOC 四级	能够完成复杂句子以及简短语篇的转述/复述, 所述内容逻辑正确, 关键信息完整。
SETOC 五级	能够完成复杂句子以及较长语篇的转述/复述, 所述内容逻辑正确, 关键信息完整。
SETOC 六级	能够完成复杂句子以及较复杂语篇的转述/复述, 所述内容逻辑正确, 关键信息完整。
SETOC 七级	能够完成复杂句子以及复杂语篇的转述/复述, 所述内容逻辑正确, 关键信息完整。
SETOC 八级	能够完成复杂语篇的转述/复述, 所述内容逻辑正确, 关键信息完整, 细节丰富。
SETOC 九级	能够完成复杂语篇或学术类语料的转述/复述, 所述内容逻辑正确, 关键信息完整, 细节丰富。

5.4.3.4 语言复杂度——Complexity

表12 英语口语能力等级计算机测评口头表达能力描述（语言复杂度）

能力等级	语言复杂度能力描述
SETOC 一级	无要求。
SETOC 二级	能够使用基础句式结构完成简单口头表达任务, 词汇及句子变化较小。
SETOC 三级	能够使用多种句式结构完成常见口头表达任务, 词汇较丰富, 句子长度根据表达目的不同而有所变化。
SETOC 四级	能够使用多种句式结构完成常见口头表达任务, 部分使用较复杂句式来突出或强化表达内容, 词汇变化多, 多处表达中使用复杂词汇。
SETOC 五级	能够使用多种句式结构完成较复杂话题任务, 部分使用较复杂句式来突出或强化表达内容, 词汇变化多, 多处表达中使用复杂词汇。
SETOC 六级	能够使用多种句式结构完成复杂话题任务, 部分使用较复杂句式来突出或强化表达内容, 词汇变化多, 多处表达中使用复杂词汇。
SETOC 七级	能够使用多种句式结构完成复杂话题任务, 根据需要使用较复杂句式来突出或强化表达内容, 词汇变化多, 多处表达中使用复杂词汇。
SETOC 八级	能够使用多种句式结构完成复杂话题任务, 多处使用复杂句式来描述或突出表述细节, 重点清晰, 词汇变化多, 多处表达中使用复杂词汇。
SETOC 九级	能够使用多种句式结构完成复杂话题或学术类表述任务, 多处使用复杂句式来描述或突出表述细节, 重点清晰, 词汇变化多, 多处表达中使用复杂词汇或专业术语。

5.4.3.5 流利度——Fluency

表13 英语口语能力等级计算机测评口头表达能力描述（流利度）

能力等级	流利度能力描述
SETOC 一级	在日常问候等常见生活场景中，能够自然、流畅地应答或发起讨论。
SETOC 二级	在日常交流或简单表述任务中能够做到语流连贯、自然、语速适当，断句合理，允许有部分语言使用上的重复或修正。
SETOC 三级	在日常交流或常见主题表述任务中能够做到语流连贯、自然、语速适当，允许有部分语言使用上的重复或修正。
SETOC 四级	在常见生活场景交流或多种话题表述任务中能够做到语流连贯、自然、语速适当，允许有部分语言使用上的重复或修正。
SETOC 五级	在常见生活场景交流或较复杂话题表述任务中能够做到语流连贯、自然、语速适当，允许有部分语言使用上的重复或修正。
SETOC 六级	在常见生活场景交流和较专业话题表述任务中能够做到语流连贯、自然、语速适当，允许有部分语言使用上的重复或修正，整体表达具有一定韵律度。
SETOC 七级	在各类场景交流和较专业话题表述任务中能够做到语流连贯、自然、语速适当，整体表达韵律度明显。
SETOC 八级	在各类场景交流和专业话题表述任务中能够做到语流连贯、自然、语速适当，整体表达韵律度明显。
SETOC 九级	在各类场景交流、专业或学术专题等表述任务中能够做到语流连贯、自然、语速适当，整体表达韵律度明显。

6 英语口语能力等级计算机测评（SETOC）测试题型

6.1 SETOC 考查语音输出能力

6.1.1 题型

语料朗读、听读。

6.1.2 可选语料形式

词汇、短语、句子、语篇。

6.1.3 出题形式

系统以文本或音频的形式提供要求朗读的内容，考生参照朗读。

6.1.4 系统采分点

词汇音准、重读音节、词汇重读、弱读、连读、省略、语调。

6.1.5 可输出测评结果

音素得分、词汇得分、句子总分、语篇总分、流利度分、完整度分、韵律度分。

6.1.6 可输出结果形式

根据考生的测试结果对所使用测试文本以不同颜色进行区分标注，通过字体加粗等形式对重读进行标注，通过不同符号的使用对连读、弱读、省略和语调等进行标注。

6.1.7 题型解读

本题型从不同形式的文本层面考查考生的语音面貌。考生需要清晰、准确地读出屏幕上显示的词汇、句子语篇，在朗读过程中，需要注意元音、辅音的发音规范以及连读、重读、语调等发音技巧的运用。

