

图 目

图 1.1	研究阶段与目标	10
图 2.1	阅读（认知加工）模式	19
图 2.2	二语阅读补偿模式	20
图 2.3	TOEIC 阅读测试成绩报告样例	37
图 2.4	SAT 预考批判性阅读成绩报告样例	38
图 2.5	认知属性层级关系类型	39
图 2.6	SAT 批判性阅读测试成绩报告样例	40
图 2.7	TOEFL 阅读测试成绩报告样例	46
图 3.1	研究方法总览	50
图 5.1	阅读能力（二阶）五维模型	90
图 5.2	阅读能力（二阶）六维模型	90
图 5.3	阅读能力单维模型	91
图 5.4	（二阶）六维模型拟合状况	94
图 6.1	第 36 小题的 π_i^* 参数马尔可夫链后验分布图	111
图 6.2	第 27 小题的 r_{ia}^* 参数马尔可夫链后验分布图	112
图 6.3	在 30000 预烧期时收敛缓慢的 r_{ia}^* 参数马尔可夫链	112
图 6.4	第五次参数估计后第 21、26 小题 r_{ia}^* 参数后验分布图	123
图 6.5	预测与实测成绩累积比例分布比较	126
图 6.6	正确比例分数的实测值与预测值差异比较	129
图 6.7	（依照第一次参数估计）题目掌握者与未掌握者的正确比例分数	131

图 6.8	(依照第五次参数估计) 题目掌握者与未掌握者的 正确比例分数·····	132
图 6.9	(依照第一次参数估计) 题目高水平未掌握者与低水平 未掌握者的正确比例分数·····	132
图 6.10	(依照第五次参数估计) 题目高水平未掌握者与低水平 未掌握者的正确比例分数·····	133
图 6.11	技能(属性)掌握个数与总分关系·····	136
图 7.1	英语阅读成绩报告(考生个人版)正面·····	152
图 7.2	英语阅读成绩报告(考生个人版)反面·····	153
图 7.3	英语阅读成绩报告(教师版)第 1 页·····	155
图 7.4	英语阅读成绩报告(教师版)第 2 页·····	156
图 7.5	英语阅读成绩报告(教师版)第 3 页·····	157

表 目

表 3.1	考生基本情况	52
表 3.2	教师基本情况	52
表 3.3	专家基本情况	53
表 3.4	被试基本情况	54
表 3.5	阅读测试结构	55
表 4.1	阅读测试描述性统计	65
表 4.2	阅读测试小节的难度与区分度	67
表 4.3	阅读材料（文本）基本信息	69
表 4.4	阅读试题类型	70
表 5.1	测量目标与试题对应关系	76
表 5.2	探索性和验证性 DIMTEST 维度检测结果	77
表 5.3	验证性题组效应 DIMTEST 检测结果	78
表 5.4	验证性 DETECT 题组效应检测结果	79
表 5.5	探索性 DETECT 题目聚类分析结果	80
表 5.6	第五题选项比例	81
表 5.7	第六题选项比例	82
表 5.8	试题与主要测量目标的对应关系	87
表 5.9	结构方程模型拟合度指标及标准	92
表 5.10	三种模型的拟合度比较	93
表 5.11	六维模型与修正模型拟合指标比较	95
表 6.1	Q 矩阵（依照 CFA 与专家判断结果）	100

表 6.2	有声思维编码类别	103
表 6.3	初始 Q 矩阵 (依照有声思维分析结果微调)	107
表 6.4	第一次参数估计值	114
表 6.5	第二次参数估计值	116
表 6.6	第四次参数估计值	118
表 6.7	第五次参数估计值	121
表 6.8	阅读技能掌握概率、实测分及预测分举例	124
表 6.9	正确比例分数的实测值与预测值	127
表 6.10	两次参数估计产生的“异常考生”	135
表 6.11	技能个数与得分相关性	137
表 6.12	三位“极端个案”考生的属性掌握概率值以及原始 作答情况	138
表 6.13	诊断信息指数	139
表 7.1	35 位考生的阅读技能掌握概率及总分	144
表 7.2	得 27 分的四位同学的技能掌握情况举例	147
表 7.3	阅读试题难度标定	149
表 7.4	学生作答情况举例 (以前四位同学为例)	151
表 7.5	阅读技能掌握概率认可度	159

缩略语

AHM	Attribute Hierarchy Model 属性层级模型
APA	American Psychological Association 美国心理学学会
AUA	Assessment Use Argument 测试使用论证（框架）
CAE	Cambridge English: Advanced 剑桥高级英语证书考试
CDA	Cognitive Diagnostic Assessment 认知诊断测评
CDMs	Cognitive Diagnosis Models 认知诊断模型
CFA	Confirmatory Factor Analysis 验证性因子分析
CPE	Cambridge English: Proficiency 剑桥熟练英语证书考试
CTT	Classical Test Theory 经典测量理论
DELNA	Diagnostic English Language Needs Assessment 诊断性英语需求测试
DELTA	Diagnostic English Language Tracking Assessment 诊断性英语追踪测试
DIALANG	Diagnostic Language Test 诊断性语言测试
DIF	Differential Item Functioning 项目功能差异
DINA	Deterministic Inputs, Noisy and Gate Model “确定性输入、噪音与门”模型
DINO	Deterministic Inputs, Noisy or Gate Model “确定性输入、噪音或门”模型
EFA	Exploratory Factor Analysis 探索性因子分析
ETS	Educational Testing Service 美国教育考试服务中心
FCE	Cambridge English: First 剑桥第一英语证书考试

