

# 基于课程标准的数学学科能力评价研究

## ——以某学区七年级测试工具开发及实施为例

郭 衍, 曹 鹏, 杨 凡, 刘金花

(北京师范大学 数学科学学院, 北京 100875)

**摘要:** 基于学科内容和学科能力的二维评价工具不仅是国际研究的主流趋势, 也是中国教育改革与发展的实际需求. 以7年级测试工具的开发和实施为例, 介绍基于课程标准的数学学科能力评价的二维框架(内容维度及能力维度)、评价过程(命题组卷、访谈测试、小规模预测及正式测试)、技术指标(难度、区分度、信度、效度等)及结果反馈形式(学区、学校、学生等). 总结了学科能力测试在内容参照、知识选取、能力导向、命题过程和结果呈现等方面有异于一般的学业考试和能力测试.

**关键词:** 数学学科能力; 教育评价; 内容维度; 能力维度

**中图分类号:** G40-03 **文献标识码:** A **文章编号:** 1004-9894(2015)02-0017-05

### 1 前 言

教育的目的是促进学生的发展, 而能力发展是学生发展的主要标志与核心内容. 具体到学科教学, 自然就需要发展学生的学科能力. 林崇德指出: 学科能力是能力与特定学科的有机结合, 是能力在特定学科中的具体体现<sup>[1]</sup>. 美国数学教师协会早在1989年提出了数学学科核心能力的思想, 并在2000年发布的《学校数学教育的原则和标准》中提出了5大数学能力<sup>[2]</sup>, 2010年《美国统一州核心课程标准》也提出: 理解、推理、质疑、建模等8项能力目标, 并鼓励以此开发统一的全方位的学生表现评价体系<sup>[3-4]</sup>; 澳大利亚的数学课标在具体内容标准描述之前都会对所应达到的能力水平做明确说明, 包括: 理解、熟练性、问题解决、推理<sup>[5-6]</sup>; 德国学科能力测试也基于其国家教育标准<sup>[7-8]</sup>和能力表现水平的二维框架<sup>[9]</sup>. 此外, 国际知名大型学生测评项目 PISA (Program for International Student Assessment) 和 TIMSS (Trends in International Mathematics and Science Study) 也都不约而同的使用了学科内容和学科能力(认知、素养)的测试框架<sup>[10]</sup>. 环顾国际先进实践经验不难看出, 以学科内容和学科能力构建二维模型已成为学业水平评价的主要方向.

2013年6月, 教育部发文推进中小学教育质量综合评价改革, 指出学业发展水平应主要考查学生对各学科课程标准所要求内容的掌握情况, 并从知识技能、思想方法、实践能力、创新意识等关键性指标进行评价. 同年11月, 十八届三中全会《决定》提到深化教育领域综合改革, 推行初高中学业水平考试和综合素质评价, 探索招考相对分离. 由此可见, 设计科学、实施客观、结构多维的学生评价不仅仅是国际主流趋势, 更是中国深化教育改革和推进教育发展的实际需求.

然而现实中针对具体学科的学生评价理论和技术发展缓慢<sup>[11]</sup>, 实践层面则更是缺乏理论指导和技术支持, 测试的命题和评价工作大多依靠主导教师的经验, 试题质量和评价

结果也基本没有经过科学检验及分析. 通过以某学区7年级数学学科能力测试工具的开发及实施为例, 探索更为科学合理的评价测试方法, 希望为数学教育评价研究与实践工作提供的参考与借鉴.

### 2 测试结构

#### 2.1 测试形式

常见的评价测试形式有: 在线测试(或计算机辅助测试)、纸笔作答和访谈观察. 访谈观察法适用于一对一或小规模评价, 通过实地观察和互动交流能够全方面地了解教学效果并给出评价分析, 多用于课堂教学效果评价和学生案例研究<sup>[12]</sup>; 计算机辅助测试将题库、组卷、答卷、评分、记录等环节信息化处理, 使得测试过程更加方便和科学. 但考虑到测试学区的客观情况和测试规模, 这里还是采用最常见的纸笔测试形式. 设计测试时间为90分钟, 学生在答题卡上完成答卷.

