

高校师生对原始分制与标准分制的感知对比

□ 林 椿 肖云南

摘要: 我国高校课程考试很少采用标准分来报道分数。通过在高校师生中开展“关于在高校大学英语课程考试中采用标准分来报道分数的可行性”的问卷调查,发现师生对标准分的总体评价普遍较高,支持利用标准分推断学生原始分数在团体中所处位置,使用标准分的意愿也较强烈,并且支持对传统原始分制进行改革。同时发现,师生在实际的教/学的过程中不知如何根据标准分提供的信息来调整教/学的方法与内容,师生对原始分与标准分的感知在三个问题上存在显著性差异。本文通过调查结果分析了标准分难以在高校课程考试中推广的原因,从而提出可行的解决策略,藉此为进一步改革原始分制与完善标准分制提供一些参考。

关键词: 课程考试; 学生评价; 原始分; 标准分; 描述性统计; 独立样本t检验

中图分类号: G642.47

文献标识码: A

文章编号: 1672-0717(2017)03-0056-07

收稿日期: 2016-10-23

基金项目: 湖南省教育厅科学研究项目“基于动态评估的英语教育绩效研究”(15C1438); 湖南省哲学社会科学基金项目“基于认知诊断理论的ESL分级测试体系研究”(16YBA392)。

作者简介: 林椿(1984-),女,海南海口人,湖南大学外国语学院与国际教育学院博士研究生,讲师,主要从事语言测试研究;肖云南,湖南大学外国语学院与国际教育学院教授、博士生导师。

一、引言

学生的考试分数可为反映教学情况、奖惩评比、推荐选拔人才、用人单位考察学生素质等方面提供重要依据。原始分记分制与标准分记分制是两种比

较常用的记分制。使用哪种记分制公平合理地评价学生,是语言测试需要认真研究的课题。美国2014年版《教育与心理测量标准》继续规定了关于考试记分、分数可比性、分数报道等方面的标准,足以说明其对记分制的一贯重视^[1]。Bachman和Palmer在2010年提出了测试使用论证(Assessment Use Argument, AUA),其最大特点是强调了考试的使用决策和使用影响,并利用使用决策和使用影响来指导考试项目的开发和效度验证^[2]。考试为决策提供依据^[3],而决策的依据是考试的分数,因此科学合理的记分制极为重要^[4]。科学合理的教育测量所提供的考试分数能对教育现状给予精确客观的描述,为教育提供丰富的反馈信息,供教育者分析教育中的优点和不足,改进教育方法,提高教育质量,从而更好地培养人才。这与“教育是培养人”的理念标准相一致,由此可见科学合理的记分制的重要性^[5]。

在不同记分制对考生排名的影响方面,有的研究显示考生按原始分和按标准分排列的名次无显著差异^[6];而有的研究显示考生按原始分和按标准分排列的名次是不相同的^[7-9]。在采用哪种记分制来报道学生成绩方面,有的研究者推荐对原始分数进行标准转化后再报道学生的成绩^[10-13];而有的研究者建议综合采用原始分和标准分来报道学生的成绩^[8]。Mahon提议作文成绩单上要同时报道原始分数与相应的字母成绩等级(Letter grade)^[14]。美国学校是很少直接使用考试原始分数来报道学生分数的,他们大量使用量表分数,其中最常用的是标准分数和百分位等级,此外还有标准九分数、四分位等级等^[15]。国内外著名的大规模标准化英语能力考试采用标准分记分制的有:中国大学英语四六级考试、中国汉语水平考试、美国研究生入学考试、美国大学水平考试、托福、雅思等。为保证考试的信度和效度,大规模标准化考试采用标准分或者转换后的标准分来报道分数,确保了测量的精度和公平性。这使得考试的分数不但可以解释,而且可以准确地