6.2 SETOC 考查语境交流能力

6.2.1 题型

情景应答。

6.2.2 可选语料形式

句子/短对话/长对话。

6.2.3 出题形式

本题是通过预设一个场景任务，让考生根据所设条件补全信息、选择正确答语或完整应答。

6.2.4 系统采分点

词汇音准、词汇重读、连读、句子语调、内容准确度、流利度、完整度。

6.2.5 可输出测评结果

词汇得分、句子总分、流利度分、完整度分、韵律度分。

6.2.6 可输出结果形式

根据考生的测试结果对所使用测试文本以不同颜色进行区分标注，通过字体加粗等形式对重读进行标注，通过不同符号的使用对连读、弱读、省略和语调等进行标注。

6.2.7 题型解读

本部分题目实际属于半开放题型，通过屏幕以文本形式给出的既定场景，将口头表达的范围限制在一个非常小的区域内。通常情况下，考生所需回答的所有信息在提示语中已经给出，只需重新组织语言“复述”出来即可。

6.3 SETOC 考查口头表达能力

6.3.1 题型

话题表达。

6.3.2 出题形式

本题通过限定一个主题，并给出部分信息参考词的形式，让考生根据所给信息组织话语完成话题表达。

6.3.3 系统采分点

内容准确度、流利度、语言复杂度。

6.3.4 可输出测评结果

内容总分、语言复杂度分、流利度分、完整度分、韵律度分。

6.3.5 可输出结果形式

具体分数及相关维度能力评价。

6.3.6 题型解读

本部分题目属于开放题型，考查考生口语综合应用的能力。在所给信息相对较少的情况下，考生须组织语言，在很短的时间内用符合表达规则和习惯底的话语完成表达任务。

7 英语口语能力等级计算机测评（SETOC）测评维度

7.1 SETOC 语音输出能力的计算机测评维度

7.1.1 单词测评维度

词汇总分、音素得分。

7.1.2 句子测评

句子总分、准确度、流利度、完整度、各单词发音得分、重读、停顿、句末升降调、连读等。

7.1.3 语篇测评维度

语篇总分、准确度、流利度、完整度、每个句子的发音得分、每个单词的发音得分。

7.2 SETOC 语境交流能力的计算机测评维度

语音分、流利度、词汇准确度、语义准确度、内容完整度和语法使用。

7.3 SETOC 口头表达能力的计算机测评维度

内容总分、语音分、内容相关度、内容完整度、语句应用、流利度、语法使用等。

8 英语口语能力等级计算机测评（SETOC）系统

8.1 SETOC 系统定义

英语口语能力等级计算机测评（SETOC）系统是实现英语口语测评从命题、制卷到机器阅卷、自动评分的全过程操作，辅助开展大规模、常态化的英语口语测评的软、硬件集合。

8.2 SETOC 系统架构

系统包括考务管理系统、命题制卷系统、阅卷评分系统和考试系统。

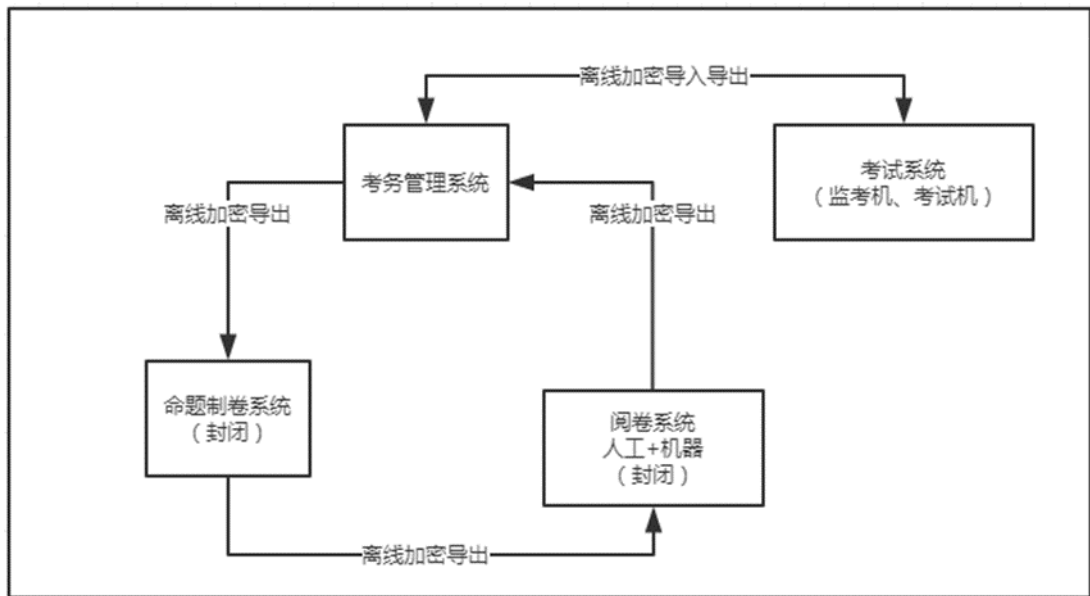


图1 英语口语能力等级计算机测评系统架构图

8.3 SETOC 考务管理系统

8.3.1 SETOC考务管理系统用户

系统管理员、考办和考点管理员。

8.3.2 SETOC考务管理系统功能

