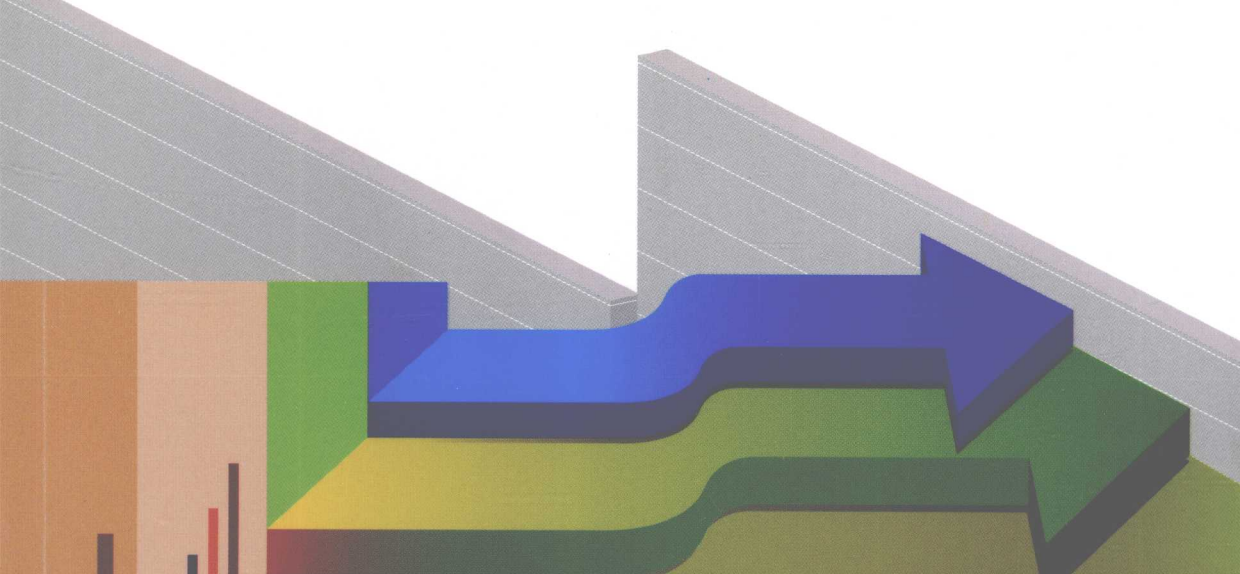


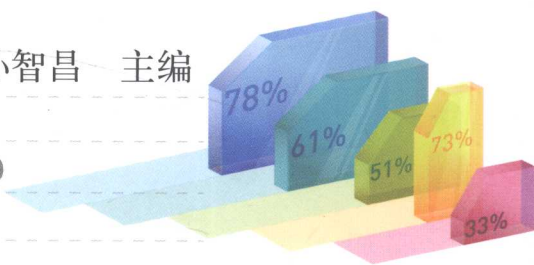


中国小学生学业成就 测评报告与测试工具

——以小学六年级四门学科为例

田慧生 孙智昌 主编

ZHONGGUO XIAOXUESHENG
XUEYE CHENGJIU CEPING BAOGAO
YU CESHI GONGJU



教育科学出版社

Educational Science Publishing House

中国小学生学业成就 测评报告与测试工具

——以小学六年级四门学科为例

田慧生 孙智昌 主编

教育科学出版社

· 北京 ·

出版人 所广一
责任编辑 谭文明
版式设计 杨玲玲
责任校对 贾静芳
责任印制 曲凤玲

图书在版编目 (CIP) 数据

中国小学生学业成就测评报告与测试工具: 以小学
六年级四门学科为例 / 田慧生, 孙智昌主编. —北京:
教育科学出版社, 2012. 6

ISBN 978 - 7 - 5041 - 6618 - 0

I. ①中… II. ①田… ②孙… III. ①小学生—学业
评定—研究 IV. ①G622. 47

中国版本图书馆 CIP 数据核字 (2012) 第 124505 号

中国小学生学业成就测评报告与测试工具

ZHONGGUO XIAOXUESHENG XUEYE CHENGJIU CEPING BAOGAO YU CESHI GONGJU

出版发行 教育科学出版社

社 址 北京·朝阳区安慧北里安园甲 9 号

市场部电话 010 - 64989009

邮 编 100101

编辑部电话 010 - 64981277

传 真 010 - 64891796

网 址 <http://www.esph.com.cn>

经 销 各地新华书店

制 作 北京金奥都图文制作中心

印 刷 北京中科印刷有限公司

版 次 2012 年 6 月第 1 版

开 本 169 毫米×239 毫米 16 开

印 次 2012 年 6 月第 1 次印刷

印 张 20.75

印 数 1 - 5 200 册

字 数 311 千

定 价 45.00 元

如有印装质量问题, 请到所购图书销售部门联系调换。

前言