在日常教学的考试中, 教师经常采用选择型测试(如选择题、判断题), 因为这种题型耗时较短、易于批改且评价结果方便量化记录和表达. 但通常教师自编的选择型题目往往集中于对知识点记忆与理解的考查, 缺乏推理过程, 而且学生作答可能存在猜测成分或“选项技巧”, 对学生思维过程和真实学习效果的考查效果不佳.

测试采用的题型为填空题和解答题, 不使用选择题. 试题数量为18道, 其中填空题10道, 解答题8道, 部分填空题和解答题有多个小空(问), 实际测试题量为28小题.

#### 2.2 内容维度

测试内容的选择与确定是测试工具的开发需要考虑的最基本问题. 测试以《义务教育数学课程标准(2011年版)》(以下简称“课程标准”)为依据, 兼顾测试学区中考大纲和教材.“课程标准”将教学内容划分为数与代数、图形与几何、统计与概率、综合与实践, 从测试内容来说, 应从数与代数、图形与几何、统计与概率3个方面进行考查. 总体设

收稿日期: 2014-12-26

基金项目: 国家自然科学基金——用于多级评分认知诊断的统计模型研究(11171029)

作者简介: 郭衍(1988—), 男, 江苏扬州人, 博士研究生, 主要从事数学教育及教育测量与评价研究.

计选择了函数、方程与不等式、图形的性质、统计与概率这4个模块重点考查,由于该学区7年级采用人教版新教材,并未学习一次函数相关内容,所以只考查方程与不等式、图形的性质、统计与概率(详见表1)。

表1 数学能力测试内容维度具体内容

一级指标	二级指标	知识点	
方程与不等式	一元一次方程	一元一次方程的概念	
		一元一次方程求解	
		一元一次方程应用题	
		解降次化为一元一次方程的一元三次方程	
	二元一次方程(组)	二元一次方程(组)的概念	
		二元一次方程组求解	
		列方程解应用题	
	一元一次不等式(组)	消元思想	
		一元一次不等式的概念	
		不等式的性质	
		一元一次不等式(组)的解法	
		不等式组解集的数轴表示	
	图形的性质	点、线	不等式组的特殊解
			一元一次不等式的实际应用
			点动成线、线动成面、面动成体
角		两点之间线段最短	
		两点确定一条直线	
		(邻)补角、(邻)余角	
相交线与平行线		对顶角	
		角平分线	
		垂线	
		三线八角	
三角形		平行线判定	
		平行线的性质	
		三角形三边关系	
		三角形内角和	
		多边形及其内角和	
数据的收集整理与描述	三角形外角性质		
	三角形三线		
	数据收集方式的选取		
	相关概念(样本容量、用样本估计总体)		
统计与概率	频数分布直方图		

### 2.3 能力维度

能力维度主要包括学习理解、实践应用、创造迁移3个一级指标,每个一级指标下又分为3个二级指标(详见表2)。

**学习理解能力**。指学生在数学学习过程中:能够及时、顺利、准确提取的数学知识;能够用多种表征方式概括数学知识,并在多种表征方式之间进行自由的转换;能够建立数学知识之间的横向和纵向的联系,发展、完善知识网络;能够顺利完成获得数学知识的数学推理。

**实践应用能力**。指学生在数学学习过程中:能够在数学情境中把实际信息与相关知识进行对应,根据数学的规律和原理完成数学运算类任务和作图类任务;能够结合实际情境把实际信息与相关知识进行对应,利用数学知识解决现实的问题,并建立数学模型解决实际问题。

