

学生学力测试与学校质量监测

——如何评价学生所获得的数学能力的发展
以及学校数学课程的发展

北京师范大学教育学院
张春莉 教授


一、学生学力测试的理论基础


- 1、什么是学力
- 学力是指通过学习获得的能力，是人的能力的基础部分。学力是在有意识、有计划、有组织的学校教育中，通过师生的相互作用形成的。学力体现了客体与主体的统一、外化与内化的统一、适应和创造的统一。




2、学力的结构

- 1956年，美国教育学家布卢姆发表了《教育目标分类学——认知领域》，把学力结构分为：认知、情感、动作技能三大领域。


- 
- 1965年，美国心理学家加涅发表了其名著《学习的条件》，并不断修订与完善其关于学习结果的研究。1985年，加涅在《学习的条件和教学论》中提出五种学习的结果，言语信息、智慧技能、认知策略、动作技能和态度。





- 1983年，美国哈佛大学教授加德纳提出了多元智能的理论，认为每个人至少有七种智能，即语言、逻辑——数学、空间、身体——运动、音乐、人际关系、自我认识智能等。随着研究的深入，他于1996年又提出了第八种智能——自然观察者智能。上述八种智能又可归为三个更一般的领域：与“物体有关”的智能，与“物体无关”的智能以及“与人有关”的智能。



- 1966年，美国耶鲁大学心理学家斯腾伯格在其《成功智力》一书中提出成功智力包括分析性智力、创造性智力和实践性智力三个方面。分析性智力用来解决问题和判定思维成果的质量；创造性智力帮助形成好的问题和想法；实践性智力用来解决实际工作中的问题。


- 
- 1996年，国际21世纪教育委员会向联合国教科文组织提交历时三年的研究报告《教育——财富蕴藏其中》。报告指出：面向未来社会发展，教育必须围绕四种基本的学习能力来重新设计、重新组织。这四种基本学习能力被称为是教育的四大支柱。

- 
- 1、学会认知，即掌握认识世界的工具。
 - 2、学会做事，即学会在一定的环境中工作。不仅是实际的动手能力，而且包括处理人际关系、社会行为、主观能动性、处理信息、解决矛盾、敢于承担风险等多种能力。
 - 3、学会共同生活，即通过增进对他人以及历史、传统和精神价值的了解，培养在人类活动中参与合作精神。
 - 4、学会发展，即对自己的有更深入的了解，能够适应和改进自己的环境。这是前三种学习成果的主要表现。



随着知识经济对人的基本素质要求的不断提高，学者们对学力结构的研究也不断深化、拓展。学力已不再局限于传统教育所特别重视的语言与数理逻辑能力，而更强调能力的多元性，重视知行统一，重视实践与创造，以应对飞速变化新环境的挑战。

3、（PISA）数学素养的定义



数学素养是指在当前或未来的生活中为满足个人成为一个会关心、会思考的市民的需要而具备的认识，并理解数学在自然、社会生活中的地位的能力，作出判断的能力，以及参与数学活动的能力。

- 数学素养包括若干运用数学能力的水平层次，从标准数学运算到数学思维能力和观察力。
- 它还要求学生理解和应用一定范围内的数学知识，例如：可能性、变化和增长、空间和形状、定量推理、不确定性和从属关系等。这些包括数学课程的特定领域，如代数、算术和几何。

(PISA) 数学素养的组成

数学素养由两个主要领域和两个次要领域组成。


两个主要领域是：数学能力、数学的主要思想；两个次要领域是：数学课程内容、情境与背景。

- 两个主要的领域是为了用来描述评价的范围和熟练程度。而次要的领域则是为了确保评价有足够的覆盖率，并在一系列已选择的评价任务中起到一个平衡作用。
- 数学能力是指一般的技能和能力如问题解决，数学语言和数学模型的使用。
- 数学的主要思想体现的是出现在真实情境中的一连串的有相互关联的数学概念。
- 数学课程内容是指在学校课程中所教授的数学内容。对于OECD/PISA来说，数字、测量、估计、代数、几何、函数、概率、统计、离散数学都是应该涉及的内容。
- 情境与背景指的是数学问题提出所依托的背景，如：教育、职业、公众以及个人背景。

