

语 言 测 试 的 信 度 与 效 度

刘忠见

(华北矿业高等专科学校 外事办,北京 101601)

摘要: 语言测试是语言教学的重要环节。就语言测试的两个最重要的标准信度和效度进行了较详尽的阐述,以期对语言教师的命题及评价测试更具科学性有所裨益。

关键词: 语言测试;标准;信度;效度

中图分类号: G 642.474

文献标识码: A

文章编号: 1004-8154(2002)01-0073-02

语言测试在教学中的主要目的是对学生所掌握的知识 and 所具有的能力进行评价。然而,测试能否是客观的、公正的、甚至是权威的评估,却取决于测试的质量。谈到测试的质量,必然在涉及到测试的标准。那么,测试的标准是什么呢?国内外有许多语言学家都作过论述。他们认为,评估语言测试的主要标准有:信度、效度和可行性,其中信度与效度是极其重要的两个标准。换言之,一个测试如果失去了信度与效度,其它标准便无从谈起。本文拟就信度与效度作一阐述,以便对语言测试有较深入的了解。

一、信度(Reliability)

1. 信度的概念

信度也称可靠性。即当被测量对象本身发生变化,用同样的“尺子”去重复测量时,总是获得类似的结果。因此,信度也被称为一致性(consistency)。

如果说一个测试的信度高,便是指一个考生的成绩序列经反复测试都大致相同。如果说某次测试完全可靠,那么便指这次测试排除了一切误差,绝对的准确。即一个考生的成绩经反复测试后完全保持一致,一组考生经反复测试后,其成绩序列完全相同。其实,绝对的准确是不可能的,因为人的因素是不稳定的。

2. 信度的评估

信度的评估一般用考试结果的相关系数来表示。相关系数等于1表示该卷完全可靠,相关系数等于0则说明该试卷完全不可靠。在实践中,人们往往根据具体情况对试卷的信度提出不同的要求。通常是客观题的信度高于主观题。客观题的信度系数一般定在0.99以上。如果一份试卷既有客观题又有主观题(如:EPT),信度系数最好不低于0.80。在英语测试中,一份好的试卷,其词汇、结构和阅读部分的信度系数一般应在0.80~0.89之间,口试的信度系数一般在0.70~0.79之间。

对信度系数的评估方法主要有三种:再测信度、平行试

卷信度和内部一致信度。

(1) 再测信度(Test-Retest)。这种方法是指用同一份试卷在相同的条件下对同一批考生在不同的时间内进行两次测试。两次考试结果的相关系数则是再测信度。

(2) 平行试卷信度(Parallel Forms)。平行试卷信度(也可译为复本信度的)是对再测信度的改进,它是指用两份或几份在格式、内容、难度和题目数量等方面都平行、一致或等值的试卷,在不同的时间内对同一批考生进行测试,两次测试结果的相关系数便是平行试卷信度。这种计算信度的方法与计算再测信度的公式相同。

这种方法虽然不是前一次测试的重复,其测试的内容不同,并且两次测试的时间间隔可以很短(一般在两周内),但在命题时要使两套试卷完全等值是非常困难的。

(3) 内部一致信度(Internal Consistency Reliability)。这种方法使用的最广泛。它是指应用每个题目的方差来评估整份试卷的信度。

3. 影响信度的因素

影响信度的因素主要有:(1)考生水平的差异。考生水平越接近,试卷信度越低,考生水平差异越大(标准差也往往很大),信度越高。(2)题目数量。试题数量越少,信度越低,数量越多,信度越高。(3)题目的区分度。题目区分能力强的越多,信度越高,反之亦然。通常,命题人员用提高试题题目的区分度和增加试题题目数量提高信度。当然,试题题目数量也不能无限的增加,因考试时间是有限的。所以,为了提高考试信度,还必须考虑测试质量的另一个极其重要的标准——考试的效度。

二、效度(Validity)

1. 效度的概念

考试的效度(有效性)指考试是否测量了它所测量的东西,是否达到了它所预定的目的。效度是一个相对的概念。换言之,一个考试的效度可有高低之分,但不能简单地

收稿日期:2001-03-31

作者简介:刘忠见(1963-),男,华北矿业高等专科学校副教授,1995年和1997年曾两次赴美国、加拿大作短期校际交流和业务培训。

说全部有效或全部无效。除非给有效与无效规定效度系数的界限。这主要是因为个人或群体的语言能力特征,只能通过其行为样本间接推测,而不是绝对有效。从这个意义上讲,效度也可以说是一个程度上的概念,它反映根据考试分数作出推论或预测的准确性程度。

2. 效度的评估

一般地说,效度可以从以下几个方面评估。

(1) 内容效度(Content Validity)。内容效度是指考试的内容是否具有代表性和综合性,或者说是否考了应考的内容。一份试卷往往不可能包括所有要考的内容,所以,选择内容的方法非常关键。例如:有30个题目是考英语语法的,而有20个题目是考主谓一致性方面的,这就很难完全反映出考生的语法能力。因为英语语法覆盖面很广,还包括动词时态、动词短语、介词短语、不定式短语、分词短语,等等。所以这样的考试,其内容效度必然很低。