设置当前考试计划、管理考点信息、管理考生编排、考打印务资料、管理考生成绩。

8.4 SETOC 命题制卷系统

8.4.1 命题制卷系统用户

命题管理员、命题专家、命题负责人、审查员外等。

8.4.2 SETOC命题制卷系统功能

导入考试计划、命题制卷、导出试卷、预览试卷、验收试卷。

8.5 SETOC 阅卷评分系统

8.5.1 SETOC考试系统用户

系统管理员/某级考办、阅卷管理员、命题专家、阅卷专家、阅卷教师、仲裁员、复审员。

8.5.2 SETOC考试系统功能

准备数据、人工评分、智能评分、发布成绩。

8.6 SETOC 考试系统

8.6.1 SETOC考试系统用户

T/C11A xxx—xxxx

监考老师、参考学生。

8.6.2 SETOC考试系统功能

8.6.2.1 监考机功能

导入数据、登录、控制考试开始和结束。

8.6.2.2 考试机功能

登录并核对考试信息、检测设备、答题。

8.7 SETOC 系统运行环境

表14 SETOC系统配置

场景	设备	性能要求
考试现场	监考机	CPU: Intel i5 (或者同等性能) 内存: 2G 硬盘: 100G 接口: USB×1 操作系统: Windows XP SP3, Windows 7或以上
	考试机	CPU: Intel i3 (或者同等性能) 内存: 2G 硬盘: 100G 接口: USB×1 操作系统: Windows XP SP3, Windows 7或以上
	网络带宽	100M以上
	耳麦	接口: USB 声卡: 集成声卡 类型: 头戴护耳 线长: 2.8米 喇叭频率范围: 20Hz~20KHz 喇叭灵敏度: 98dB ±3dB 喇叭阻抗: 32Ω±15% 麦克风类型: 单指向 麦克风频率范围: 20Hz~16KHz 麦克风灵敏度: -42±3dB 麦克风阻抗: 0.68KΩ
人工评分	服务器	CPU: Intel E5-2670 (或者同等性能) 内存: 8G 硬盘: 100G 接口: USB×1 操作系统: Windows 2008或以上
	客户端	CPU: Intel i3 (或者同等性能) 内存: 2G 硬盘: 100G

		接口：USB×1 操作系统：Windows XP SP3, Windows 7或以上
	耳麦	接口：USB 声卡：集成声卡 类型：头戴护耳 线长：2.8米 喇叭频率范围：20Hz~20KHz 喇叭灵敏度：98dB ±3dB 喇叭阻抗：32Ω±15% 麦克风类型：单指向 麦克风频率范围：20Hz~16KHz 麦克风灵敏度：-42±3dB 麦克风阻抗：0.68KΩ
	网络	100兆
智能评分	任务节点	CPU：Intel i5（或者同等性能） 内存：4G 硬盘：500G 操作系统：CentOS或以上
	评分节点	CPU：Intel E5-2670（或者同等性能） 内存：8G 硬盘：500G 操作系统：CentOS或以上
	操作节点	CPU：Intel i5（或者同等性能） 内存：4G 硬盘：500G 接口：USB×1 操作系统：Windows 7或以上
	网络	1000M或以上