学生学业成就是指在教师的指导下，学生在先前经验的基础上通过学习活动在知识、技能以及情感态度等方面达到的发展水平。在世界各国普遍重视并探索和进行宏观监控教育质量的背景下，在我国实现“两基”，全面推进素质教育，积极推进课程改革及已基本解决办学条件和经费投入等问题的前提下，开展学生学业成就调查研究具有非常重要的理论意义和实践价值。一方面，它可以在国家宏观监控我国学生的学业成就现状，便于我们更加全面、客观、真实地了解 and 把握我国基础教育质量的情况，从而真正有效地推进素质教育的落实，为教育行政部门决策和管理提供切实可靠的依据，促进教育公平和教育均衡发展目标的实现。另一方面，它也可以为课程政策的调整、课程标准的制定以及课程实施等提供科学依据，从而发挥评价促进学生发展、教师提高和改进教学实践的功能，切实推动基础教育改革的深入和教育质量的提高。

本调查的对象为小学六年级学生，调查科目是语文、数学、科学、品德与社会四个学科，主要目的是了解学生在语文阅读、数学、科学、品德与社会方面的学业成就，并配合学生问卷和学校问卷来了解影响学生学业成就的相关因素。

本调查采用标准参照评价模式,以全日制义务教育课程标准(实验稿)(以下简称课程标准)为研究依据,从学科内容和学科能力两个维度确定各学科评价的基本框架,形成双向细目表,并根据 SOLO (Structure of the Observed Learning Outcome) 分类理论关于学习结果的分类编制各学科测试工具。测试工具的开发经历了准备、编制、修改、预试、再修改及最后定稿的过程。整个测试过程建立了严密的培训、组织和管理体系,保证了测试结果的科学性。

本调查采用分层随机抽样的方法从全国东中西部 8 个省共 31 个区县抽取 18000 多名小学六年级学生,对其在语文阅读、数学、科学、品德与社会四门学科的学业成就状况进行了调查,考查其学业质量达到课程标准的情况,并深入分析影响学生学业质量的原因,以为教育行政部门的决策提供科学依据,为学科教学和学生提供反馈信息,进而改进和提高学科教学质量,促进学生进行自我导向、自我监控的学习。

本研究是课题团结合作的成果。课题主持人田慧生研究员自始至终组织和领导了课题的研究,课题组副组长兼秘书长孙智昌协助课题主持人开展研究,课题组核心成员有陈琴、刘芳、蔡永红、任春荣、胡军、陈晓东、杨莉娟、张鹏举、马晓强、马延伟、杨宝山、江明、冯新瑞、王晓霞、李嘉骏等。本报告撰写者分工如下:前言、学生学业成就测评总报告由田慧生、陈琴撰写;语文阅读测评报告由张鹏举、王晓霞撰写;数学测评报告由陈晓东撰写;科学测评报告由胡军、杨宝山撰写;品德与社会测评报告由杨莉娟撰写;学业成就影响因素调查报告由任春荣撰写;测试工具由各学科组和总课题组反复研究完成。报告整体由田慧生、孙智昌、陈琴负责统稿。尽管我们做了很大努力,但错漏和不足之处在所难免,敬请各位方家批评指正。

本报告的出版得到教育科学出版社的大力支持,编校工作者的严谨认真使本报告生色不少,在此一并表示感谢!