推论出考生的语言水平^[16]。

科学合理的记分制是确保测试可靠、公平、有效的关键。原始分与标准分优点各异,原始分直接反映了考生对教学内容的掌握程度,而标准分反映了一个原始分数在团体中所处的位置,使不同科目、不同时间的考试分数具有可比性,并能提供更准确的统计信息。前人的研究大多是关于原始分记分制与标准分记分制的对比研究^[17-18],关于师生对原始分与标准分的感知差异的调查研究很少。尽管标准分有很多优点,然而现实中高校课程考试很少采用标准分来报道分数。针对这一现象,本研究在国内某一高校的师生中展开了“关于在高校大学英语课程考试中采用标准分来报道分数的可行性”问卷调查,旨在找出教师与学生对原始分与标准分的感知是否存在差异,希望为进一步改革原始分制与完善标准分制提供一些参考。

二、研究方法

1. 研究问题

本研究拟研究以下问题:

第一,教师与学生对原始分与标准分的感知情况如何?

第二,教师与学生对原始分与标准分的感知是否存在显著性差异?

第三,标准分难以在高校课程考试中推广应用的原因是什么?

2. 研究对象

本研究的调查对象为天津某高校的教师与学生。我们随机邀请120名师生进行问卷调查,在收回的有效问卷中,按照师生1:1的比例,随机选取共100份(教师50份,学生50份)有效问卷进行统计分析。随机抽样对象兼顾到不同院系、不同专业的教师与学生。

3. 调查问卷

进行问卷调查之前,我们对调查对象进行了非正式的、历时大约5分钟的简短培训,简单介绍原始分与标准分的概念,以及这两种记分制的优缺点。调查问卷总共设计15个问题,第1~13题为单选题,选项都采用莱克特五级量表(Likert Scale)格式;第14题为多选题,提供四个选项;第15题为提供四个选项的半开放式问题。问卷调查的目的是了解师生对原始分与标准分的看法,找出教师与学生对原始分与标准分的感知是否存在显著性差异,以了解在高校大学英语课程考试中采用标准分来报道分数的可行性。

4. 分析方法

本研究采用SPSS22.0软件做描述性统计和独立样本T检验。数据分析首先采用SPSS22.0做描述性统计以了解教师与学生对原始分与标准分各项内容的打分的均值,然后做独立样本T检验以找出教师与学生对原始分与标准分各项内容的打分的均值是否存在显著性差异。

表1 教师与学生对问卷第1~10题的回答的统计结果

问卷问题	教师			学生		
	N	M	SD	N	M	SD
Q1. 培训前对原始分的了解程度	50	4.08	.853	50	3.86	1.143
Q2. 培训前对标准分的了解程度	50	3.48	.863	50	3.20	.990
Q3. 用原始分推断学生位置	50	2.32	.844	50	2.42	.950
Q4. 用标准分推断学生位置	50	4.12	.521	50	3.82	.596
Q5. 原始分的合理性	50	2.68	.741	50	2.48	.953
Q6. 标准分的合理性	50	3.64	.749	50	3.68	.587
Q7. 继续使用原始分的意愿	50	3.60	.808	50	3.32	.653
Q8. 建议使用标准分的意愿	50	3.80	.571	50	3.62	.602
Q9. 完全使用原始分制	50	3.60	.639	50	3.40	.728
Q10. 支持改革并使用标准分制	50	3.60	.495	50	3.56	.577
Q11. 能否接受标准分的缺点	50	3.96	.727	50	3.44	.812
Q12. 根据标准分调整教/学的方法等	50	2.68	1.269	50	2.38	.780
Q13. 标准分有积极反馈与促进作用	50	2.92	1.066	50	2.56	.577
对原始分制总体评价	50	16.2800	1.94873	50	15.4800	1.85428
对标准分制总体评价	50	18.6400	1.63832	50	17.8800	1.80295

三 结果讨论

1. 描述性统计结果

教师与学生对问卷第1~13题的回答的描述性统计结果(见表1):

