

## 第四节 项目功能差异与测量偏差



①项目功能差异（differential item function, DIF）是指不同群体对同一项目的答对概率（或得分率）不同，即不同群体在同一项目上得分存在差异。

- ↳可能源于群体间本身水平存在差异（良性DIF）；
- ↳也可能由其它原因（如文化、性别、地域、种族、职业、年级、SES）所致（不良DIF）。

②良性DIF不存在测量偏差，不良DIF存在测量偏差（test bias）。

1

Li, J. School of Psychology, Beijing Normal University

### ①一致性DIF & 不一致性DIF

↳一致性DIF是指受测者的能力水平与其组别之间不存在交互作用，即在所有的能力水平上，一组受测者回答某一项目的正确率都大于另一组。

↳当受测者的能力水平与其组别之间存在交互作用时，则表现为非一致性DIF。

2

Li, J. School of Psychology, Beijing Normal University

## DIF及测验偏差侦查的一般步骤



### （1）受测者群体的确定及数据资料的准备

↳根据测验特征，确定哪两类受测者群体可能会存在DIF。然后再随机抽取这两类受测者群体的测验反应数据。所抽取的受测者总量应足够多（一般不少于1000人），但也不能太多（一般不多于2000）。

3

Li, J. School of Psychology, Beijing Normal University

### （2）目标组和参照组的确定

↳根据项目特征，确定项目对以上确定的哪类群体有利，若第（1）步确定为男生和女生两个群体，且认为项目可能对男生有利，那么一般把男生样本作为参照组，女生样本作为目标组。

4

Li, J. School of Psychology, Beijing Normal University

### （3）匹配变量的确定

↳一般以测验总分作为匹配变量，将两个群体中相同测验总分的受测者一一匹配，采用适当的DIF侦查方法进行分析，找出并剔除存在DIF的项目，从而组成一个不含DIF的子测验。然后再以该子测验的总分作为匹配变量，再对子测验进行DIF分析，找出并剔除存在DIF的项目，如此反复，直到找到一个不含DIF项目的子测验，并以受测者在该子测验上的得分作为最终的匹配变量。

5

Li, J. School of Psychology, Beijing Normal University

### （4）DIF的探索性分析

↳根据第（3）步确定的最终的匹配变量对测验的所有项目进行正式的DIF分析。

### （5）DIF成因分析

↳成立专家小组对项目产生DIF的原因进行分析，并确定哪些项目是真正存在偏差（bias），即存在不良DIF。这需要学科专家、所测特质研究专家、测量学专家多方面人员共同参与完成。

6

Li, J. School of Psychology, Beijing Normal University

## DIF侦查的常用方法

分为实际得分方法和潜在特质方法。

- 实际得分方法包括散点图法、卡方方法、MH法、标准化方法（STND）、SIBTEST方法、逻辑斯蒂回归方法等。
- 潜在特质方法包括采用验证性因素分析和项目反应理论两类方法。

7

Li, J. School of Psychology, Beijing Normal University

## 例：MH方法（Mantel - Haenszel Procedure）

用于侦查0、1记分项目的DIF，以测验总分作为匹配变量。MH方法统计量的计算建立在一张 $S \times 2 \times 2$ 的列联表中，其中S是测验总分的水平数（由研究者根据需要自行确定），对于其中任一水平K，可得出两子群体得/失分的 $2 \times 2$ 列联表。

群体	题目得分		总计
	1	0	
参照组	$f_{1rk}$	$f_{0rk}$	$n_{rk}$
目标组	$f_{1jk}$	$f_{0jk}$	$n_{jk}$
总计	$n_{1k}$	$n_{0k}$	$n_k$

8

Li, J. School of Psychology, Beijing Normal University

根据S个列联表，计算如下统计量：

$$\alpha_{MH} = \frac{\sum_{k=1}^S \frac{f_{1rk} \cdot f_{0jk}}{n_k}}{\sum_{k=1}^S \frac{f_{0rk} \cdot f_{1jk}}{n_k}}$$