8.8 SETOC 系统性能

8.8.1 容量

单机房同时在线考生20-50人。

8.8.2 学生端响应速度

学生端每秒上报考试状态，上传答案时间小于10秒。

8.8.3 教师端响应速度

教师端下载任务时间小于10秒，上传答案到云平台的时间小于30秒。

8.8.4 安全

学生端、教师端、云平台的数据双重备份；任务包、答案包等敏感数据加密处理。

8.8.5 同时在线人数

一天2万人。

8.8.6 测评响应速度

从考试结束到输出成绩，机器智能评分24小时；异常音频处理4天；生成和发布成绩报告1天。

8.8.7 与人工测评的一致性

Pearson相关系数（反映计算机测评与人工测评之间相关关系密切程度的统计指标） $\geq 75\%$ ；容错一致度（20%） $\geq 75\%$ 。

9 英语口语能力等级计算机测评（SETOC）考务

9.1 SETOC 考试准备

主要步骤包括：成立考试管理机构、组建考务队伍、培训考务人员、部署人机考试环境、准备考场、准备候考室、准备签字处、测试考试系统、命题制卷、制作考务考编。

9.2 SETOC 正式考试

主要步骤包括：组织考生进入候考室、组织考生进入考场、宣布开始考试、监督考生答题、处置考场异常情况、督促考生交卷、督促考生离开考场、上报考试数据。

9.3 SETOC 考后评分

9.3.1 抽样人工评卷

按照本次机考的试题内容，划分不同科目、不同题型的打分任务；定义本次英语听说机考的评卷模式（双评加仲裁）；设置本次评卷允许评卷专家所评试卷的误差；设置评卷给分板的给分模式（直接给分、分布给分或按当次给分）。

9.3.2 智能评分

设置智能评分子系统的用户角色，建立不同的用户组，分配相应的用户权限。

设置考生样卷的抽取比例，智能评分子系统自动从全部考生答题数据中随机抽取答题数据，建立考生定标集；评卷专家组组织评卷专家对考生定标集进行人工定标评分，掌握评分进度；将定标集作为机器学习的初始化数据集，为后期智能评分做准备。

根据人工定标结果，自动训练和优化智能评分策略；对全部考生答题数据开始自动评分试评，评卷专家组可以随机提取评分结果进行人工比对，如果发现误差较大可以重新调整智能评分策略，对考生答题结果进行重新评分。

9.4 SETOC 成绩发布及复查

9.4.1 成绩发布

智能评分结束后，由考点办公室统一公布考生考试结果。

9.4.2 成绩复查

考生如对本人成绩有疑问,在有效时间内办理申请查分手续,各考点办公室将申请复查考生名单汇总后上报进行复查,复查结果由各考点及时通知考生本人。

9.5 SETOC 区块链测评成绩管理系统

9.5.1 在SETOC测评成绩管理中应用区块链技术的意义

随着英语口语计算机测评在普及,成绩管理的重要性日益突出。目前的成绩管理系统多为中心化的。虽然其信息处理流程简单,效率较高,但存在单点故障风险和单点突破风险。此外,还存在内部人员非法修改测评成绩的风险。

针对上述风险,可以在测评成绩管理系统中采用区块链技术。区块链技术是在多方无需互信的环境下,通过密码学技术让系统中所有参与方协作,来共同维护可靠数据日志的方式。它具有分布式、去中心化和安全可信的特点,可以利用该技术提高测评成绩管理系统的安全性。

9.5.2 SETOC区块链测评成绩管理系统的组成

区块链测评成绩管理系统由三大模块组成:测评数据及元数据管理模块、用户身份管理及验证模块、用户利用信息管理模块。这三部分的相互关联需要通过智能合约来触发。用户如需查询测评数据,首先需要校验其身份信息及获取测评数据的相应权限,通过审查后调取测评数据及元数据,而其查询测评数据的痕迹将被记录在区块链上,以便将来进行用户行为统计和分析。结合测评数据的管理内容和流程以及实际利用需求,可将这三部分模块可以细化成多个更小的模块。

9.5.3 SETOC区块链测评成绩管理系统的主要功能

9.5.3.1 注册功能

9.5.3.1.1 学生注册

系统为学生提供注册界面。打开系统后,选择学生角色,输入姓名、学号、学校、地区等相关信息后,SETOC区块链测评成绩管理系统利用密钥生成器生成公私钥对,私钥由学生保存,公钥提交给学校或教育主管部门。注册完成之后,根据学生的注册信息,创建虚拟资产,并且把虚拟资产提交给学校或教育主管部门。