二、学力测试的特点


- 1、学力测试的目的
- 两种不同性质的测验：
 - 学业成就测验（又称学绩测验）
 - 学业能力测验（进一步还可分为能力倾向测验和能力水平测验）。
- 成就测验和能力测验都要测量已形成的能力，但除此之外，能力倾向测验还要能测量学生的学习潜能，预见学生未来学习的潜力。

2、学力测试的特点



学力测试的特点是常模参照测验与效标参照测验相结合的特点。因为常模参照测验对于测量学生通过不同途径解决复杂问题的综合能力特别有效，但它也有它的局限性，它的结果不能说明学生是否准备好了学习更高级的内容。比如，一次数学测验中，一名学生处于全班的前3%，但这并不能表明这个学生已经具备了学习下一阶段的数学的能力，可能班级中的每一个人对数学的理解都是不足的。而效标参照测验测量的是对相对具体的目标的掌握情况，测验的结果要与一个既定的准则或成绩标准做比较，以判断学生是否具备了进入下一阶段的学习的能力。

三、学力测试的原则



1、学力测试所涉及的知识应尽量避免考查只需死记硬背的内容

- 有些内容是学生学习过程中必须知道的一些基本事实，但即使学生不理解，只要死记硬背就会做，这样的试题就不应该放入学力测试题中，因为它考察的不是学力，而是记忆力和具体的信息。

例：判断当一张纸被切碎成小条后，面积比原来大、小、还是相等？（图形显示一张完整的纸，以及切碎后的纸）

例：连续5次抛一枚均匀的硬币，落地后都是正面朝上，知道再抛一次的结果是正面朝上还是反面朝上。

2、学力测试要涉及与该能力密切关系的不同的内容维度，同一内容维度要体现出不同的难度要求


PISA从以下三个维度评价数学素养


- (1) 数学内容。首先它主要指的是在数学思考的意义下具有更广泛意义的数学内容（如可能性、变化和增长、空间和形状、推理、不确定性和从属关系），其次才是有关的具体的课程内容（如数、代数和几何等）。
- (2) 数学过程。它是根据一般的数学能力来进行定义的。
- (3) 数学情境。包括从个人生活到与科学和公共事务有关的一系列问题。




数学能力包括

- **a. 思考和推理：**包括提出数学问题；知道各种数学答案；能够区分不同的数学表达（定义、定理、推理、假设、实例、条件命题）；并且能够理解和掌握所给数学概念的外延和局限。

- 
- **b. 论证：**包括知道数学上的依据是什么，它与别的数学推理有何区别；能够理解和评判不同种类的数学论证；能够提出一些启发式的问题（可能（不）会发生什么？为什么？）；能够构建自己的论证方法并将其表达出来。

- 
- **c. 交流：**对于数学情境中的某一事件，能够以各种不同的方式（口头的或书面的）将其表达出来，而且能够理解他人对这一事件的陈述（口头的或书面的）。

- 
- **d. 建模：**将要建模的情境简化为现实的模型；将“现实”模型翻译成数学模型；根据其现实性来解释该模型；用数学方法来处理该模型；对该模型进行验证；对该模型及其解进行反思、分析和评论；针对该模型及其解进行交流（包括解的局限）；并对建模的过程进行监控。



- e. 问题提出和解决：能够提出、识别各种不同的数学问题（例：纯数学问题、应用问题、开放题、封闭题）；用各种不同的方法解决不同类型的数学问题。



- f. 表达：能够解释和辨别数学对象、数学情境的各种不同表达方式，以及不同表达方式之间的相互关系；能够根据不同的情境和目的在各种不同的表达方式之间进行选择 and 转换。



- **g. 使用符号化、正规化、技术化的语言和运算：**能够将符号化、正规化的语言解释翻译成生活中的语言；能够将生活中的语言转换成符号化、正规化的语言；能够掌握带有符号、公式的叙述和表达式；运用变量解方程、进行计算。



- **f. 运用辅助工具：**知道并且能够运用各种辅助工具（包括信息工具），并且知道这些辅助工具的局限性。



各种能力不可能完全独立，也不可能分开来测量。

- 比如，PISA测试中，它并不会应用不同的试题来分别考察数学的各种能力，因为在解决一个“现实”数学问题时需要若干的数学能力。因此测试题是根据这些能力的水平而组织的。



