

(上接第一版)

告内容包括每个受测考生的总成绩和子分数,以及解释性信息;而收费版分数报告包括:试题信息报告(逐题分析考生表现)、按个体需求形成个性化报告、自定义参考组分析报告等。

【国内发展现状】

近年来,国内教育考试机构积极开发考试评价服务,北京、上海、浙江等省市已开展相关研究和实践。

1. “云海工程”。2011年,教育部考试中心在云南和海南两省试点开展“云海工程”,其中一项重要内容就是为每个考生提供学生版“成绩分析报告”。报告主要包括三大内容:一是各科考试标准分成绩、

综合分成绩和百分等级,二是考生在各科考查的学科内容和能力结构上的分项表现,三是升学指导测验结果。



2. 利用高中学业水平考试成绩研发的个人报告。上海市教育考试院近年来研发了“高中学业水平考试个人成绩报告”(含“等级考”和“合格考”),

从各科目的“知识内容”和“能力目标”两个维度对学生在各分项上的得分率和百分位进行评价,分为中文和英文两个版本。接下来还将在推行学考机考的过程中,积极应用大数据对考生进行动态性和过程性评价。浙江省教育考试院近年来也推出考生“高中学业水平考试个性化成绩诊断报告”(含“学考”和“选考”),包含考生的试题得分情况、学科知识掌握情况、学科能力情况、考核目标达成情况四个方面,并划分五个等级来反映考生个人在全体学生中的相对水平。

3. 利用中考数据开发分数报告。青岛市招生考试

院着眼于中考数据于2017年开始研发面向初中毕业生的网络版诊断性分数报告——“青岛市初中毕业生学业水平考试成绩报告单”。诊断性分数报告包括整体报告和学科报告两个部分,整体报告包含学科学业发展评价和综合能力与素养发展评价两个维度。学科报告包含9个学科,每个学科包含知识领域评价和能力层级评价两个维度。郑州市教育局开发了“学业评价分析报告系统”,向初中生和高中生提供日常学业测试的评价报告。报告包含学生考试整体报告及各科的分析报告,考试整体报告涉及学生的整体成绩概况、学业等级、优劣势科目分析等,单科分析报告涉及成绩综述、卷面作答情况、卷面得失分情况、知识点掌握情况、认知水平情况等,并为学生提供了对应的提升技巧。

4. 自学考试领域的评价服务。北京市教育考试院于2009年首次在自学考试领域开展诊断性评价服务,为“计算机应用基础”课程考生提供个人成绩分析报告,报告包括考生学习上的优势、不足和建议等,在全国高等教育自学考试尚属首次。在此基础上,又向2010年“英语水平考试(一)(笔试)”课程的全部考生提供了成绩分析报告。考生不仅能看到自己的总分,还能得到相关课程的考核标准、试题考核点及个人知识掌握状况和专家建议。

【发展趋势】

当前国内外考试机构面向学生开发的评价服务主要体现出三个方面的发展趋势:一是维度更为多元。从分数、百分等级等基本描述,到对能力结构、学科素养等潜能的评价,逐步向错因分析、学习建议、发展规划等深层次认知诊断发展。二是数据更为可靠。既有终结性评价数据,又有过程性评价数据;既有学业数据,又有非学业数据;既有结构化数据,又有非结构化数据,实现对考生考试测量情况的精准深入分析。三是呈现更为智能。借助计算机网络技术,开展线上智能化评价,形成线上分数报告系统,为学生快速提供分析诊断,同时依据强大的数据库,帮助学生提供更具个性化的分析报告。

(本文由高柳萍提供)

2021年第4期(总第76期)

