



心理
测量学

PSYCHOLOGICAL
TESTING

主讲：黎 坚
单位：北京师范大学心理学院
信箱：JianLi@bnu.edu.cn


School of Psychology Beijing Normal University



第二章 测验的编制

- ①第一节 确定测验目的
- ①第二节 项目的编制
- ①第三节 测验的编排与组织
- ①第四节 测验的标准化
- ①第五节 测验的鉴定与使用说明


Li, J. School of Psychology, Beijing Normal University



第一节 确定测验目的

- ①1.1 明确测验的对象
- ①1.2 明确测验的用途
- ①1.3 测验的目标
- ①1.4 测验编制计划


3 Li, J. School of Psychology, Beijing Normal University



1.1 明确测验的对象

- ①了解团体的组成和特点。
- ①年龄、智力水平、社会经济和文化背景以及阅读水平等。


4 Li, J. School of Psychology, Beijing Normal University



1.2 明确测验的用途

测验用于描述、诊断、预测、选拔还是安置？

5 Li, J. School of Psychology, Beijing Normal University



1.3 确定测验的目标

- ①测验将用于测量什么心理构想或行为表现，例如目标是测量学业成就、智力、特殊能力倾向，还是人格。
- ①编制者首先需要对所要测量的心理构想进行概念界定（概念化），然后需要说明该构想所包含的内容或维度将通过什么行为表现出来或怎样进行测量（操作化）。

6 Li, J. School of Psychology, Beijing Normal University

1.4 制定测验编制计划

① 测验目标确定后，需要进一步制定编制计划，列出测量内容大纲。

① 不同测验的编制计划，其重点不同

① 用于招聘选拔的测验重视收集工作分析资料

① 智力测验注重理论基础或者选取实证效标

① 成就测验需要明确测验的内容范围

①

7

Li, J. School of Psychology, Beijing Normal University

(例) 成就测验的编制计划

① 需要确定能表征测量目标的内容或技能范围，并制定各部分比例（即编写双向细目表），以便行为取样。例如：

① 测验内容与测验目标（或题型）之间关系的双向细目表

① 题型与难度、测验内容之间关系的双向细目表

8

Li, J. School of Psychology, Beijing Normal University

小学自然常识测验编题计划双向细目表
——测验内容与认知目标

测验 教材 内容	知识 Knowledge	理解 comprehension	应用 application	分析 analysis	综合 synthesis	评价 evaluation	合计
生物世界	6	5	6	3	2	0	22
资源利用	2	3	3	1	1	0	10
动力和机械	2	3	4	2	0	0	11
物质、物性与能量	5	6	8	3	2	0	24
气象	2	4	3	2	2	0	13
宇宙	2	5	4	1	0	0	12
地球	2	2	2	1	1	0	8
合计	21	28	30	13	8	0	100

注：表中的数字代表每一类题目所占的百分比

制定测验编制计划的方法

- 回顾以往的研究成果
- 考虑时代的特点
- 了解受测群体实际情况
- 向有关专家、资深者咨询和请教

10

Li, J. School of Psychology, Beijing Normal University

第二节 项目的编制

②.1 搜集资料

②.2 命题的一般原则

②.3 确定项目的形式

11

Li, J. School of Psychology, Beijing Normal University

2.1 搜集资料

为编制测验而搜集的资料应当内容丰富，覆盖面广，代表性强，并带有一定趣味性。来源可包括：

- 已出版的标准测验
- 理论和专家的经验
- 临床观察和记录

12

Li, J. School of Psychology, Beijing Normal University

2.2 命题的一般原则

- ① 内容方面：题目的内容符合测验目的；内容取样有代表性；题目间内容相互独立，互不牵连。
- ① 文字方面：使用当代语言，语句简明扼要，最好一句话说明一个概念，尽量不用双重否定句。
- ① 理解方面：题目应有确切的答案，但不能有答题倾向线索，题目的内容不要超出受测团体的知识水平和理解能力，题目格式要容易理解。
- ① 社会敏感性方面：尽量避开社会敏感问题，如涉及社会禁忌或个人隐私的题目不应使用。