Contents 目 录

前言

第一部分 学生学业成就测评总报告 / 1

- 一、学业成就测评——深化基础教育改革，提高教育质量的客观需要 / 1
- 二、本学业成就测评的研究依据与主要内容 / 3
- 三、研究方法 with 工具 / 4
- 四、学生学业成就测评研究的实施过程 / 8
- 五、本学业成就测评的科学性分析 / 9
- 六、学生各科学业成就状况的分析 / 11
- 七、影响学生学业成就的因素分析 / 20
- 八、研究结论 / 54
- 九、对策与建议 / 55

第二部分 学科测评报告 / 58

语文阅读测评报告 / 58

- 一、研究背景 / 58

- 二、研究的理论框架 / 59
- 三、研究方法 with 工具 / 63
- 四、阅读测试结果分析与讨论 / 74
- 五、主要发现 / 84
- 六、对策与建议 / 87

数学测评报告 / 89

- 一、研究背景 / 89
- 二、研究的理论框架 / 89
- 三、研究方法 with 工具 / 101
- 四、测试结果分析与讨论 / 113
- 五、主要发现 / 126
- 六、对策与建议 / 128

科学测评报告 / 130

- 一、研究背景 / 130
- 二、研究的理论依据和主要分析 / 141
- 三、研究目的和方法 / 147
- 四、测试结果分析与讨论 / 167
- 五、主要发现 / 181
- 六、对策与建议 / 184

品德与社会测评报告 / 186

- 一、研究背景 / 186
- 二、关于学生社会性发展评价的理论架构 / 187
- 三、评价工具的研制与测试 / 195
- 四、学生社会性发展的现状分析与讨论 / 219
- 五、主要发现 / 232
- 六、对策与建议 / 235

第三部分 学业成就影响因素调查报告 / 239

- 一、前言 / 239
- 二、结果与发现 / 240
- 三、结论 / 285
- 四、调查中发现的问题 / 286

第四部分 学科测评工具 / 289

- 一、语文学科样题 / 289
- 二、数学学科样题 / 295
- 三、科学学科样题 / 300
- 四、品德与社会学科样题 / 307
- 五、学生问卷 / 315
- 六、学校问卷 / 323

第一部分

学生学业成就测评总报告

一、学业成就测评——深化基础教育改革，提高教育质量的客观需要

近年来，随着各国政府教育质量意识的提升，以科学的研究和测评体系获得教育质量及学生发展状况真实、全面的评价信息，更好地为教育改革提供决策基础，为学生个性化发展、可持续发展提供依据已成为当今世界教育改革发展的—大趋势。对教育质量即教育水平高低和效果优劣的程度的考查主要集中在学习者的发展水平层面，而学业成就是反映其发展水平的核心指标。

学业成就是指学生在教师指导下，在先前经验的基础上，通过学习活动在知识、技能以及情感态度等方面达到的发展水平。

有关学生学业成就的调查与测评，国内外都进行了大量的理论研究和实践探索。目前在国际上比较有影响的学生学业成就调查研究主要是国际教育成就评价协会（The International Association for the Evaluation of Educational Achievement，简称 IEA）和经济合作与发展组织（Organization for Economic Cooperation and Development，简称 OECD）这两个国际组织开展的国际比较研究。IEA 主持进行的国际数学与科学教育成就趋势调查（Trends in International Mathematics and Science Study，简称 TIMSS）及其重复研究（又称录像研究、后续调查或再研究，简称 TIMSS-R）项目主要考查的是学生在科学和

数学学科方面的学业能力。测评的对象包括三类，第一类是9岁的学生（三年级或四年级），第二类是13岁的学生（七年级或八年级），第三类是中学最后一个年级的学生。其中，TIMSS-R主要研究第二类学生群体。而OECD主持进行的国际学生评价项目（Programme for International Student Assessment，简称PISA）则主要侧重于学生成年后的实际生活能力，包括阅读素养、数学素养和科学素养。测评的对象是15岁的学生。这两项研究因其评价维度、指标和工具的科学性与严密以及评价程序的规范等，在国际上产生了很大的影响。除此之外，还有不少国家自己定期组织进行各学科的学生学业成就调查，建立国家常模，开展纵向和横向的比较研究，如美国的全国教育进展评价调查（The National Assessment of Educational Progress，简称NAEP），英国的学生成绩评估（The Assessment of Performance Unit，简称APU）调查，日本的国研调查，澳大利亚、新西兰、瑞士等国周期性国家常模修订等。这些研究在了解学生学业成就状况，通过评价来引导学生学习方面积累了丰富的经验。