调查结果显示,师/生对标准分的总体评价($M=18.64/17.88$)高于对原始分的总体评价($M=16.28/15.48$)。教师对原始分、标准分的了解($M=4.08$ 、 $M=3.48$)都比学生对它们的了解($M=3.86$ 、 $M=3.20$)多,而且师/生对原始分的了解比对标准分的了解都多;师/生都更加支持利用标准分推断学生原始分数在团体中所处位置($M=4.12/3.82$);师/生皆认为标准分($M=3.64/3.68$)比原始分更具合理性,而且学生更加认可标准分的合理性,不过教师群体对于这一问题的看法差异比较大($SD=0.749$)。师/生都更加愿意使用标准分($M=3.80/3.62$),师/生支持对传统原始分记分制进行改革并考虑使用标准分($M=3.60/3.56$)。

传统的原始分记分制一般不会出现负数或者小数点,但标准分即 z 分数有正负值(z 分数一般取值大于-3,小于+3,也可能出现“0”分即代表原始考分为所有原始分的平均分)、且有时候有小数点。笔者担心教师、学生、家长会误认为标准分为0分“就表示该生一点都不会,相当于交白卷”;担心他们会不理解标准分的这些缺点,从而不信任标准分及其理论。为此,针对标准分的这些缺点,问卷提出了问题11“传统原始分一般不会出现负数或者小数点,但标准分即 z 分数有正负值(z 分数大于负3,小于正3)且有时候有小数点,你能接受标准分这一形式么?”,调查结果显示80%的教师与48%的学生表示“完全能够”“能够”接受标准分的负值与小数点

形式,没有人表示“完全不能够”接受。问题12“你会根据标准分提供的信息(如平均分、标准差、原始分数所处位置等),调整教学(或学习)的方法与内容吗?”,调查结果显示64%的教师与56%的学生不知如何根据标准分提供的信息(如平均分、标准差、原始分数所处位置等)来调整教学(或学习)的方法与内容。问题13中“标准分能使学生了解自己在团体中所处的水平。它是否能够促使学生做好学习计划以促进自己综合能力的提高呢?”,48%的教师与54%的学生认为,标准分不一定能够督促他们做好学习计划以促进综合能力的提高。教师与学生对问卷第11~13题的回答的统计结果见表2。

调查表明,相比对原始分的感知,虽然师生对标准分有更高的总体评价,但在实际的教学与学习的过程中,教师(或学生)不知如何根据标准分提供的信息(如平均分、标准差、原始分数所处位置等)来调整教学(或学习)的方法与内容。同时,他们对标准分是否能够督促他们做好学习计划以促进综合能力的提高存在一些质疑。非正式访谈发现,师生都承认标准分能为教师的“教”与学生的“学”提供很多信息,但是师生都认为教师想要提高教学质量、学生想要提高综合能力,教师应在教学方法与内容上、学生应在学习方法与内容上改进与调整。

问题14“以标准分报道分数的形式,你认为较难理解的内容有哪些?(多选题)”,66%的教师认为已知标准分,使用正态分布函数表查学生排名 P 值比较难;72%的学生认为已知标准分,不需查表,如何使用软件计算排名 P 值比较难。问题15“你认为,教务系统采用何种形式来报道学生分数,更加科学、合理、可行呢?(半开放式问题)”的回答显示,52%的教师和50%的学生认为教务系统采用标

表2 教师与学生对问卷第11-13题的回答的统计结果

题号	分组	人数	完全能够(%)	能够(%)	不一定(%)	不能够(%)	完全不能够(%)
11	师	50	20	60	16	4	0
	生	50	8	40	40	12	0
12	师	50	20	0	16	56	8
	生	50	0	6	38	44	12
13	师	50	16	0	48	32	4
	生	50	0	2	54	42	2