13

Li, J. School of Psychology, Beijing Normal University

若必须涉及敏感问题（性、犯罪、自杀）

- ① 怎样鼓励受测者作出真实的回答（菲利普）：
 - ☞ 命题时假定受测者具有某种行为，使其不得不在确实没有该行为时才否定，可避免否定过多的倾向。
 - ☞ 命题时假定规范不一致。
 - ☞ 指出该行为是常见的，虽然是违规或违法的。

① 随机反应技术（Random Response Technique, RRT）

14

Li, J. School of Psychology, Beijing Normal University

2.3 确定项目的形式

- ① 根据对受测者的要求不同来分，可以分为两大类：选择型题目和构造型题目。
 - ☞ 选择型（selected-response format）题目要求受测者在有限的几个答案中选择正确的答案，如选择题、是否题、匹配题等。
 - ☞ 构造型（constructed-response format）题目要求受测者给出正确答案，如论文题、简答题、填充题等。

15

Li, J. School of Psychology, Beijing Normal University

根据评分的客观性分类

- ① 客观题（objective test）
 - ☞ 把受测者的反应与事先定好的正确答案相比较，评分有很强的客观性，不同评分者结果一致。
 - ☞ 注意：既可以是构造型也可以是选择型。
- ① 论述题（essay test）
 - ☞ 要求受测者针对题目写出相当长的答案，然后由教师或其他评价者对这些答案进行主观评定。
 - ☞ 论述题都属于构造型题目。

16

Li, J. School of Psychology, Beijing Normal University

2.3.1 判断题（是非题、正误题）

true-false item

- ① 向受测者提供一个陈述句，让受测者判断它正确与否，做出“是”或“否”、“正确”或“错误”的判断。

17

Li, J. School of Psychology, Beijing Normal University

编制原则

- ① 应绝对正确或错误
- ① 测量重要的而不是细节性的问题
- ① 一题只包含一个概念
- ① 避免反面叙述，尤其是双重否定
- ① 避免含混不确定的描述
- ① 避免使用限定词或暗示性的词汇
- ① 一般是与非的题数应该相等，且随机排列

18

Li, J. School of Psychology, Beijing Normal University

2.3.2 评定量表 (rating scale)



① Likert量表: 要求受测者对一系列陈述句表示看法, 有程度的区分。通常该尺度有5或7点。了解受测者的习惯、态度、兴趣等。

19

Li, J. School of Psychology, Beijing Normal University

2.3.3 选择题 (multiple-choice item)



① 选择题由题干 (Stem) 和选项 (alternatives) 两部分组成, 题干可以是不完全陈述句, 也可是直接疑问句。选项包含一个或几个正确的答案以及若干错误答案 (诱答)。

② 通过分析受测者选择的诱答项目, 可以获得有用的诊断信息。

20

Li, J. School of Psychology, Beijing Normal University

编制原则



1) 选项之间相互独立, 避免逻辑上的包含关系; 有时需要列出“其它”选项。

2) 题干的表达要简洁、明确, 适合受测者的文化程度。

21

Li, J. School of Psychology, Beijing Normal University

- 3) 选项一般4个或5个为宜, 各选项长度尽量相等。
- 4) 正确答案的顺序要随机排列在各个位置上, 次数大致相等。
- 5) 选项要简短、干练, 每个选项中都重复遇到的字词放到题干中。
- 6) 若是找最合适答案, 应强调“哪个答案最合适”。
- 7) 选项的形式统一, 例如都是人名、时间等。
- 8) 少用诸如“以上都不对”、“以上都对”的选项。

22

Li, J. School of Psychology, Beijing Normal University

诱答的编制方法——似是而非



① 应阐明不知道正解的受测者会选择该诱答的理由

- ✎ 使用最常见的错误
- ✎ 能反映出受测者缺乏某种知识或技能
- ✎ 使用那些在口语中与题干相联系的词
- ✎ 大众认可但实际错误的概念或者部分正确的句子
- ✎ 使用书面语言或者其他具有真理性的措辞
- ✎ 诱答在内容上与正解同质或者相似
- ✎ 诱答与正解在长度、结构和内容的复杂度上相似

23

Li, J. School of Psychology, Beijing Normal University

匹配题 (matching item)