我国由于种种原因一直没有整体参与IEA和OECD等国际评价组织开展的学生学业成就调查研究。国家和地区层面尽管有高考和中考，但它们发挥的功能主要是甄别和选拔。从20世纪90年代开始，我国开始在学生学业成就的调查方面有了一些进展。原国家教委基础教育司与联合国儿童基金会、教科文组织联合开展的8省市抽样调查，涉及四年级和六年级小学生共2.4万名、近1300所小学和6000多名教师，是我国第一次对小学教学质量进行的宏观监测^①。近年来，北京市教育科学研究院承担的教育部“十五”重点课题“小学生学业成就评价改革研究”以及北京市教育委员会委托的“北京市义务教育教学质量监控与评价系统”项目，上海市黄浦区教育局承担的教育部“十五”重点课题“中小学实施素质教育中的学业管理和评价研究”，教育部基础教育司承担的联合国儿基会项目“东亚太平洋地区学生学业评价研究”以及北京师范大学的教育部基础教育监测中心进行“基础教育质量监测”项目等都在学生学业成就评价方面进行了探索和研究，但这些研究或缺失后续延伸，或侧重微观层面，或正在进行之中，距离建立反映义务教育学

^① 沈白愉，孟鸿伟．我国小学生学习质量现状[J]．云南教育，1996（5）：20．

生学业成就状况的国家常模和数据库，还有很大距离。

综上所述，目前我国缺乏国家层面对学业成就的调查研究和监测；现有的学业成就评价还主要是“为考而学，为考而教”；心理与教育测量理论在学生学业评价中的运用不充分，评价工具的编制多凭经验，理论性、规范性欠缺^①；评价多停留于现状调查，对如何促进教学改进和学生发展关注少。因此，在世界各国普遍重视并探索和进行宏观教育质量监控的背景下，在我国实现“两基”，全面推进素质教育，积极推进课程改革及已基本解决办学条件和经费投入等问题的前提下，开展学生学业成就的调查研究，研制出体现本土学生发展状况的相对完善的测评工具，进行全国基础教育阶段学生学业成就调查，分析影响学生学业发展的因素，具有非常重要的理论意义和实践价值。

一方面，学生学业成就测评可以填补我国在国家宏观层面学业成就调查数据的空白。基础教育质量作为提升国家竞争力的决定性因素，是一个国家综合国力的重要标志之一。全面、客观、真实地了解和把握我国基础教育的质量，可以真正有效地推进素质教育的落实，为教育行政与管理部门决策和管理提供切实可靠的依据，促进教育公平和教育均衡发展目标的实现。另一方面，它可以促进教育教学的改进和学生的发展。通过大规模的学生学业成就调查可以对学生的发展水平、课程目标的达成度、课程教学等作出评估，从而为课程政策的调整、课程标准的制定以及课程实施等提供科学依据，真正发挥评价促进学生发展、教师提高和改进教学实践的功能，切实地推动基础教育改革的深入，提高教育质量。

二、本学业成就测评的研究依据与主要内容

对于学生学业成就水平的评价主要有两种模式：常模参照评价和标准参照评价。常模参照评价是联系先前其他学生在该测验上的表现来解释现在学生的成绩，先前的那些学生被叫做常模组，其对分数的解释是相对的；标准

^① 辛涛：新课程背景下的学业评价：测量理论的价值[J]。北京师范大学学报：社会科学版，2006（1）：56。

参照评价依赖测验内容对某一评价领域的代表程度作出分数解释，评价不涉及他人，因此又称为绝对评价。本学业成就调查研究主要采用标准参照评价模式，以课程标准为研究依据，对全国范围内不同地域、城乡、性别的小学生的学业成就状况进行调查，考查其学业质量达到课程标准的情况，并深入分析影响学生学业质量的原因，以期为教育行政部门的决策提供科学依据，为学科教学和学生学习提供反馈信息，进而改进和提高学科教学质量，促进学生进行自我导向、自我监控的学习。

本学业成就调查研究的主要内容是：（1）研制适合我国国情的小学六年级学生语文阅读、数学、科学、品德与社会四个学科学业成就评价的指标体系和测评工具；（2）在全国范围内进行小学六年级学生语文阅读、数学、科学、品德与社会学科学业成就水平的调查研究，建立小学生学业成就的数据库，为最终形成小学生学业成就评价的国家常模奠定基础；（3）对影响小学六年级学生学业成就的相关因素进行分析，为学科教学和学生学习提供反馈，促进教学改进和学生学习的提高。