9.5.3.1.2 教师注册

系统为教师提供注册界面。打开系统后,选择教师角色,输入姓名、工号、班级、学校等信息后,SETOC区块链测评成绩管理系统利用密钥生成器生成公私钥对,私钥由教师保存,公钥提交给学校或教育主管部门,并记录在册。

9.5.3.1.3 学校或教育主管部门在注册过程中的作用

学校或教育主管部门也是作为一个节点存在,是区块链的交易主体。学校或教育主管部门不必注册,在系统安装时直接生成学校或教育主管部门的公私密钥对,并且公开学校的公钥。

9.5.3.2 成绩上传及修改功能

教师在系统上将成绩整理好以后,创建虚拟资产,将该交易提交给学校或教育主管部门。如果要修改成绩,则创建相应的虚拟资产,并且把该交易发送给学校或教育主管部门。若收到学校或教育主管部门的同意信息,则重新上传修改后的成绩;若收到学校或教育主管部门的拒绝信息,则不修改成绩。

学校或教育主管部门接收到教师修改成绩的交易后，首先审核，如果同意修改，则创建同意的虚拟资产返回给教师；如果不同意修改，则创建拒绝的虚拟资产返回给教师。

9.5.3.3 成绩查询功能

SETOC区块链测评成绩管理系统方便学生查询成绩，学生可以使用浏览器通过HTTP请求来查询交易的数据，系统中设计的数据查询接口可以方便的查询交易的数据、区块的数据、节点信息以及区块状态等信息。对学生成绩的查询以学生的公钥为查询条件，如果学生的账户没有收到来自学校或教育主管部门的交易，说明目前还没有录入的成绩，成绩不可查询；如果教师已经成功提交成绩，并且被区块链成功验证，成绩可查询。