内容效度的确定,一般不靠统计手段,而是命题人员或审题人员对试卷的内容、题目的难易度、区分度等进行严格的分析。

(2) 效标关联效度(Criterion-Related Validity)。效标关联效度又可分为预测效度(Predictive Validity)和共时效度(Concurrent Validity)。

预测效度是指考试的结果和预言是否有效。例如,要通过考试选拔学生进一步深造,该考试是否选拔了应该选拔的学生,有没有选错或选漏?这就要看考试起到了预测作用没有。一份具有很好的预测效度的试卷,应该能够正确地预言学生未来的行为。例如:把学生高考的英语成绩和大学一年级末的英语成绩作比较,观察其相关性,就可以了解高考英语试卷的预测效度。有些考试(如:学能考试、分班考试、水平考试)由于与将来的学习有关,所以应特别注意预测效度。预测效度的评估可用计算再测信度所用的 Pearson 的积距率公式来计算其相关系数。

共时效度是用来将新的考试和已经公认的考试作比较,以便证明新的考试的效度。例如:在较短的时间内(一般不超过两周)让同一组学生参加 EPT 和 TOEFL 考试,如果考试结果的相关系数非常高,说明两者的相关性很高。如果说 TOEFL 是一个公认的标准化考试,那么 EPT 也是一个标准化考试。共时效度的评估仍然可用 Pearson 的积距率法来计算相关系数。

共时效度与预测效度的关系是:两者都以某种独立的而且可靠的效标作为参照量,把所测试的分数与效标分数作比较,计算其相关程度。预测度和效标是考生将来的成绩,共时效度的效标是另一次标准化考试。因此,预测效度和共时效度也可统称为效标关联效度。两者的区别就在于测试时间上的差异。共时效度要求相比较的两次考试由同一组考生在同一天或时间相隔很近(一般不超过两周)的情况下完成。预测效度则要求相比较的两次考试的间隔是半年、一年甚至更长时间,其测试对象是同一批考生。

(3) 结构效度(Construct Validity)。结构效度指一个考

试所测量的能力是否符合语言、语言学习和语言行为理论中所假设的能力。这里说的结构是任何关于语言理论中所假设的能力或特征。这些理论可以是心理学方面关于语言学习或语言习得的理论。结构效度可以帮助建立心理学或语言理论中的假设,也可以推翻这些理论中的假设。对于结构效度的评估,也必须采取统计手段。鉴于手段复杂,一般都由计算机进行运算。加上篇幅所限,故在此不作介绍。

3. 影响效度的因素

影响效度的因素主要有:(1)题目和试卷的指令不清楚。这样考生会因为考试要求不明确而降低成绩。(2)题目的难度水平不合适。换言之,没有在合适的水平上测试所要测试的内容。这样,一份过易或过难的试卷会把中等和上等的考生放在同一水平上。(3)选择项目模棱两可,含糊不清。这与上述(1)中的影响一样会使考生不明确意义而降低成绩。(4)题目与测试目的不符或没有足够的题目数量来代表要测试的内容。(5)题目开始就很难,致使考生一开局就严重受挫,从而导致后面的题目也能放弃。(6)在多项选择里,正确答案很容易被认出来。

三、信度与效度的关系

一方面,信度与效度均可通过一系列(其中也有相同的)方法来测量,有时两者可以是互补的。另一方面,信度高不一定就效度高,然而效度高却必然信度高。或者说,信度只是必要条件,而不是充分条件。以我所见,信度是效度的前提,效度比信度更重要。

四、结束语

语言测试是语言教学的重要环节,系属狭义应用语言学(教学法)所探讨的具体内容。针对语言测试中两个重要的标准,笔者从宏观角度进行了阐述。我认为,一旦从宏观上掌握这两个标准,评估语言测试时,就不再只凭经验或印象,而是基于科学的依据——测试内容和方法应符合测试的目的要求,测试成绩要稳定一致,即信度高。可见,了解、掌握它们,对语言教师命题、评价测试具有多么重大的指导意义。

参考文献:

- [1] 桂诗春. 标准化考试——理论、原则与方法[M], 广州:广东高等教育出版社,1986.
- [2] Bachman, L. F. *Fundamental Consideration in Language Testing*. Oxford University Press, UK, 1990.
- [3] Butler Christopher. *Statistics in Linguistics*. Basil Blackwell Ltd., Oxford, UK, 1985.
- [4] Hughes Arthur. *Testing for Language Teachers*. Cambridge University Press, UK, 1989.
- [5] Weir, Cyril J. *Communicative Language Testing*. Prentice Hall International Ltd., UK, 1990.
- [6] Widdowson, H. G. *Teaching Language as Communication*. Oxford University Press, UK, 1978.