三、研究方法 with 工具

（一）各学科评价框架的确定

本学业成就调查以各学科课程标准中对学生在学科内容和学科能力两个维度的发展要求作为依据来确定评价的基本框架。其中，学科内容是指课程标准中明确规定学生应掌握的各学科的基本内容，学科能力是课程标准中要求学生达到的掌握水平。各学科在内容和能力领域的评价框架见表 1-1。

表 1-1 小学六年级各学科评价的基本框架

学 科	内容维度	能力维度
语文	字词	认读
	句段	理解
	篇章	运用

续表

学 科	内容维度	能力维度
数 学	数与代数 空间与图形 统计与概率 综合运用	知识技能 数学思考 解决问题
科 学	生命世界 物质世界 地球宇宙世界	呈现 应用 探究
品 德	公民与社会 健康与安全 历史与文化 地理与环境	认知 理解 运用

(二) 测评的技术路线——SOLO 分类理论和方法

本学业成就调查采用 SOLO 分类理论作为各学科测评的基本技术路线，以此研制和开发各学科测评的指标体系和测试题。

SOLO 分类理论依据学生对于某一学科问题的回答所表现出的思维结构的复杂程度不同，将学生的表现由低到高划分为前结构水平、单一结构水平、多元结构水平、关联结构水平和拓展抽象结构水平五个结构层次。

1. 前结构水平 (Pre-structural Level, P)

学生错误地理解问题，缺乏回答问题所需的简单知识，为以前所学的无关知识所困扰，关注问题中某些偶然的不相关的信息，回答问题逻辑混乱或同义重复。

2. 单一结构水平 (Uni-structural Level, U)

学生关注主题或问题，但只使用一个相关的线索或资料，找到一个线索就立即跳到结论上。

3. 多元结构水平 (Multi-structural Level, M)

学生使用两个或多个线索或资料,却不能觉察到这些线索或资料之间的联系,不能对线索或资料进行整合。表现为回答问题时能联系多个孤立事件,但却缺乏有机整合的能力,常常给出一些支离破碎的信息。

4. 关联结构水平 (Relational Level, R)

学生能够使用所有可获得的线索或资料,并将它们编入总体的联系框架中,总体成为在已知系统中内在一致的结构。表现为能够联想多个事件,并将多个事件联系起来回答或解决较为复杂的具体问题。

5. 拓展抽象结构水平 (Extended Abstract Level, EA)

学生超越资料进入一种新的推理方式,并能概括一些抽象特征。表现为会归纳问题,在归纳中概括考虑了新的和更抽象的特征;结论具有开放性且更抽象。

处于小学阶段的学生的思维结构主要表现为单一、多元和关联三种结构水平。各个学科设计和编制反映这三种结构水平的测试题对学生的学业成就状况进行检测,从而了解和获得学生在该学科上的学业成就状况。

(三) 双向细目表的编制

根据评估框架,各学科从本学科课程标准的内容领域中,筛选出重点的内容目标,结合课程标准中的能力要求将内容目标转化为学生学习的表现期望,并根据 SOLO 分类法标出其所属的能力类别,形成双向细目表(见各学科报告),编制测试题目。依据 SOLO 分类法的要求,各学科双向细目表中单一(U)、多元(M)、相关(R)结构水平题目的比例是 2:2:1。

(四) 测试工具的开发

本学业成就调查测试工具的开发经历了准备、编制、修改、预试、再修改及最后定稿的过程。



明确研究目的和研究方法，学习 SOLO 分类法，查阅 IEA、PISA 等国内外相关项目的研究成果，研究新课程改革的理念和各学科课程标准，开展前期讨论交流。



依据课程标准的内容要求和 SOLO 分类法，编制小学六年级学生在各学科的能力结构框架，包括涉及的内容领域，课程标准的要求，对学生的能力要求，以及按一定比例设计相应的单一、多元和关联结构水平题目。



组织专家进行讨论，对认知结构框架中的内容和估计难度结构设计进行调整和修改，对一些较难把握的题目，根据学生试测的结果进行修改完善，并对题目的顺序、题目中选择答案的排序进行调整。



在北京市朝阳区选取若干所小学进行预试，根据预试结果，对认知结构框架中的内容进行再次修改、筛选，删掉学生不易理解和难度过大的题目；将 1.3 倍的题量减至适量，题目结构水平作相应调整。正确答案的排列顺序调整合理，完善评分标准。