10 英语口语能力等级计算机测评（SETOC）报告

10.1 SETOC 学生测评报告

10.1.1 基础分数数据

学生测评总分、各大题得分、各小题得分、分项维度得分等。

10.1.2 基础分数对比统计

单项排名、综合排名、标准差等。

10.1.3 音频数据

标准示范音、学生音频。

10.1.4 基础语音测评结果

以颜色区分朗读文本、以符号标记重读及升降调等信息。

10.1.5 基础语音数据统计

词汇统计、音素统计等。

10.1.6 基础语音能力评价

语音能力综合评价、语音能力分项评价、分项能力雷达图分析等。

10.1.7 语境交流能力数据统计

分项能力值、标准差等。

10.1.8 语境交流能力评价

语境交流能力综合评价、语境交流能力项评价。

10.1.9 口头表达能力数据

分项能力值、标准差等。

10.1.10 口头表达能力评价

口头表达能力综合评价、口头表达分项能力评价。以上所有数据可接受定制。

10.2 SETOC 班级测评报告

10.2.1 基础分数数据统计

班级总分平均分、各大题平均分、各小题平均分、最高分和最低分、数据提交结果、测评完成情况等。

10.2.2 基础分数数据对比统计

总分排名、分项排名、标准差等。

10.2.3 基础语音数据统计

学生个人报告链接、词汇统计、音素统计等。

10.2.4 基础语音能力评价

语音能力综合评价、语音能力分项评价。

10.2.5 语境交流能力数据统计

分项能力平均值、最高值和最低值、分项排名、标准差等。

10.2.6 语境交流能力数据评价

语境交流能力综合评价、语境交流能力分项评价。

10.2.7 口头表达能力数据统计

分项能力平均值、最高值和最低值、分项排名、标准差等。

10.2.8 口头表达能力评价

口头表达能力综合评价、口头表达分项能力评价。

10.2.9 学生数据横向对比

区间值统计、分项能力雷达图分析、优秀率等。以上所有数据可接受定制。

10.3 SETOC 学校测评报告

10.3.1 基础分数数据统计

总分平均分、各大题平均分、各小题平均分、最高分和最低分、各分数段学生分布、班级分数段对比等。

10.3.2 基础语音能力评价

语音能力综合评价、语音能力分项评价。

10.3.3 基础语音数据统计

各能力值区间学生分布、学生能力强弱雷达图分析、班级分项数据对比、班级报告链接、学校综合词汇统计、学校综合音素统计等。

10.3.4 语境交流能力数据统计

T/CIIA xxx—xxxx

分项能力平均值、最高值和最低值、分项排名、标准差、班级分项数据对比等。

10.3.5 语境交流能力综合评价

语境交流能力综合评价、语境交流能力分项评价。

10.3.6 口头表达能力数据统计

分项能力平均值、最高值和最低值、分项排名、标准差、班级分项数据对比等。

10.3.7 口头表达能力评价

口头表达能力综合评价、口头表达分项能力评价。以上所有数据可接受定制。

10.4 SETOC 地区测评报告

10.4.1 基础分数数据统计

总分平均分、各大题平均分、各小题平均分、最高分和最低分、各分数段学生分布、学校分数段对比等。

10.4.2 基础语音能力评价

语音能力综合评价、语音能力分项评价。

10.4.3 基础语音数据统计

各能力值区间学生分布、学生能力强弱雷达图分析、学校分项数据对比、学校报告链接、地区综合词汇统计、地区综合音素统计等。

10.4.4 语境交流能力数据统计

分项能力平均值、最高值和最低值、分项排名、标准差、学校分项数据对比等。

10.4.5 语境交流能力综合评价

语境交流能力综合评价、语境交流能力分项评价。

10.4.6 口头表达能力数据统计

分项能力平均值、最高值和最低值、分项排名、标准差、学校分项数据对比等。

10.4.7 口头表达能力评价

口头表达能力综合评价、口头表达分项能力评价。以上所有数据可接受定制。

11 英语口语能力等级计算机测评（SETOC）题库建设

11.1 SETOC 通用题库建设

11.1.1 题型

11.1.1.1 封闭题型

a) 音标识读：朗读屏幕显示的音标；

- b) 单词朗读：朗读屏幕显示的单词；
- c) 单词听读：听单词录音，然后朗读该单词；
- d) 看图说单词：看屏幕上的图片，然后说出该图片所指事物的单词；
- e) 短语/句子朗读：朗读屏幕显示的短语/句子；
- f) 短语/句子听读：听短语/句子录音，然后朗读该短语/句子；
- g) 短文朗读：朗读屏幕显示的短文；
- h) 短文听读：听短文录音，然后朗读屏幕显示的该短文；
- i) 听对话选择正确的答案：听对话录音，然后读出正确的答案。

11.1.1.2 半开放题型

- a) 情景问答（阅读文字提示，回答问题）：阅读屏幕显示的情景提示，然后依照屏幕提示（或语音提示）依次口头回答问题；
- b) 情景问答（听录音，回答问题）：听短文（或对话）的录音，然后依照屏幕提示（或语音提示）依次口头回答问题；
- c) 信息转述：听短文录音，根据所听到的内容，记录相关信息，然后口头转述包含若干个要点的内容；
- d) 看图（或视频）回答问题：看图（或视频），然后依照屏幕提示（或语音提示）依次口头回答问题；
- e) 角色扮演：看视频听对话，然后扮演其中一个角色进行对话，再扮演另一个角色进行对话；
- f) 交际问答：听句子录音，说出相应的答语；
- g) 开放题型；
- h) 话题简述：听短文录音（或无此步骤），阅读屏幕上的中文（或英文）提示，然后用英语进行话题简述；
- i) 故事复述：听短文录音，然后用英语复述。

11.1.2 难度

难度是指试题的难易程度，是评价考试的一个非常重要的一个指标。一个题目，如果大部分考生都能答对，那么这个题目的难度就小；如果大部分考生都不能答对，那么这个题目的难度就大。

11.1.2.1 难度系数的获取方法

11.1.2.1.1 收集数据

- a) 数据来源：一是从以往的英语口语计算机测评中获取数据，尤其是大规模考试（如中考）的英语口语计算机测评，因为人数越多，各种偶然因素互相抵消的可能性就加大，其带来的不利的影响就越小；二是抽样，即当条件不成熟，无法获取历史数据时，从要考察的班级、学校或地区中分别抽出各个成绩段的学生若干名参与测评，从中获取具有代表性的数据；
- b) 数据收集范围：取决于题库服务的对象，即我们要获取哪个层面的难度系数，如班级、学校、地区等的试题、试卷的难度系数；
- c) 数据维度：包括考试时间、考试人数、班级名称、学校名称、地区名称、各学生的总分、各学生的每小題得分等。