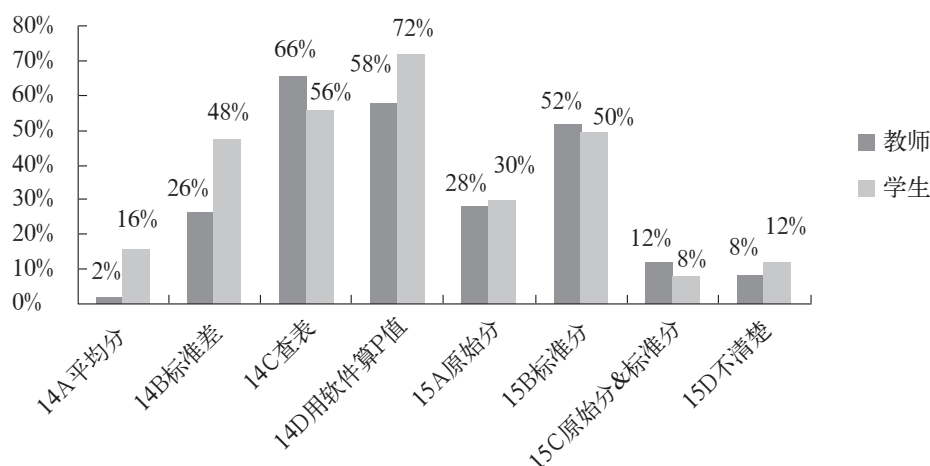


图1 师生对问题14和15的回答的对比分析

准分形式来报道学生分数更加科学、合理、可行。师生对问题14和15的回答的对比分析见图1。

2. 独立样本T检验结果

上节统计分析表明,教师与学生对原始分与标准分这两种记分制的感知存在差异,但是这种差异是否显著呢?为回答这一问题,需进行独立样本T检验。教师与学生对两种记分制的感知的独立样本T检验结果见表3。

研究发现,教师与学生对原始分与标准分的感知在三个问题(问题4、11、13)上存在显著性差异

($p=0.009$ 、 0.001 、 $0.038 < .05$)。问题4的回答显示,教师比学生更能根据标准分提供的信息推断考试分数在团体中所处的位置($M=4.12/3.82$),而且这种差异很显著。问题11的回答显示,教师比学生更能接受标准分的缺点比如负值与小数点形式($M=3.96/3.44$),而且这种差异具有显著性。问题13,关于标准分对学习的积极反馈与促进作用,虽然师生存在一些质疑,但是教师比学生更加肯定标准分对学习的积极反馈作用($M=2.92/2.56$),而且这种感知差异具有统计性意义。

表3 教师与学生对两种记分制的感知的独立样本T检验结果

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
1	1.091	98	.278	.220	.202	-.180	.620
2	1.508	98	.135	.280	.186	-.089	.649
3	-.557	98	.579	-.100	.180	-.456	.256
4	2.682	98	.009*	.300	.112	.078	.522
5	1.172	98	.244	.200	.171	-.139	.539
6	-.297	98	.767	-.040	.135	-.307	.227
7	1.906	98	.060	.280	.147	-.012	.572
8	1.533	98	.129	.180	.117	-.053	.413
9	1.460	98	.148	.200	.137	-.072	.472
10	.372	98	.711	.040	.108	-.173	.253
11	3.373	98	.001*	.520	.154	.214	.826
12	1.425	98	.157	.300	.211	-.118	.718
13	2.100	98	.038*	.360	.171	.020	.700

* $p \leq 0.05$.