形成最终小学六年级学生四门学科测试的认知结构框架，测试题目，评分标准，订正排版印制中的问题，并作好正式测验前的一切准备。

本学业成就调查测试工具的开发遵循以下原则：

1. 遵循学生的认知规律和年龄特点；
2. 全面考虑课程标准中各个领域的内容，选择各领域中的重点要求；
3. 题干陈述的语言简洁、明了，附图清晰，尽量使所有水平的学生能够读懂题目要求，选项答案文字精准且数量相当；
4. 合理设计题目的结构水平、比例、题量、用时、评分标准。

(五) 测试工具的内容

本学业成就调查的工具主要包括语文、数学、科学、品德与社会四门学

科试卷、学生问卷和学校问卷。其中，语文试卷 22 题，数学试卷 37 题，科学试卷 42 题，品德与社会试卷 42 题。学生问卷 27 大题共 132 小题，学校问卷共 30 题。语文和数学的测试时间为 60 分钟，科学和品德为 50 分钟。学生问卷和学校问卷的填写时间分别为 40 分钟和 20 分钟。

四、学生学业成就测评研究的实施过程

（一）研究对象的选取

本学业成就调查采用分层随机抽样的方法，从全国东中西部 8 个省共 31 个区县中抽取了 18600 名小学六年级学生，最终参与测试的学生为 18226 人。其中，城市学生占 68.5%，农村学生占 31.5%；男生占 46.6%，女生占 53.4%。每名学生都参加语文、数学、科学、品德与社会四个学科的测试。学生问卷的测试对象为所有参与学科测试的学生，学校问卷的测试对象为所有参与学科测试的学校，共 372 所。本学业成就调查的施测时间为 2009 年 5 月。

（二）施测程序

本学业成就调查的测试建立了严密的培训、组织和管理体系。

1. 两级测试培训体系

首先，由项目组专家进行项目测试实施的全国培训。培训的具体内容包括：语文、数学、科学、品德与社会的测试目的，出题思路与测试框架；测试组织工作要求；测试样本的构成及学生编码的形成；测试实施过程中的操作细节与问题处理；测试问卷的回收与寄送；个别地区测试问题答疑。其次，由各区县级组织对参与测试学校及参与者的测试组织工作进行培训，培训内容包括：测试组织工作的要求、学生编码方案、监考安排及要求等内容。

2. 严密的组织和管理体系

制定了包括工作准备、正式测试、测试后的核查、整理和汇总等一系列工作的组织和要求，以保证试卷的有效回收，确保测试的真实、科学。

(三) 数据统计与分析

本学业成就调查运用项目反应理论 (Item Response Theory, 简称 IRT) 计算试卷参数和学生能力参数。在项目反应理论指导下的整卷层面的质量分析以单维单参数计分模型 (简称 Rasch 模型) 为基础, 对测试的质量进行分析。试卷分析的指标主要包括信度和效度、题目难度、区分度、题目特征参数等。学生学业成就的分析主要是根据学生的能力参数对学生的总体能力水平、不同学科的学业成就状况、不同地域、城乡和性别学生的学业成就差异等。

五、本学业成就测评的科学性分析

(一) 测试问卷的回收率

本次调查共发放语文、数学、科学、品德与社会四科试卷以及学生问卷各 18226 份, 学校问卷 372 份。四科试卷的回收率分别为 92.2%、98.2%、89.9% 和 98.9%, 学生问卷的回收率为 94.6%, 学校问卷的回收率为 83.3%。

表 1-2 测试问卷的回收率 (%)

	语 文	数 学	科 学	品 德	学生问卷	学校问卷
回收数	16799	17898	16378	18024	17244	310
回收率	92.2	98.2	89.9	98.9	94.6	83.3

(二) 各科试卷的信效度、题目难度、区分度和题目特征参数分析

本次调查结果表明, 语文、数学、科学、品德与社会四门学科试卷都具有较好的信度和效度。其中, 四门学科的信度均在 0.80 以上; 拟合指数的平均值均为 0.99 或 1.00, 表明试卷的内容结构效度都很高, 符合 Rasch 模型能力单维性的假设; 各学科之间的相关系数也达到了非常显著的水平 (见表 1-4), 表明各科试卷的测试结果反映的是学生的同一种能力。