11.1.2.1.2 统计

按照数据收集范围（班级、学校、地区等）统计出同一试卷每道题的平均得分。

11.1.2.1.3 计算

试题的难度系数的计算公式如下：

$$NDn = 1 - PFn/MFn$$

式中：

NDn——某试题的难度系数；

PFn——该试题的平均得分；

MFn——该试题的满分。

NDn越大，表明该试题的难度越高，反之则越低。

试卷的难度系数的计算公式如下：

$$NDsz = \sum_{n=1}^m Kn \times NDn$$

式中：

NDsz——试卷的难度系数；

Kn——某道试题的分数在整个试卷分数中的权重；

M——试卷的题数。

NDsz越大，表明该试卷的难度越高，反之则越低。

难度系数只能从总体上反映试题、试卷的难度，因受其它因素的影响，如学生在考试中发挥失常等，就会使获得的难度系数失真。随着历史数据积累得越多，难度系数就相应地更加接近真实情况。可以挖掘大量考试实践的数据，利用数据分析方法分析不同考题与难度等级的相关性，逐步调整其难度系数，直到趋于稳定。

11.1.2.2 难度系数在题库建设中的应用

题库中试题、试卷的难度系数标定之后，就可应用于针对各种考试的题库组卷。

由于英语口语计算机测评的同一次考试中会使用多套试卷，为保证测评的公平性，同一次考试中不同批次的试卷中各题型的难度系数应基本一致。

一般将难度值小于和等于0.3的试题定为容易题；大于0.3和小于0.6的定为中档题；大于和等于0.6的试题定为难题。

以高考为例，难度以适中为宜，单个试题的难度系数一般在0.3—0.7之间，试卷的难度系数一般在0.5—0.6之间。

11.1.3 区分度

区分度反映的是试题、试卷对考生水平的区分能力，区分度高的试题、试卷能将不同水平的考生区分开来，水平高的考生得高分，水平低的考生得低分。

11.1.3.1 区分度的获取方法

11.1.3.1.1 收集数据

类似难度系数的数据收集。

11.1.3.1.2 统计

按照数据收集范围（班级、学校、地区等）分别统计试题、试卷得分处于前27%的考生的平均分、得分处于后27%的考生的平均分。

11.1.3.1.3 计算

试题的区分度的计算公式如下：

$$QFn = \frac{Xh - Xl}{MF_n}$$

式中：

QFn——某试题的区分度；

Xh——该试题的得分处于前27%的考生的平均分；

Xl——该试题的得分处于后27%的考生的平均分；

MF_n——该试题的满分。

QFn越大，表明该试题的区分能力越好，反之则越差。

试卷的区分度的计算公式如下：

$$QFsz = \sum_{n=1}^m Kn \times QFn$$

式中：

QFsz——试卷的区分度；

Kn——某道试题的分数在整个试卷分数中的权重；

M——试卷的题数。

QFsz越大，表明该试卷的区分能力越好，反之则越差。

11.1.3.2 区分度在题库建设中的应用

区分度可应用于甄别哪些试题可进入题库。试题、试卷的区分度若在0.4以上，说明其区分度很好；若在0.3-0.39之间，说明其区分度较好；以上两类试题、试卷可进入题库。若区分度在0.2-0.29之间，说明其区分度不太好，需要修改；若在0.19以下，说明其区分度不好，应从题库中剔除。

与难度系数一样，为了保证英语口语计算机测评的公平性，同一次考试中不同试卷的区分度要基本一致。在组卷时，应尽量选择那些经过历史检验证明区分度基本一致的试题、试卷。

区分度也不是越大越好，如果太大，说明低分段与高分段的差别太大，容易使低分段的学生丧失信心。

造成区分度过高的三种情况，一是低分段太低，二是高分段太高，三是上述两种情况同时出现，其原因主要还是试题、试卷的难度过大。

11.1.4 信度

11.1.4.1 信度的定义

信度是指使用同一试卷对考生重复测验时，或两个平行试卷对考生测验时，所得测验分数的一致性和稳定性程度。

11.1.4.2 信度的获取方法

11.1.4.2.1 收集数据

类似难度系数的数据收集。

11.1.4.2.2 统计

统计帮忙不同时间、不同范围（班级、学校、地区等）的每道试题、每份试卷的平均分。

11.1.4.2.3 计算

试题的信度的计算公式如下：

$$XDn = \frac{(Max - Min)}{MFn}$$

式中：

XDn——某个试题的信度；

Max——该试题在不同时间、相同地点（班级、学校、地区等，下同）的考试中得分平均值的最大值；

Min——该试题在不同时间、相同地点的考试中得分平均值的最小值；

MFn——该试题的满分。

XDn越大，表明该试题的信度越低，反之则越高。

试卷的信度的计算公式如下：

$$XDsz = \sum_{n=1}^m Kn \times XDn$$

式中：

XDsz——试卷的信度；

Kn——某道试题的分数在整个试卷分数中的权重；

M——试卷的题数。

XDsz越大，表明该试卷的信度越低，反之则越高。

不同时间、相同地点（如2017年、2018年的初一年级）的平时成绩差异较大时，上述公式计算的信度的失真也是较大的；只有平时成绩差异较小时，利用上述公式计算出来的信度才比较有意义。