FM	Fusion Model 融合模型
GT	Generalizability Theory 概化理论
IELTS	The International English Language Testing System 雅思考试
IRT	Item Response Theory 项目反应理论
KET	Cambridge English: Key 剑桥英语入门考试
MCMC	Markov Chain Monte Carlo 马尔可夫链蒙特卡罗 (算法)
MELAB	Michigan English Language Assessment Battery 密歇根英语考试
MFRM	Many-Facet Rasch Model 多面拉西模型
MTMM	Multi-Traits Multi-Methods 多质多法
NAEP	National Assessment of Educational Progress 美国国家教育进展评估
NC-RUM	Non-Compensatory Reparameterized Unified Model 非补偿性再参数化统一模型
PET	Cambridge English: Preliminary 剑桥初级英语考试
PPM	Posterior Probability of Mastery (属性) 后验掌握概率
PSAT/NMSQT	Preliminary Scholastic Assessment Test/ National Merit Scholarship Qualifying Test 大学入学预考 / 国家奖学金申请资格考试
PTE	Pearson Test of English 培生英语考试
RSM	Rule Space Model 规则空间模型
SAT	Scholastic Assessment/Aptitude Test 学术能力评估测试 (美国高考)
SEM	Structural Equation Modeling 结构方程模型
TOEFL iBT	Test of English as a Foreign Language Internet-Based Test 托福 (网考)
TOEIC	Test of English for International Communication 托业考试
TRT	Testlet Response Theory 题组反应理论
UM	Unified Model 统一模型

目 录

前 言	I
图 目	III
表 目	V
缩略语	VII
第一章 引言	1
1.1 研究背景	1
1.1.1 诊断信息的作用	1
1.1.2 高考英语改革的现实需求	3
1.1.3 测试效度验证的理论要求	5
1.1.4 阅读能力的诊断	7
1.2 研究目的	10
1.3 研究问题	10
1.4 本书结构	12
第二章 文献综述	14
2.1 阅读能力研究	14
2.1.1 构念与构念界定	14
2.1.2 母语阅读模式与能力	16
2.1.3 二语阅读模式与能力	18
2.1.4 二语阅读测试构念界定	21
2.1.5 二语阅读能力维度（及构念效度）的验证方法	24

2.2 认知诊断模型诞生的背景	28
2.3 认知诊断模型的选择	34
2.4 融合模型与语言测试	40
2.4.1 融合模型的数学表达	40
2.4.2 融合模型在语言测试中的应用	44
2.5 小结	47
第三章 研究方法	49
3.1 研究方法	49
3.1.1 研究参与者	51
3.1.2 研究工具	54
3.2 数据收集与分析步骤	57
3.2.1 第一阶段	57
3.2.2 第二阶段	57
3.2.3 第三阶段	59
3.2.4 第四阶段	62
3.3 小结	63
第四章 阅读测试质量分析	64
4.1 阅读测试样题基本质量分析	64
4.1.1 整份样题的质量参数	65
4.1.2 小题的质量参数	66
4.2 阅读材料特征分析	69
4.3 阅读试题特征分析	70
4.4 学生问卷结果分析	71
4.5 教师问卷结果分析	72
4.6 小结	72
第五章 阅读能力维度分析	74
5.1 命题者视角	75

5.2 试题单维性分析	77
5.3 试题维度数量及聚类分析	78
5.4 学生问卷结果分析	80
5.5 专家判断结果分析	86
5.6 验证性因子分析	89
5.7 小结	97
第六章 诊断信息的提取	98
6.1 Q 矩阵构建	99
6.1.1 CFA、专家判断结果与 Q 矩阵构建	100
6.1.2 有声思维分析结果与 Q 矩阵构建	102
6.2 马尔可夫链收敛检验	108
6.3 融合模型参数估计与检验	112
6.4 融合模型拟合优度检验	123
6.4.1 模型预测得分与实测得分比较	123
6.4.2 实测正确比例分数与估计正确比例分数	127
6.4.3 技能掌握者和未掌握者对题目作答反应情况比较	129
6.4.4 阅读技能掌握情况和阅读成绩之间的关系	136
6.5 “极端个案”分析与阅读技能（属性）的“诊断力”	137
6.6 小结	141
第七章 诊断信息的呈现	143
7.1 成绩报告的内容与参照依据	144
7.1.1 诊断信息的参照依据	144
7.1.2 试题难度的参照依据	147
7.1.3 其他内容的参照依据	150
7.2 成绩报告的制作	151
7.2.1 考生个人成绩报告	151
7.2.2 任课教师成绩报告	154
7.3 成绩报告的评价	158

7.3.1 问卷分析	158
7.3.2 讨论分析	159
7.3.3 任课教师对成绩报告（考生个人、教师版）的意见	162
7.4 小结	163
第八章 结论	164
8.1 研究发现	164
8.1.1 阅读样题质量	164
8.1.2 阅读能力维度	165
8.1.3 诊断信息提取	166
8.1.4 诊断信息呈现	168
8.2 研究启示	168
8.2.1 诊断信息提取的新思路	168
8.2.2 高考成绩报告的新形式	169
8.2.3 正面后果证据的新途径	169
8.2.4 构念效度验证的新方式	170
8.2.5 阅读能力维度的新视角	171
8.3 研究局限与展望	172
8.4 结语	174
参引文献	176
附录	198
附录 1 英语阅读测试样题	198
附录 2 阅读试题评价问卷（学生卷）	211
附录 3 阅读试题评价问卷（教师卷）	214
附录 4 阅读试题评价问卷（专家卷）	216
附录 5 有声思维实验研究工具	218
附录 6 英语阅读成绩报告（教师版）第 4、5 页	223
附录 7 英语阅读测试成绩报告评价问卷	225