

## 思考题

1. 这篇文章分析的考试“复旦英语水平考试(FET)”是什么类型的考试？为什么？这项考试是标准参照还是常模参照？
2. FET 考试分为哪几部分？考试任务和题型有哪些类型？与传统考试相比，FET 有哪些优越性？
3. 作者对 4 个研究问题的回答分别用了哪些统计方法分析数据？
4. 现有数据分析的结果表明这次考试成绩的分布情况怎样？信度和难度如何？
5. 文章中的扁度是什么？扁度值的大小意味着什么？
6. 作者如何解释考生口试成绩高于其它三部分的成绩？这个结果还可能与什么因素有关？
7. 在双因素方差分析（性别×专业）的结果中，虽然性别主效应显著，然而图 1 的结果是否显示任何专业的男女生成绩都有显著差异？虽然专业主效应不显著，可是作者的事后多重比较结果是否显示各专业的学生成绩都不存在显著差异？这些结果能说明什么问题？
8. 作者使用探索性因子分析的目的是什么？因子分析的结果是否符合考试设计者的初衷？为什么？
9. 作者用了哪些方法来收集 FET 考试的效度证据？总体来讲，FET 考试的效度如何？作者是如何通过这项研究的结果来解释信度和效度的关系的？
10. 考试开发和考试效度研究有无先后顺序？如果有的话，应该遵循什么顺序？
11. 针对这项考试还可以做哪些后续研究？
12. 根据这篇文章的研究方法和分析结果，我们能对期末考试成绩做哪些数据分析？数据分析结果有哪些用处？
13. 这篇文章对您以后的相关研究有什么启示？