**创新迁移能力**。指学生在数学学习过程中:能够对新颖的数学方法有创造性的体会和认识;能够展开想象,积极猜想,进行合情推理;对数学学习过程中的不同的解决问题方法进行批判性思考,并给出个人的评价;能够从已有知识和

技能出发,通过合情推理主动建立相关知识之间的联系,以及对未知的问题进行深入地探索。

表2 数学能力测试能力维度具体内容

一级指标	二级指标
A 学习理解	A-1 观察记忆
	A-2 概括理解
	A-3 说明论证
B 实践应用	B-1 分析计算
	B-2 推理解释
	B-3 简单问题解决
C 创造迁移	C-1 综合问题解决
	C-2 猜想探究
	C-3 发现创新

“课程标准”用4个行为动词对结果目标进行描述,分别为:了解、理解、掌握、运用,在该学区的中考大纲也使用了类似的行为动词对考试内容进行能力划分,但大部分知识点的要求低于课标中的表述。另外,4个行为动词难以清晰地界定开,这里使用的“三分法”相比课标的“四分法”更加易于操作。

## 3 测试与评分

### 3.1 命题与组卷

内容维度和能力维度的交叉形成了二维的命题框架(双向细目表),两者的交点标定了题目的内容特征及能力要求,从而为编制或筛选题目提供依据。

首先是进行内容维度和能力维度的讲解培训,然后向数学教育硕士、一线教师、教研员广泛征集题目,并通过命题小组多次筛选、讨论、协商组成一套1.5~2倍于正式测试题目的预测试卷,通过访谈测试和预测试删减、调整题目,最终形成正式测试卷。

### 3.2 访谈测试

访谈测试又被称为“出声思维”,是让做题者口述做题时的想法、思路及解题过程,通常每道题需要访问6名考生(2性别×好中差3类),以充分了解每类学生对于试题的理解是否符合出题者的意图,试题表述是否存在歧义等。

### 3.3 小规模预测

同样选取比较有代表性的考生通过类似正式测试的纸笔作答形式作为预测。由于预测试卷比正式测试的题量大,同时也为使每道题都获得完整的反馈信息,所以小规模预测的时间较长,保证每个学生都能充分作答所有题。操作时,可以将试卷做均衡处理(分AB卷,调整题目顺序,防止排在后面的题目学生没有认真作答),在条件允许的情况下也可以适当增加预测试考生人数。

通过小规模预测,可估计题目的难度、区分度等,筛选题目组成最终测试卷,并进一步优化试题表述及评价标准。

### 3.4 正式测试

正式测试采用分层随机抽样的方法,根据该学区提供的学校分组名单,从3个组中随机抽取3~4所学校,共11所学校,在学校内部再采用随机抽样的方法分年级从每个班级中抽取5~6名学生参加正式测试。

### 3.5 评分

在填空题评分中,若考生作答与答案相同,则该空记1分,

否则为 0 分。在解答题评分中，采用多级评分法分步给分。

最终采用经典测量理论 (CTT) 和项目反应理论 (IRT) 两种方法计算参测者得分。CTT 得分主要用来检测试题的难度、区分度、信效度等指标；而 IRT 得分则用来呈现参测学生在各方面的能力表现。

#### 4 技术指标

技术指标主要包含两大类：一类是针对题目的指标，包括难度、区分度，另一类是测试总体的指标，包括信度和效度。

##### 4.1 难度

试题难度常常被定义为通过率或得分率，该定义下难度值越大，实际难度越低。

就填空题而言，只区分对错，难度值用答对人数的百分比表示：

$$P = \frac{R}{N}$$

其中， $N$  为全体考生人数， $R$  为答对该题的人数。

就解答题而言，评分并非只有对、错两种结果，而是从 0 分至满分的多种可能，应以下式计算难度值：

$$P = \frac{\bar{x}}{x_{\max}}$$

其中， $\bar{x}$  为全体考生在某题上的平均得分， $x_{\max}$  为该题的满分。

难度值参考范围： $P < 0.3$  难； $0.3 < P < 0.6$  偏难； $0.6 < P < 0.75$  中等难度； $0.75 < P < 0.85$  偏易； $P > 0.85$  容易。经计算，试题难度分布合理，且每类题的前两题难度值均在 0.7 以上，便于考生做题时快速进入状态。