11.1.4.3 信度在题库建设中的应用

越是重要的考试，如中、高考，越要求试题、试卷的信度要高，缺少信度或信度较低都无法向考生和社会交代。所以针对这类考试的题库，试题、试卷的信度也是甄别其能否入库的一个重要指标。当然，对于利害性较低的考试，可适当降低对信度的要求。

11.1.5 题量

说明：确定题量的前提是，为防止出现抄袭、泄题现象，考试批次之间的试卷内容不能重复；每个考试批次需用若干套试卷（一般来说，每个考试批次用4套试卷，就基本能保证不出现抄袭现象）。

用于某个等级（如中考）一次考试第n个题型需具备的题量TLn：

$$TLn = PC \times TS \times SL$$

式中：

PC——批次；

TS——试卷套数；

SL——试卷中第n个题型的考题数量。

用于某个等级（如中考）一次考试需具备的总题量ZTL：

$$ZTL = \sum_{n=1}^m TLn$$

T/CIIA xxx—xxxx

式中：

M——试卷的题型数。

11.1.6 题的难度确定

11.1.6.1 依据教材

以大、中、小学教材中包含的音标、单词、句子、短文及语法、语境为依据，确定选题范围。

11.1.6.2 专家初定

由有丰富教学、测评经验的专家在上述选题范围中，初步设计考题及其难度系数。

11.1.6.3 实践调整

挖掘大量考试实践的数据，利用数据分析方法分析不同考题与难度等级的相关性，逐步调整其难度系数，直到趋于稳定。

11.1.7 组卷规则

11.1.7.1 类似题型互斥

为保证全面测评各个知识点，同一套试卷中类似题型不能重复出现。如封闭题型目前共有9种，同一套试卷中只能出现其中一种；半开放、开放题型亦然。

11.1.7.2 难度系数接近

为保证测评的公平性，同一次考试中不同批次的试卷中各题型的难度系数应基本一致。

11.1.7.3 分值占比变化

为适应对不同能力等级的考察要求，根据能力等级之不同调整各题型分值占比。低等级的试卷中，封闭题型的得分占比一般较大；随着等级的不断升高，半开放题型、开放题型的得分占比应逐步加大；考虑到地区教育水平的差异，可控制开放题型的得分占比，教育水平较高的地区开放题型的得分占比可适当提高，反之教育水平低有地区开放题型的得分占比可适当降低。

11.2 SETOC 自适应题库建设

11.2.1 自适应题库的使用对象

特殊需求（如选拔、竞赛，或教学改革等，下同）的口语测评。

11.2.2 题库题型

根据特殊需求，不作统一要求，根据测评需求具体确定。

11.2.3 题库题量

题库中每个题型的每个难度等级的考题题量至少应为考卷相应难度等级考题题量的40倍以上，可随着时间不断扩充。

11.2.4 考题难度分级

T/C11A xxx—xxxx

根据特殊需求,将每个题型的考题分为若干个难度等级。考题难度分级由专家确定,并通过考试实践进行一致性验证,不断修正。

11.2.5 组卷规则

根据对考生口语能力级别的预估(或考生自我选择),生成由相应难度等级考题所组成的初试试卷,试卷题型及题量根据特殊需求确定;

初试结束后,如果某个题型的得分为80-100分(按照百分制计分,下同),则二试试卷该题型的难度上升一个级别;如果某个题型的得分为60-79分,则二试试卷该题型的难度不变;如果某个题型的得分为0-59分,则二试试卷该题型的难度下降一个级别;按照上述规则生成二试试卷;

依此类推,二试结束后,相继生成三试、四试、……试卷并完成考试,直到相邻三份试卷的得分误差小于10%,或达到考试组织方式所希望的次数,最终确定考生针对不同题型的实际能力。

参考文献

- [1] 中华人民共和国教育部国家语言文字工作委员会 [2018] 《中国英语能力等级量表》
- [2] 中华人民共和国教育部 [2019] 《义务教育英语课程标准（2019版）》
- [3] 中华人民共和国教育部 [2017] 《普通高中英语课程标准（2017版）》
- [4] 中华人民共和国教育部 [2000] 《高职高专教育英语课程教学基本要求（试行）》
- [5] 中华人民共和国教育部 [2004] 《大学英语课程教学要求（试行）》