水平一：常规的数学评价中出现的简单计算或定义。

- 这一水平的考察内容在许多标准化考试和国际比较研究中经常采用的，包括关于事实的知识，表述，再现，回忆，按照常规解题的能力。这部分题目多数是选择题，也有少部分开放题。这部分题目多与符号、公式以及计算技巧有关。
- 例1：四则运算、解方程等。
- 例2：7，12，8，14，15，9的平均数是多少？

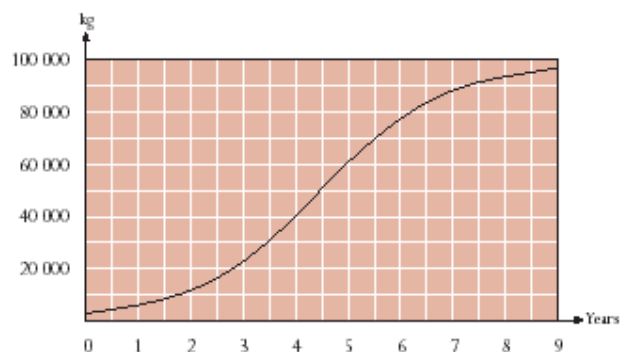
水平二：要求建立知识之间的联系，解决简单的问题。

- 在这个能力水平上，注重的是不同内容与领域之间的联系以及为了解决简单问题而整合信息的能力。学生必须会选择他们的策略和工具。虽然这些问题并非常规问题，但是只需要相对低水平的数学能力。这部分题目需要学生具有证明的能力，同时也需要建模的能力，以及提出问题、解决问题的能力。
- 例3：你已经驾车行驶了 $\frac{2}{3}$ 的路程，油箱的油量从一开始的一箱消耗到现在的 $\frac{1}{4}$ 箱，你能行驶完全程吗？
- 例4：Mary和Martin分别距学校3千米和5千米，Mary距Martin多远？

水平三：包括数学思维、概括和洞察力，要求学生分析、识别一个情境的数学本质，提出自己的问题。

- 这个水平的题目要求学生对情境进行数学化处理，能够从具体的情境中识别、抽象出数学问题并解决：分析，解释说明，形成自己的方法、策略，提供证明，包括论据和概括，这其中还包括批判性思维、分析和反思。在这个能力水平，不仅要求学生能够解决问题而且要提出问题，并能够洞察数学的本质。
- 这一水平是数学和数学素养的核心，同时也是很难进行测试的部分。多项选择题不太适合这部分的评价，开放题是比较合适的，但是无论是题目的设计还是答案的评分都是非常困难的。

例5：在一个池塘中引种一些鱼苗，池塘中鱼的总重量的增长模型如下图所示。假设渔民计划若干年后开始捕捞，则多少年后开始捕捞，以后每年的收益最大？说明你的理由。



3、学力测试题在内容和形式上都应充分鼓励学生构建自己的回答。

- 学生自己编写答案，答案的形式可以是短文、问题解决的方案、图或表，而不是简单的选择（一个且惟一）正确答案。这样的测验会测量到那些较高水平和发散性的思维。



- 例：一个三位数的三个数字之和是10,其中个位和十位的数字相同,这个数可能是多少?
- 把你能想到的都写出来.



- 例：商场的李叔叔承包的柜台有如下三种车辆,自行车600元,电动自行车900元,摩托车1800元.如果李叔叔今天的销售收入共5400元钱,请你想一想,李叔叔这一天可能卖出哪几种车?各卖了多少辆?




4、学力测试的结果要能够预测学生未来的学习能力

- 测试结果要能够预测学生未来的学习能力即学习潜能。既然学生目前还没有实现的能力如何能测量出来呢？这里就涉及到一个心理学的概念——“最近发展区”。所谓最近发展区，是指孩子不能独立解决但可以在他人的指导下解决问题的机能领域。最近发展区是一个范围，这样范围越大，就表明这个孩子的学习潜能越大。



关注学习过程而非结果

- 乔·坎皮恩（Joe Campione）和安·布朗（Ann Brown）提出的方法是，当孩子解决一个问题时，为他或她提供循序渐进的提示。这些提示被书写出来，开始是一般的提示，最后是星系说明如何发现这个答案。孩子使用提示的方式和测验情境下的学习方式为其学习潜力提供了证据。
- 这些测验的任务反映了儿童从阅读学习材料或解题指导中受益的程度，关注的是孩子学习的过程，而不是关注孩子的当前水平。