##### 4.2 区分度

常用的区分度计算方法包括鉴别指数、点二列相关系数、二列相关系数、积差相关系数等，测试使用鉴别指数 ( $D$ ) 来计算区分度：

$$D = P_H - P_L$$

其中， $P_H$  和  $P_L$  为高分组和低分组在该题目上的得分率。一般情况下，根据测试总分将测试者排队，取前 27% 为高分组，取后 27% 为低分组。但当测试人数过少（少于 100 人）时，则不宜使用 27%，而应使用 50% 为分界点。

鉴别指数值参考范围<sup>[13]</sup>： $D > 0.4$  很好； $0.30 < D < 0.39$  良好； $0.20 < D < 0.29$  尚可，需要修改； $D < 0.2$  差，淘汰。经计算，试题区分度均在 0.2 以上，除个别小问外，大部分在 0.4 以上。

##### 4.3 信度

信度是衡量测试质量高低的重要指标。简单的说，信度就是测试结果稳定性的程度，即（不同的施测者使用）某测量工具反复测量考生的某一特质，结果间的一致性。

在测试中，对信度的考查包括评分者信度和测试的整体信度。

###### 4.3.1 评分者信度

评分者信度 (score reliability) 是指多个评分者给同批答卷进行评分的一致性程度。测试的解答题部分需要学生书写解答过程，评分者根据解答情况分步给分，因此，对这类

题目评分的准确把握是对试卷信度的保障。

阅卷组由 20 名数学教育方向的研究生及一线教师组成。首先，对阅卷组成员进行培训，保证每个阅卷者都能准确理解评分标准。其次，进行集体出声预评，抽取 10 份试卷，阅卷组集体评分，并对评分标准理解不一致的地方进行讨论，保证每个阅卷者对分步评分的把控一致。最后，再抽取 10 份试卷，阅卷者两两分组评分后计算评分一致性系数：

$$\text{评分一致性系数} = \frac{\text{评分一致小题数}}{\text{试卷小题总数}} \times 100\%$$

保证评分一致性系数在 90% 以上，再由阅卷者独立阅卷。期间，阅卷组长还会抽检 10% 的试卷。

###### 4.3.2 整体信度

测试使用内部一致性来衡量试卷的整体信度。

内部一致性也称为同质性信度 (homogeneity reliability)，是指测试内部所有题目间的一致性程度。在教育测量中一般使用克龙巴赫  $\alpha$  系数来刻画内部一致性，若 Cronbach  $\alpha$  系数  $\geq 0.70$ ，可视为信度良好<sup>[14]</sup>。

经计算，7 年级测试卷的内部一致性系数为 0.866，表明测试卷具有良好的信度。

##### 4.4 效度

效度是指测试实际能测出所需要的能力特质的程度，效度是相对于测量目的而言的，因此在衡量测试效度前要明确测量目的。2006 年出版的美国权威教科书《教育测量》指出，效度检验是对自己提出的观点收集证据进行证明的过程<sup>[15]</sup>。

###### 4.4.1 内容效度

评估内容效度的常用方法是专家评议法，请专家判断测试内容是否恰当。

测试通过以下方法来保证内容效度：首先，请数学教学论专家及该区的教研员把关测试框架及指标体系；其次，在测试结构和指标体系的基础上编制双向细目表，依据细目表命题；最后，在命题、试测、评标商定的过程中邀请一线教师、教研员、数学教育专家、教育测量专家参与，并根据修改意见不断完善试题和评价标准。

###### 4.4.2 结构效度

因子分析法是探索和验证测试结构的有效方法<sup>[16]</sup>。在测试前，已经对试卷有了详细的设计方案（内容维度、能力维度）；测试实施后，分别采用探索性因子分析 (EFA) 和验证性因子分析<sup>[17]</sup> (CFA) 两种方法检验结果效度。