3. 标准分记分制难以推广的原因

问卷调查了解到高校师生对标准分记分制持肯定态度,但是在实际高校课程考试中,标准分仍较难以推广应用。真正影响标准分实施的问题是什么?调查结果显示,首先,关于能否接受标准分的缺点比如负值与小数点形式(问题11),虽然多达64%的师生表示“完全能够”或者“能够”接受,但是仍然有36%的师生表示“不一定”或者“不能够”接受;关于标准分制哪些内容较难以理解(问题14),65%的师生认为“如何使用软件计算排名P值比较难”,说明师生对标准分的计算方法不太熟悉。标准分有自身的缺点,比如标准分在分数形式上有正值且可能出现负值、小数点;标准分的计算方法需要学习才能掌握且计算过程相对麻烦;标准分不能直接反映教学目标达成程度。这是影响标准分实施的內因。其次,关于培训前对标准分的了解程度(问题2),虽然共有58%的师生表示“非常了解”、“了解”“一般了解”,但是仍有42%的师生表示“不了解”“根本不了解”。很多教师和教学管理人员不理解标准分,这给标准分的推广增加了难度。标准分的满分不是100分,这与传统的记分制很不相同,很多教学管理人员和教师以标准分不便于记入学生成绩档案为由而不使用标准分。这是影响标准分实施的外因。

标准分面临着难以在高校课程考试中推广应用的问题,那到底还要不要使用标准分?答案是肯定的。标准分有很多优点,比如:能反映考试分数在团体中的位置,使不同科目、不同时间、难度不一的考试分数具有可比性、可加性,无需考虑试卷的难易程度及考分的离散程度;更加稳定、科学、合理;能提供更多、更准确的统计信息。在百分制中,不同科目考试之间分数可比性较差,而标准分可以比较不同单位分数^[20]。大规模标准化考试一般不采用原始分(即考生的卷面得分)来报道考生成绩,因为这样做分数的信息量太小,而通常进行一系列的分数转换,使最后报道的分数不但信息量大且具有可解释性^[16]。标准分给人才选拔与决策提供更多的信息,若我们的课程期末成绩用于保送研究生,如果用标准分显示就可以更精确地比较同年级学生的考分所处位置,更好地评价学生,择优推荐。当然,标准分也可以通过建立模型,在形式上设计

为百分制,但是意义上不同于百分制的原始分了。

4. 讨论与建议

针对标准分难以在高校课程考试中推广应用的问题,结合此次调查结果并参考1994年中华人民共和国教育部颁发的标准分数制度实施方案^[21],本研究提出以下建议:(1)明确告知教学管理人员、教师、学生高校课程考试建立标准分制的目的,以及标准分制的基本内容,并由教学管理人员、教师负责标准分的实施。高校课程考试建立标准分制的主要目的有三点:实现考试分数标准化,克服原始分数的局限性,使考试分数更科学、准确地反映学生水平;为高校研究生招生录取提供更科学、准确的依据,方便录取;为进一步开发利用考试信息,充分发挥考试的作用,建立标准分制既有利于高校教学、又有利于用人单位选拔人才。(2)研究制定统一的转换软件等具体操作细则。专业人员需教会教学管理人员、教师如何更快捷计算标准分、如何更好地使用标准分。由于计算器、计算机的发展和普及,教师可借助工具计算得到原始分数的标准分;教学管理人员可请计算机专业人员在教务系统中编程录入概率密度函数积分公式,由计算机自动生成概率P值,在分数报道单上,教师与学生只要了解标准分Z值与概率P值的意义即可。已知随机变量标准分Z服从正态分布,计算标准分Z值的概率P(即考分位置)即对标准分Z的概率密度函数求积分,具体计算公式及公式解释可参见陶敏力论文^[22]。

(3)编制宣传材料,加强宣传工作,培训各级教学管理人员、教师。加大对标准分的宣传力度,能让教师、教学管理人员、学生进一步了解标准分的优点,使得教师、教学管理人员愿意逐步对传统原始分(百分制)记分制进行改革并科学合理地使用标准分。(4)采取实验、总结、推广的科学态度开展工作。参考全国大学英语四、六级考试中心给各校反馈的本校所有考生在某次大学英语四、六级考试中的成绩报告单形式,可在教务系统中同时用两种记分制来报道高校学生课程成绩。