江苏省教育考试院编印

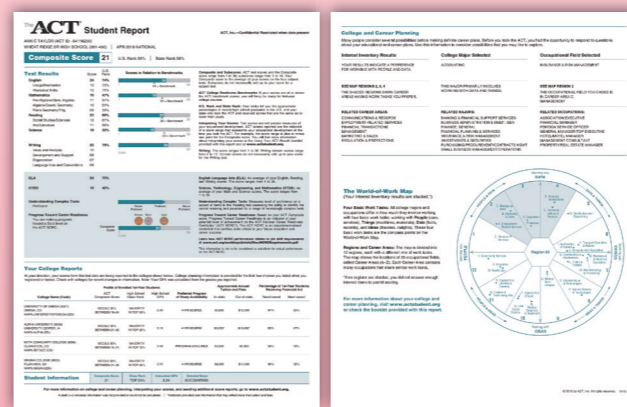
学生分数报告国内外发展现状与趋势

考试作为教育评价的主要形式,其诊断、反馈与矫正功能日渐受到关注。2020年10月,中共中央、国务院印发《深化新时代教育评价改革总体方案》,明确提出“改进结果评价、强化过程评价、探索增值评价、健全综合评价”。考生分数报告作为结果评价的形式之一,既承载了各种先进的测量、评价技术,也是教学主体采取干预和改进措施的主要依据。本期月报通过梳理国内外学生分数报告的发展现状与趋势,交流研究和实践经验与成果,以期为进一步深化教育评价改革,推进和完善个体评价提供思路。

【国外发展现状】

国外大规模教育考试,如美国的ACT和SAT,都基于考生的成绩研发了学生分数报告。

1. 美国ACT学生分数报告的框架为:(1)考生基本信息;(2)考试成绩说明,包括考生总成绩概况、考生排名情况和分项科目成绩细节;(3)成绩的递送与使用;(4)专业和职业生涯规划。该成绩报告单囊括丰富的信息、易懂的图表和标准化的量表分数,可以反映考生的成绩、专业和职业规划等诸多信息,在考生和大学之间架起了沟通的桥梁。



2. 美国SAT学生分数报告的设计紧紧围绕考试目的,根据不同层级,提供18个分数,主要包括5个

部分:(1)总分;(2)2个分项分数,分为阅读写作和数学两个部分;(3)2个跨科目分数,分别是科学分析分数和历史/社科分析分数;(4)3个科目分数,分别是阅读、写作以及数学;(5)7个子维度分数,包含情境中的词义、论据的掌握、观点表达、标准英语规范、代数、问题解决与数据分析、高等数学。另外,作文部分还包括三个分数:阅读能力、分析能力和写作能力,这三个分数将单独报告,不计入总分。同时,根据考生的各项成绩表现,从考生的答题表现中收集“证据”,为考生提供了关于大学学习和职业生涯准备的丰富信息。

3. 澳大利亚教育评价中心(EEA)为考生提供了丰富的诊断性报告,不仅报告考生的考试分数,还报告每个考生的优势和弱项、每道试题所测的能力、考生的百分等级以及考生群体和个人的答题情况,帮助学生诊断学习障碍。同时,借助于计算机网络技术,还研制了网上成绩报告系统,可个性化生成分析报告。

4. 美国教育考试服务中心(ETS)开发了很多具有影响力的测试产品,也为考生提供丰富的分数报告。如专业领域测试(Major Field Tests)考生分数报告包含免费版分数报告和收费版分数报告,免费版分数报

(下转第四版)

国际经验及启示

周云在《美国SAT考试分数报告的设计及启示》

一文中介绍,美国新一轮SAT考试改革中重新设计了分数报告,紧扣考试测量目标和用户使用需求,完整反映考试内容和考查能力,并通过量表转换进行分数合成并呈现考生排名情况,以反映考生能力倾向及大学学习和职业生涯准备情况。对我国教育考试分数报告设计提供的启示有:1.要设计行之有效的分数报告,采取测量什么报告什么的原则,为解释分数意义建立框架,搜集考生在考试过程中的证据,为分数报告提供解释。2.加大教育测量理论和技术的应用,通过将原始分数转换成具有参照点和单位量表的数值,使不同原始分数可比。3.建立适合我国国情的分数报告制度,以此更好探测到学生的内在知识结构和能力水平,识别其优势和劣势,从而对学生学习、教师教学、教育行政部门决策、高校招生录取、课程评价改革以及考试开发提供有效帮助。4.考试机构应从我国的实际需求和教育发展特点出发,开展实证研究,建立更适合国情的分数报告制度。