① 分左右两列呈现前提条件 (premise) 和反应选项 (response), 要求受测者以最佳方式将两者对应关联起来。

② 儿童可以划线连接, 成人可以直接标注编号。

③ 反应选项的数量应该多于前提条件数量, 以避免受测者的猜测。或者在指导语中声明每个反应选项可能对应多个、单个或零个前提条件。

④ 注意: 每个前提条件只能对应一个反应选项。

24

Li, J. School of Psychology, Beijing Normal University

排序题 (rearrangement item)



- ① 受测者需要按照事先确定的固定类别重新排列一组选项。

Li, J. School of Psychology, Beijing Normal University

2.3.4 填空题/简答题



(completion item / short-answer item)

- ① 让受测者对用直接疑问句表述的问题用几个字 (填空题) 或几句话 (简答题) 进行回答的一种题型。



✎ 不适合测量复杂的教育目标

✎ 可能存在一个以上的正确答案, 因此评分并非完全客观

26

Li, J. School of Psychology, Beijing Normal University

编制原则



- ① 宜用问句, 若为填空, 空格尽量在句尾。
- ① 对不完整的答案, 应事先规定评分标准。
- ① 若为填空, 一题的空不宜过多, 且需填答的内容应该是关键词句。

27

Li, J. School of Psychology, Beijing Normal University

编制原则 (续)



- ① 答案要明确、唯一, 简短而具体。

- ① 题目要对答案范围加以明确限制, 例如指明数字型答案的单位, 但不能给予提示。

28

Li, J. School of Psychology, Beijing Normal University

2.3.5 论述题 (essay type test item)



- ① 包括论述题和作文题, 能对受测者的知识、能力全面考察, 能测量高层次的认知目标, 考查受测者的整体能力。

✎ 优点:

- ① 答案能够反映考生解答问题的思维过程
- ① 可以促使受测者重视所有知识之间的内在联系
- ① 编题花费的时间较少
- ① 减少受测者仅靠猜测答对的可能性

29

Li, J. School of Psychology, Beijing Normal University

✎ 不足:

- ① 评分缺乏客观性
- ① 对整个教育内容的取样缺乏广泛的代表性
- ① 易高估擅长言辞但学科知识贫乏者的实力
- ① 阅卷时费时费力

Li, J. School of Psychology, Beijing Normal University

编制原则

- ①能不用，则不用
- ①若用，出“活题”
 - ↳ 不出偏题、怪题，不脱离考试的范围和要求
- ①所有受测者接受同样的论文题，没有选择作答的机会
- ①一个论文题所占比重适当，不宜太小，也不宜太大，数量以2-3个为宜
- ①编制题目的同时，制定明细合理的评分标准

31

Li, J. School of Psychology, Beijing Normal University

2.3.6 操作题

①让受测者进行实际操作，如画图、走迷津、装配物体等，介于纸笔测验和实际活动评价之间的测验方式。

↳ 四项基本原则

- ①明确测量目标并将其操作化
- ①选择真实性程度高的项目
- ①指导语要简明扼要
- ①制定评分细则和计分方法

32

Li, J. School of Psychology, Beijing Normal University

第三节 测验的编排与组织

③3.1 合成测验

③3.2 预测与项目分析

33

Li, J. School of Psychology, Beijing Normal University

3.1 合成测验

1) 选择试题形式

↳ 考虑测验的目的和材料的性质

①再认 vs. 回忆 vs. 综合

↳ 考虑接受测验的团体的特点

①对有言语缺陷的人（聋哑/口吃）尽量采用操作测验

↳ 各种实际因素：个体 vs. 团体

34

Li, J. School of Psychology, Beijing Normal University

2) 审定题目：自审 & 他人审

↳ 各类题目比例是否适当（应与双向细目表的计划一致）；

↳ 题目是否表述清楚；

↳ 题目是否适合施测对象和施测条件；

↳ 难度和区分度是否恰当（要结合测验用途）；

↳ 题目是否相互独立，有无重叠题目；

↳ 题目数量应为最终题本的2倍或更多，以备筛选和编制复本。

35

Li, J. School of Psychology, Beijing Normal University

3) 测题的编排

↳ 基本原则

①难度排列应由易到难。

①同类型的题目放在一起，统一说明。

①根据题型本身特点调整排列顺序。

36

Li, J. School of Psychology, Beijing Normal University



常见的排列方式

- ① 并列 - 直进式：按题目的性质归为若干个分测验，在同一分测验内部按题目难度由易到难排列。
- ② 混合螺旋式：将各种类型的题目按难度划分为若干个层次，然后将同等难度、不同性质、不同类型的题目组合在一起，再按难度水平从易到难排列。