考试有选拔功能,比如有些高校的教务部门根据学生课程成绩排名情况优先保送读研,根据排名提出推免生名单。采用标准分对学生课程成绩排名更加科学与合理,当然我们不提倡用平时成绩排名去打击成绩差的学生。人才评价,更需要人性的、人

文的关怀。

严格遵循决策程序,灵活运用科学的决策方式方法,是走向科学化的教育管理决策的途径之一^[5]。高校在人才培养过程中,要更多地利用其教育、激励功能,以便改进教学。标准分比原始分给人才选拔与决策提供更多的信息,标准分制是保证有效测试与有效决策的关键。

四、结语

本研究发现师生对标准分的总体评价都较高,都认为标准分记分制比较科学合理,支持对传统百分制的改革,愿意并建议使用标准分记分制。我们通过非正式访谈发现:师生都承认标准分能提供很多信息,教师认为要提高教学质量,仍然得在教学方法与内容上改革与调整;学生也认为要提高自身综合能力,仍然得在学习方法与内容上改进与调整。教育测量与评价研究的目的之一是要给语言教学与语言学习提供反馈与参考^[23]。对于一次考试,我们除了关心试卷本身的质量(其信度、效度以及各题的难度、区分度是否合乎要求),还关心考试的记分制是否科学与合理。有经验的教师总是根据考试中学生反馈的考试分数调整自己的教学,而有经验的学生也可通过考试分数认识自己的水平,改变学习方法、努力方向以及努力的程度。可见,考试反馈的考试分数影响着学生的学和教师的教,考试的记分制是否科学直接关系到教学质量的高低、学生的发展方向以及人才的合理选拔培养^[22]。

从世界知名大学的记分制可以看到级别分和标准分被广泛应用^[24]。哈佛大学对学生成绩的打分标准不是按绝对分数值(即原始分数)来打“优”或打“A、B、C”,而是将学生考分最为集中的那个分数段列为中等分数段即为B。也就是说,学生若想超过B,得到B+或是A-,则必须超过大多数人才行,因而有时候90分可能是A,有时候90分则又可能是B。在哈佛大学,学校只强调竞争的客观存在,却不强调学生的考分及名次,也从来不向学生公布其所在班上或系里的名次。这样做是为了使学生不至于出现自满或自卑的现象而影响学习。学生每次考试不只看分数高低,而能够客观地定位自己,并能够与以往的成绩定位作比较,只有自己的现在比

自己的从进步了,才算是真正进步^[25-26]。哈佛大学的记分制对学生的成绩观念有良好的引导,这也是标准分的一个优点所在^[27]。科学合理的记分法能对教育产生良好的促进作用。在标准化语言测试发展的今天,因为传统的原始分没有反映题目的难易度、区分度等重要的质量指标,提供的信息量很小,误差比较大^[4],所以高校应该采用更加科学合理的记分制比如标准分制来记分。

参考文献

- [1] Alderson, J. C., Clapham, C., Wall, D.. 语言测试的设计与评估[M]. 杨惠中, 导读. 北京: 外语教学与研究出版社, 2000: F23.
- [2] Bachman, L. F., Palmer, A. S.. Language Assessment in Practice: Developing Language Assessment and Justifying Their Use in the Real World[M]. Oxford: Oxford University Press, 2010: 25.
- [3] 李智, 肖云南. 基于论证的测试效度验证与高考英语效度验证研究[J]. 教育测量与评价(理论版), 2013(11): 57-62.
- [4] 杨惠中. 有效测试、有效教学、有效使用[J]. 外国语, 2015(1): 2-26.
- [5] 胡弼成, 彭湘韧. 论教育管理决策的科学化[J]. 湖南师范大学教育科学学报, 2014(5): 13-17.
- [6] Webber, S. A., Clinton, B. D.. Using z-scores to evaluate accounting students: a test of Weinstein's suggestion[J]. Journal of Accounting Education, 1999(4): 407-415.
- [7] 宋岭梅. 关于多科考试成绩评定的统计方法问题[J]. 教育研究与实验, 1984(3): 33-38.
- [8] Weinstein, G. P. Evaluation of accounting students[J]. Journal of Accounting Education, 1994(3): 193-204.
- [9] 张菁莉. CET评分标准改革分析[J]. 河南职工医学院学报, 2010(2): 210-212.
- [10] Thomas, C. R. Rational standard and ability adjusted standard transformed score models[J]. Educational and Psychological Measurement, 1985(4): 803-809.
- [11] Ebel, R. L., Frisbie, D. A.. Essentials of Education Measurement(5th ed.)(M). Englewood Cliffs: Prentice Hall, 1991: 289.
- [12] Ornstein, A. C. Grading practices and policies: An overview and some suggestions[J]. Nassp Bulletin, 1994(561): 55-64.
- [13] Cross, L. H. Grading students[J]. Practical Assessment, Research & Evaluation, 1995(8): 1-3.
- [14] Mahon, R. L. A Grading System for Composition Papers[J]. The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 2005(3): 102-104.
- [15] 杨志明. 中美教育考试分数的报告与使用之简要比较[J]. 教育测量与评价, 2008(7): 51-53.