从内容维度和能力维度对试题进行因子分析，结果显示试卷的结构效度良好：每个题目的因子载荷均高于截断值<sup>[18]</sup>，拟合指数 (CFI) 和 Tucker-Lewis 指数 (TLI) 均在 0.9 以上，近似误差均方根 (RMSEA) 估计值在良好 (0.05~0.08) 范围内，标准化残差均方根 (SRMR) 小于 0.08，模型拟合良好<sup>[19]</sup>。

## 5 结果呈现

### 5.1 分数转化

填空题使用的是单参数 RASCH 模型，解答题使用分步

计分模型,并通过 CONQUEST 软件对数据进行估计,再使用多维 RASCH 模型对内容维度(方程与不等式、图形的性质、统计与概率)和能力维度(学习理解、实践应用、创造迁移)进行估计。

软件默认的输出分数为标准分(又称为 Z 分数),是将原始分与测试群体评教分之差除以标准差得到的商数,该值不受原始测量单位的影响,具有可比性和可加性。Z 分数的群体均分为 0,正负表示原始分在平均分之上或之下,绝对值表示原始分与平均分之间的距离。相较原始分,Z 分数能够真实反映一个分数距离平均数的相对标准距离。

由于 Z 分数会出现负值和小数,让人在阅读时常造成困难和误解,美国测量学家 McCall 建议将 Z 分数扩大 10 倍再加 50 转化为 T 分数<sup>[20]</sup>,以消除负号和小数。但由于 T 分数在 0~100 之间,均值为 50,标准差为 10,有一半人在 50 分以下,容易被当成百分制的分数,并产生有 80% 以上学生不及格的误解。

测试选用的是美国大学入学考试委员会使用的标准分——CEEB 分数:

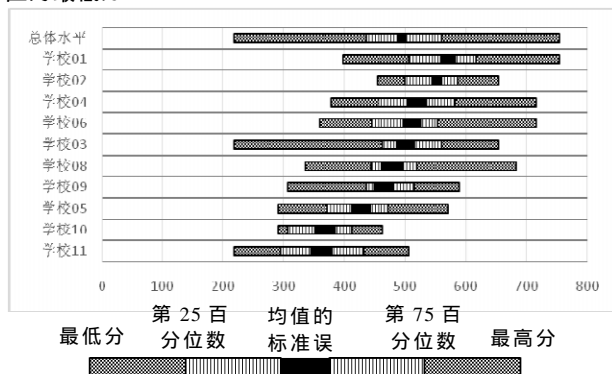
$$CEEB\text{分数} = 100 \cdot Z + 500$$

该分数的均分为 500,标准差为 100,形式上不容易造成误解。此外,国际学生评估项目(Program for International Student Assessment,简称 PISA)及国际数学和科学趋势研究(The Trends in International Mathematics and Science Study,简称 TIMSS)的报告中均采用了此标准分。

由于试题分内容维度和能力维度,所以反馈结果丰富,下文分别以学区、学校及学生个人举例,呈现部分结果。

## 5.2 学区结果反馈

从图 1 可以得到很多信息,最上方的一条为整个学区考生的学科能力估计值的分布情况,可以知道本次测试的最高分、最低分及整体的平均水平。另外,还可以看出,学校 2 的均分较高且学生能力相对集中;虽然学校 3 的平均分也不错,但是学生两极分化严重:前 75% 的学生得分优于均分较高的学校 4 和学校 6,后 25% 的学生成绩较差,甚至还有全区的最低分。



注:学校 7 只有高中部,所以未参加 7 年级测试

图 1 某区各学校学科能力估计值的分布情况

## 5.3 学校结果反馈

通过图 2 可以发现,虽然学校 5 的平均分比学区的平均分低,但是从内容维度看:方程与不等式、概率与统计的差

距并不大;从能力维度看:学习理解和实践应用的差距也较小。学校 5 和学区总体差距较大的是图形与几何及创造迁移能力,反观学校 5 内部最高分与最低分两个学生情况,最大的差异也同样出现在图形与几何及创造迁移能力,所以,学校 5 在今后的数学教学中尤其应注意不同学生这两点的发展均衡问题。