37

Li, J. School of Psychology, Beijing Normal University



3.2 预测与项目分析

1) 预测：目的在于获得受测者对题目如何反应的资料，发现未曾设想的情况，例如题意不清或引起误解。

- ↳ 预测对象来自正式测验准备应用的群体
- ↳ 预测的施测过程与情境应接近正式测试
- ↳ 时限可以稍宽，使每个受测者都能做完
- ↳ 随时记录受测者的反应情形

38

Li, J. School of Psychology, Beijing Normal University



2) 项目分析：量的分析和质的分析

↳ 难度、区分度、选项分析

- ① 为了保证项目分析的准确性，在一次项目分析，项目修订后，应取来自同一总体的另一样本再测一次，并进行第二次项目分析，比较两次结果是否一致。

39

Li, J. School of Psychology, Beijing Normal University



第四节 测验的标准化

④4.1 测验内容

④4.2 施测过程

④4.3 测验计分

④4.4 测验分数的解释

40

Li, J. School of Psychology, Beijing Normal University



4.1 测验内容

- ① 对所有受测者施测相同或等值的项目。
- ① 测验内容不同，所得结果的比较没有意义。

41

Li, J. School of Psychology, Beijing Normal University



4.2 施测过程

- ① 相同的测验情境：统一采光、桌椅、布置等，防止噪音和无关人员的干扰
- ① 相同的指导语：测验目的，如何作答，不常见的题型需要提供例题
- ① 相同的测验时限：
 - ① 人格测验相对宽松
 - ① 能力测验时限严格，其时限一般由预测决定（一般设定为90%的受测者完成全部项目的时间）

42

Li, J. School of Psychology, Beijing Normal University

4.3 测验计分 (scoring)



① 客观题计分过程应注意:

- ① 及时而清楚地记录反应情况;
- ① 要有一张记录正确反应的表格, 即记分键 (scoring key / scoring stencil);
- ① 将受测者的反应和记分键比较, 对反应进行分类。

① 对于主观题, 两个或多个评分者的评定应一致 (一般要求相关系数达到0.9以上)。

43

Li, J. School of Psychology, Beijing Normal University

4.4 测验分数的解释



① 对分数的解释必须标准化, 如果同一个分数可做出不同的推论, 测量便失去了客观性。

① 原始分数本身无意义

① 需要建立参照系统, 例如常模参照

① 将原始分转换为导出分数, 例如百分位常模、标准分常模等, 并严格根据要求对测验结果进行解释。

44

Li, J. School of Psychology, Beijing Normal University

第五节 测验的鉴定与使用说明



① 信效度检验

① 编写测验说明书

45

Li, J. School of Psychology, Beijing Normal University

5.1 信度和效度的检验



① 可先在小范围试用, 考察信效度, 初步确定可用后再进行大范围施测, 检验信效度。

① 信度和效度的评估方法多样, 具体应根据测量的性质和目的而定。

46

Li, J. School of Psychology, Beijing Normal University

5.2 编写测验说明书



1. 测验的目的和功用。
2. 测验编制的理论背景以及测验中的材料是根据什么原则、应用什么方法选择出来的, 许多手册还提供选择题目的统计指标。
3. 如何实施测验的说明。
4. 测验的标准答案和计分标准。
5. 常模资料, 包括常模表、常模适用的团体及对测验分数如何做解释。
6. 测验的基本特征, 包括难度、区分度、信度、效度和因素分析的资料, 以及这些资料取得的条件和情境, 包括调查的样本和时间。

47

Li, J. School of Psychology, Beijing Normal University



电 邮: JianLi@bnu.edu.cn

办公室: 后主楼1516



Thank You !



School of Psychology

Beijing Normal University