群体	题目得分		总计
	1	0	
参照组	$f_{1rk}$	$f_{0rk}$	$n_{rk}$
目标组	$f_{1jk}$	$f_{0jk}$	$n_{jk}$
总计	$n_{1k}$	$n_{0k}$	$n_k$

$\alpha_{MH}$ 的取值介于0至正无穷之间， $\alpha_{MH}=1.0$ 时，表示该研究项目无DIF； $\alpha_{MH}<1.0$ 时，表示研究项目对目标组有较低难度； $\alpha_{MH}>1.0$ 时，表示所研究项目对参照组有较低难度。

9

Li, J. School of Psychology, Beijing Normal University

但是由于 $\alpha_{MH}$ 的计算来自于样本数据，因此对其值是否等于1.0必须进行统计检验。

检验统计量是 $MH\chi^2$ ，其计算公式为：

$$MH\chi^2 = [ \sum_{k=1}^S f_{1rk} - \sum_{k=1}^S E(f_{1rk}) - 0.5 ]^2 / \sum_{k=1}^S Var(f_{1rk})$$

其中：

$$E(f_{1rk}) = n_{1k} \cdot n_{rk} / n_k$$

$$Var(f_{1rk}) = n_{1k} \cdot n_{0k} \cdot n_{rk} \cdot n_{jk} / [ n_k^2 (n_k - 1) ]$$

10

Li, J. School of Psychology, Beijing Normal University

## 例：性别DIF（SIBTEST法）

Schutte于1998年编制的Emotional Intelligence Scale（EIS）自陈问卷共有33个项目，采用5点记分，包括感知情绪、调控自我情绪、调控他人情绪、运用情绪等4个维度。使用华南师范大学王才康修订中文版，比较男性和女性的DIF。

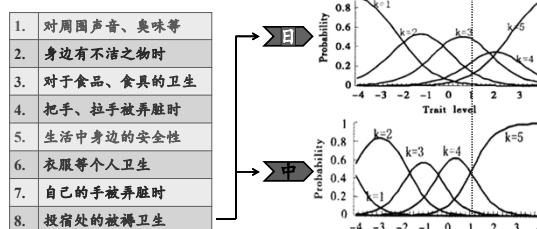
- 我期望能够做好自己想做的多数事情
- 我难以理解别人的肢体语言（\*）
- 我会去寻找一些让自己开心的活动
- 当别人消沉时，我能够进行帮助，使他感觉好些

11

Li, J. School of Psychology, Beijing Normal University

## 例：文化DIF（IRT方法）

OSHIBA 简易人格量表中“环境敏感性”维度，回答形式为“1 - 完全不介意”～“5 - 非常介意”，共5个等级，比较中日受测者。



12

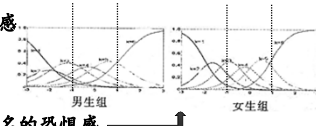
Li, J. School of Psychology, Beijing Normal University

### 例：性别DIF（IRT方法）



① 《新生适应性量表》中的“情绪适应性”分量表，共5个项目。回答形式为“1-非常符合”~“6-非常不符合”6个等级，得分越高表示被试的情绪适应水平也就越高，比较男生和女生DIF。

1. 进入大学后我有失落感
2. 我的心情很浮躁
3. 在大学里我有些自卑
4. 进入大学后我有种莫名的恐惧感
5. 进入大学后我的心情愉快充实(\*)



13

Li, J. School of Psychology, Beijing Normal University

### 不良DIF的控制



- ① 保证题目所测心理品质与全卷所测心理品质完全一致，避免题目测量了测验之外的第二种构想。
- ① 题目语言规范、无歧义，避免使用方言、俚语等。
- ① 编制好的题目应广泛征求不同群体的意见。
- ① 在正式使用前，最好先进行小规模试测，并修改或删除存有不良DIF的试题。

14

Li, J. School of Psychology, Beijing Normal University