图 2 学校 5 各维度得分与学区总体水平比较雷达图

## 5.4 学生个人结果反馈

学生个人结果的反馈能够让学生了解更多的细节信息,如表 3 所示,该考生已经进入学区前 10% 的行列,而且在某些维度上已经达到前 5% 的水平。根据这份反馈表格,该考生可以知道在今后的学习中,应该更加注重创造迁移能力的发展同时也应多花些精力在统计与概率的学习上。

表 3 学生个人各维度得分情况

	考试编号:XXXX		均值	前 20%	前 10%	前 5%
	总体表现					
知识	方程与不等式	510.34	352.18	448.40	481.23	504.57
	图形与几何	476.05	291.67	391.06	427.77	452.11
	统计与概率	559.73	520.38	587.29	611.58	640.26
能力	学习理解	640.85	532.89	606.66	627.79	642.60
	实践应用	657.75	502.62	597.10	639.00	658.61
	创造迁移	434.18	282.37	381.26	426.49	451.76

## 6 总 结

作为数学学科能力测试,主要在以下方面有别于一般的学业考试和能力测试:

第一,基于课程标准。一般的能力测试常会落脚于智力测验或能力倾向测试,其中涉及数学能力的模块包括数感、抽象推理、空间关系理解等,这类测试适用范围广,但没有明确的背景知识要求,无法满足学科能力测试的要求。数学课程标准为能力维度的划分提供了良好的参照,同时也为能力与学科的有机结合构建了理想的桥梁,使得能力、内容的二维评价框架有据可依。

第二,选取核心知识。传统的学业考试是在某一学习阶段完成后的终结性测试,目的是检测学生对所学知识的掌握情况。整个中学阶段的知识复杂繁多,即便具体到某个年级,知识点也较多,不可能单凭一份试卷准确测量。学科能力测试并不试图覆盖全部知识点,而是根据整体设计,选取核心知识模块(如函数、方程与不等式、图形的性质、统计与概率),再根据特定年级的教学内容在对应知识模块下选取核心知识点进行考查。这样不仅提高了一份试题对核心知

识点检测的准确性，同时也为跨年比较提供了途径。

第三，重视能力指向。常规的学业考试注重知识点的检测，而忽视学科能力的考查。学科能力测试内容维度与能力维度并重，包括学习理解、实践应用、创造迁移3个一级指标及9个二级指标。如学习理解，侧重学生对教材知识再现、理解的能力，而实践应用技能则侧重考查学生利用所学知识合情推理、简单问题解决的能力。所以，学科能力测试不但可以检索学生对核心知识的掌握情况，还能够了解学生在某知识上的能力发展情况，方便教师因材施教。

第四，命题过程严格。通常的学业考试往往由出题教师根据自身经验命题，对试题难度的把控也是依据对测试对象的主观了解。而学科能力测试有规范设计的命题框架，从命题、选题、试题研磨、组卷等每一环节都有质量监控措施，还会有预测甚至多轮预测以优化试题和评价标准。

第五，评价结果丰富。传统学业考试的结果通常只有总分及排名，考生无法详细了解问题所在。学科能力测试的结果包括总的能力估计值、不同内容维度上的得分还有不能能力维度上的得分，考试可以明确了解自己在核心知识上的欠缺以及具体能力的不足。此外，学科能力测试结果为标准分，消除了整体试卷难度变化对得分的影响，为得追踪得分研究学生能力发展提供了科学的依据。

致谢：对北京市育英学校郭转娜，中国人民大学附属中学王立东，北京师范大学第二附属中学王西辞、郭方奇，北京教育学院陈鹏举、刘晓婷，天津师范大学秦华，北京一零一中学贾思雨等教师在测试工具研制工作中的贡献，在此一并表示感谢！