例: $3 \times 9 = 27$

$3 \times 99 = 297$

- $3 \times 999 = 2997$
- $3 \times 9999 = 29997$
- 那么, $3 \times 9999999 = (\quad)$

- A、2999997 B、30000007
- C、29999997 D、299999997




- 例: 在玛格内行星上,人们使用migs, mags及mogs为钱币,
- $1 \text{ mags} = 8 \text{ migs}$ $1 \text{ mogs} = 6 \text{ mags}$
- 请问 $5 \text{ mogs} + 3 \text{ mags} = ?$

- (A)54 migs (B)120 migs
- (C)240 migs (D)264 migs



5、对测试结果进行解释时必须把误差的因素考虑进去。

- 每种测验都会存在误差，含有选择题越多的测试，这种误差就越不可能避免，因为学生更可能会凭运气来选择答案。如何降低误差以使实际分数更接近真实分数呢？首先在编制测验时，就要努力提高测验的信度。增加测验题目数量是提高信度的最有效方法。一般来说，题目数量多的测验比题目数量少的测验信度高。



但是受考试时间和学生的心理承受力的限制，我们不可能编排过量的试题数量，所以还需要在编制完测验之后，考虑测验的标准误，即估计在重复测试的情况下学生成绩可能发生的变化有多大。这样在解释测验分数时，就必须把标准误考虑进去，学生的分数不能看做是单一的点，而是一个包括该分数的区域。两个学生的分数之间的几分之差就不应被视为重要区别，只有当两个分数的区域之间没有重叠时我们才可以认为他们之间确实有区别。



四、数学学力测试命题原则

- 数学学生学习质量监控的命题，除了必须具备的基本要求（如内容的科学性、知识点的覆盖率、表述的简洁明了、文字和格式的规范性及内部和外部效度的有效性等）以外，还必须遵循以下基本原则：



(1) 考查内容要依据《学力要求》，体现基础性、应用性和时代性

- 考查内容要突出对学生基本数学素养的评价，试题应首先关注《学力要求》中基本、核心的数学观念、思想和方法，不在技巧性上做过高的要求。



对“数与代数”部分的评价

- 应结合具体情境考查学生对数的意义的理解，应主要考查学生对概念、法则及运算的理解与运用水平，不应单纯地考查对知识的记忆，对于运算的评价不能过分要求技巧。“数与代数”中要有探索规律、估算、鼓励学生算法多样化、考查学生符号意识和函数思想（包括数学建模）的试题。



对“空间与图形”部分的评价

- 要结合直观素材和生活情境评价学生对图形的认识，应主要考查学生对基本几何事实的理解、空间观念的发展及合情推理的能力和初步演绎推理能力的获得。



对“统计与概率”部分的评价

- 重点应放在考查学生用适当的方法描述数据的技能，考查学生读图（表）及分析数据的技能，应结合生活情境考查学生是否具有统计观念和初步的概率意识。



(2) 试题素材或情境要具有公平性和多样性

- 数学试题的素材或情境应对每一个学生而言是公平的，要避免需要特殊背景知识才能够理解的试题素材或情境。同时数学的情境应具有多样性，既要有纯的数学问题，目的在于考查学生探索数学内部问题的能力，又要有实际或拟现实的问题，目的在于考查学生应用数学概念和规则解决问题的能力，还要有一些真实的现实生活中的数学问题，目的在于考查学生对现实世界的数学洞察力，包括区分有用的数据和无关的数据的能力，用数学和非数学的方法结合起来解决有关问题的能力以及能综合考虑现实生活中的合理性和可行性提出相应的策略、方法或答案的能力，而这类试题，命题人员要格外注意，对试题情境的描述应尽量简洁明了，整张试卷不要造成学生过大的阅读量。



(3) 试题的表述和答案的选项要体现探索性和开放性

- 学力标准提倡学生通过观察、实验、猜测、验证、推理和交流等数学活动方式来学习数学，在试题的构思和设计上也应该体现解题过程的探索性和解题结果的开放性。为此，在试题条件的表述上，提倡考查学生区分有用的数据和无关的数据的能力，以及通过对实践情境的转化发现隐含的已知条件的能力。在答案的选项上，提倡考查学生对答案的合理性和多解性的审视能力，同时应避免“以上都不对”，或“A、B、C都正确”类似这样凑数的选项。另外，要尽量避免可能而不必要的解题陷阱，比如，在试题条件的表述中出现太多的“不”字或“非”字，或者要求学生找出以下答案选项中“不正确”或“错误”的选项。