- [16] 杨惠中,金艳.大学英语四、六级考试分数解释[J].外语界,2001(1):62-68.
- [17] 石伟,向远辉,黄玲娟,唐利强.标准分在高校课程考试中的应用[J].湘潭师范学院学报(自然科学版),2005(2):130-132.
- [18] 李国英.标准分与CET考试新记分方式[J].长春大学学报,2006(10):105-106.
- [19] 秦晓晴.外语教学研究中的定量数据分析[M].武汉:华中科技大学出版社,2003:19.
- [20] 黄玉石.使用标准分合理评定学生学业质量[J].辽宁工业大学学报(社会科学版),2004(4):34-35.
- [21] 中华人民共和国教育部.普通高等学校招生全国统一考试建立标准分数制度实施方案(教试厅[1994]1号)[OL].<http://www.chinalawedu.com/falvfagui/fg22598/34504.shtml>,1994.
- [22] 陶敏力.考试分数的解释[J].重庆师范学院学报(自然科学版),1986(3):99-103.
- [23] Bailey,K.M.Washback in language testing(TOEFL Monograph Series)[M].Princeton,NJ:Educational Testing Service,1999.
- [24] 王建平.从分数制到等级制:中考评价制度改革的审思[J].湖南师范大学教育科学学报,2014(5):24-27.
- [25] 肖云南,罗晓英.关于英语测试的改革[J].外语教学与研究,2002(4):294-299.
- [26] 肖云南.标准分的意义[OL].<http://blog.sina.com.cn/xiaoyunnan>,2010.
- [27] 别敦荣,贾杰静.哈佛大学的发展历程、教育理念及启示[J].大学教育科学,2011(6):76-81.

A Comparative Study of College Teachers and Students' Perceptions of Raw Score and Z-score

LIN Chun XIAO Yun-nan

Abstract: Z-score is seldom used to report students' performance in the college curriculum examination. By conducting among college teachers and students a questionnaire survey on the feasibility of reporting scores of college English course examination using z-score, it is found that generally teachers and students have a higher opinion of z-score, both show more support of using z-score to infer the position of raw score in the group, hold the view that z-score is more reasonable, present stronger willingness to use z-score, show more support for the reform of traditional centesimal system, meanwhile, teachers and students respond that in the actual teaching/learning process they do not know how to adjust their teaching/learning method and content according to the information provided by z-score; there are significant differences between teachers and students' perception of raw score and z-score in terms of three questions. By analyzing the feedback from the survey, the reasons why z-score is difficult to be popularized in reporting scores of college course examination can be found, then feasible solutions are proposed accordingly, thereby providing some hints for further reforming the raw score system and improving the z-score system.

Key words: raw score; z-score; descriptive statistics; independent samples t-test

(责任编辑 陈剑光)