#### [参 考 文 献]

- [1] 林崇德. 论学科能力的建构[J]. 北京师范大学学报(社会科学版), 1997, (1): 5-12.
- [2] 全美数学教师理事会. 美国学校数学教育的原则和标准[M]. 北京: 人民教育出版社, 2004.
- [3] 曹一鸣, 王立东, Cobb P. 美国统一州核心课程标准高中数学部分述评[J]. 数学教育学报, 2010, 19(5): 8-11.
- [4] 吴立宝, 曹一鸣. 初中数学课程内容分布的国际比较研究[J]. 教育学报, 2013, (2): 29-36.
- [5] 康明媛, 曹一鸣, Clarke D. 中、澳、芬数学课程标准中内容分布的比较研究[J]. 教育学报, 2012, (1): 62-66.
- [6] 董连春, Stephens M. 澳大利亚全国统一高中数学课程标准评述[J]. 数学教育学报, 2013, 22(4): 16-20.
- [7] 曹一鸣, 代钦, 王光明. 十三国数学课程标准评介(高中卷)[M]. 北京: 北京师范大学出版社, 2013.
- [8] 曹一鸣. 十三国数学课程标准评介(小学初中卷)[M]. 北京: 北京师范大学出版社, 2012.
- [9] 徐斌艳. 旨在诊断与改进教学的数学学科能力测评分析——来自德国的实践[J]. 全球教育展望, 2011, (12): 78-83.
- [10] 刘晓玫, 陈娟. PISA与TIMSS中有关数学评价的比较分析[J]. 外国教育研究, 2007, (2): 77-80.
- [11] 汪贤泽. 基于课程标准的学业成就评价程序研究[D]. 华东师范大学, 2008.
- [12] 王立东, 王西辞, 曹一鸣. 数学课堂教学中的学生评价研究——基于两位教师课堂录像的编码分析[J]. 数学教育学报, 2011, 20(5): 37-40.
- [13] Ebel R L. *Essentials of Educational Measurement* [M]. Prentice-Hall, Inc, 1972.
- [14] Nunnally J. *Psychometric Theory* [M]. New York: McGraw Hill, 1978.
- [15] Brennan R L. *Educational Measurement. ACE/Praeger Series on Higher Education* [M]. Greenwood Pub Group, 2006.
- [16] 戴海崎, 张峰, 陈雪枫. 心理与教育测量[M]. 广州: 暨南大学出版社, 2011.
- [17] Brown T A. *Confirmatory Factor Analysis for Applied Research* [M]. Guilford Press, 2006.
- [18] Ford J, Maccallum R, Tait M. The Application of Exploratory Factor Analysis in Applied Psychology: A Critical Review and Analysis [J]. *Personal Psychology*, 1986, (39): 291-314.
- [19] Browne M W, Cudeck R. Alternative Ways of Assessing Model Fit [J]. *Sociological Methods & Research*, 1992, 21(2): 230-258.
- [20] Mccall W A. *Measurement: A Revision of How to Measure Education* [M]. Macmillan, 1939.

### Researcher Assessment of Mathematics Competence Based on Curriculum Standards —Take the Development and Application of Testing Tools for Grade 7 Students as an Example

GUO Kan, CAO Peng, YANG Fan, LIU Jin-hua

(School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China)

**Abstract:** The 2-dimensional assessment modeled signed by content and dimension is not only the main trend of international studies, but an important requirement of educational reform and development in Chinese well. By taking the development and application of testing tools for grade 7 students as an example, this present article introduces the testing framework, evaluation process, technical indexes and feedback forms of the assessment of mathematics competence based on curriculum standards. This article also compares the differences between the assessment of mathematics competence and ordinary school work exam or ability test.

**Key words:** mathematics competence; educational assessment; content dimension; ability dimension

[责任编辑: 周学智]