(4) 试题的难度应追求有价值的难度因素

- 试题的难度要从每一水平的试题中加以控制，对于事实性知识和简单技能，难度（得分率）不能低于0.7，避免考冷门的记忆性知识，对于概念理解的试题，不提倡出过于复杂、抽象的问题，要结合具体的情境或直观形象的支撑来加以考查，对于运用规则的试题，不能出计算上过于繁难的试题，而对于解决问题的试题，要尽量避免以往的解题经验的影响，对于难度在0.5以下的难题，必须是所有学生都不熟悉的新情境，但同时也要杜绝偏题或怪题。

五、为学校数学课程的改进提供信息

- 对学校课程改进进行需求和影响因素分析时，一个核心的词语就是多元化——多元化的内容、多元化的资源以及多元化的方法。
- 1984年，Goodlad在“学校教育研究”中对学校教育的基本范围进行了分析，提出了一套具有参考意义的框架体系。

数据类型

| | | 个人的 (个体的) | 教学的 (课堂) | 机构的 (学校) | 社会的 (学校教育) |
|-----|-----------|--------------|-------------|-------------|---------------|
| 数据源 | 教师 | | | | |
| | 学生 | | | | |
| | 家长 | | | | |
| | 教师学生的交互作用 | | | | |

课程
活动
意义

| 数据类型 数据源 | 个人的 (个体的) |
|-------------|---|
| 教师 | <ul style="list-style-type: none"> •人口统计资料 •从事教育职业的原因 •教学经验 •教育信条 |
| 学生 | <ul style="list-style-type: none"> •人口统计资料 •自我概念 •学习期望 |
| 家长 | <ul style="list-style-type: none"> •人口统计资料 •生活在社会中的时间 •政治信条 |
| 教师学生的交互作用 | |

| 数据类型 数据源 | 教学的 (课堂) |
|-------------|--|
| 教师 | <ul style="list-style-type: none"> •在教学、行为控制和日常事务上分别花费的时间 •行为目标的使用 •某些学习行为的出现频率 |
| 学生 | <ul style="list-style-type: none"> •在教学、行为控制和日常事务上分别花费的时间 •课堂学习内容的难度 •某些学习活动出现的频率 •课堂“氛围” |
| 家长 | |
| 教师学生的交互作用 | <ul style="list-style-type: none"> •课堂（教师学生的交互作用） •在教学、行为控制和日常事务上花费的时间 •校正性反馈的运用 •开放式问题与封闭式问题的使用 •在整个班级、个人和小组上分别花费的时间 |

| 数据类型 数据源 | 机构的 (学校) |
|-------------|--|
| 教师 | <ul style="list-style-type: none"> •各种学校功能(社会的、学术的、个人的和职业的)的相对重要性 •学校“氛围”或者工作环境 •主要问题 •教育的公平性(能力、种族和性别) |
| 学生 | <ul style="list-style-type: none"> •学校功能的相对重要性 •评定等级 •主要问题 •教育公平性 •咨询工作的效果 •学科偏好 |
| 家长 | <ul style="list-style-type: none"> •学校功能的相对重要性 •评定等级 •主要问题 •教育公平性 •在学校活动和决策中的参与状态 •令人不快的学习材料 |
| 教师学生的交互作用 | |

| 数据类型 数据源 | 社会的 (学校教育) |
|-------------|--|
| 教师 | <ul style="list-style-type: none"> •种族隔离 •对公立(入网)学校的财政支持 •教师联合会 •最低资格 •全球教育在学校中的作用 |
| 学生 | <ul style="list-style-type: none"> •种族隔离 •学校对工作经验的看重情况 •学校的价值观 |
| 家长 | <ul style="list-style-type: none"> •种族隔离 •对公立(入网)学校的财政支持 •教师联合会 •教师薪水 •最低资格 •全球教育在学校中的作用 |
| 教师学生的交互作用 | |

促进改进的一条重要的理念

- 把信息的收集和使用看成行政人员、教师、学生或家长、个人或集体的某种自我检视。
- 如果在看法、观念和使用上都能将重视信息当作学校的“文化规则”，形成一种合作式的、基于信息的探究和评价观，学校就能在促进学校不断改进的评价过程中发挥积极的作用，并使学生真正受益。