郭浩在《TOEFL Junior 考试成绩报告单的设计与启示》一文中介绍了小托福(TOEFL Junior)的基本内容构成及其成绩报告单的设计特点,阐释了小托福如何实现“为学校、教师、家长、学生提供更客观的测评标准、更精确的英语语言能力解读以实现英语学习指导”。对我国高考或中考成绩报告提供的启示有:1.应包含丰富的反馈信息。可依据各学科课程标准,将各学科划分若干知识与能力领域,统计分析各领域涉及的题目要求、知识和能力领域发展评价雷达图,以反映学科各知识领域的表现。2.可细化各单项能力描述,给出学习建议。基于各学科发展评价体系,将各总分分值不同的学科都放到一个尺度衡量,考生优势与不足可视化,语言类考试可细化单项能力描述与学习建议。以高考英语成绩报告单为例,可具体给出学科知识领域(词汇和语法知识领域、听力、口语、阅读理解、书面表达领域等)考生应达到的标准及能力建议。3.注意表达方式。小托福依据考生整体分数划分为优秀、熟练、扩展中、进步中、形成中5个等级,高分段积极赞扬性描述,低分段鼓励性肯定性描述。避免了消极、否定的表达方式,有助于学生积极从容

应对未来的学习生活。

俞如旺等在《分数报告的国际经验译介及其本土化的思考》一文中认为,分数报告(score report)是对考试分数进行综合分析的一种考试结果反馈形式。文章分析了分数报告的内容框架和行为准则,通过对北卡罗来纳州个人分数报告和分数频次报告的译介,为本土化提供了如下启示:1.报告项目应根据报告目的、对象不同进行科学合理设置,报告对象信息、学科结构和学业水平、学生总体测量信息、学业水平诊断性评价及反馈建议等,同时注意报告的行为准则,如明确报告对象、目的、应用图表呈现数据、提供必要的解释指南等。2.应用量尺分数和成就水平报告个体成绩并提供诊断性评价。量尺分数(Scale Score)是为了解决不同测验成绩的解释与比较问题,将原始分数进行转换,从参与测试学生群体的总体成就水平中导出。应用量尺分数取代原始分数,可避免升学压力下的“分分计较”。成就水平(Achievement Level)是以学生群体为对象,以参与评价学生在多个水平百分比分布的方式来显示。如北卡罗来纳州个人分数报告将文字描述与图表呈现结合起来,帮助用户更直观读懂量尺分数和成就水平。3.以描述性统计量和频次分布表呈现学生整体的成绩分布情况。北卡应用“量尺分数的简要统计”一栏呈现中位数、标准差、众数、量尺分数等描述性统计量;“量尺分数的频次分布表”一栏则呈现考生在不同分数、不同成就水平以下的累计人数及其比例;便于报告使用人了解学生整体的分数分布比率及其差异情况。我国初高中学业水平考试具有基础教育质量监测功能,因此这两项亦可以直观反映教育质量。文章同时认为,借鉴国际分数报告时应注意:1.国际分数报告多用于教育质量评估与监测,而我国高考考试规模大、涉及面广、利害性高,分数报告可能引起误解、问责等问题,设计者应充分考虑;2.数据转化、印刷等因素可能导致报告滞后于考试,对一线教学反馈不及时,应考虑基于网络的分数报告;3.报告使用者可能因涉及的统计学数据或名词过于复杂而难以全面理解考试主办方的报告意图;4.群体分数报告应在学校教学可能存在的问题及改进上提供反馈性建议。

国内经验及启示

刘辉在《促进学习的评价:从报告分数到建立反馈机制》一文中认为,单一的分数报告不能很好地促进学生的学习,促进学习的反馈机制要从静态的分数报告转向动态的整合分析,遵循“分析—判断—诊断—教学反馈意见”的过程。1.建立信息化平台,实现更及时、便于查询、多层次、多类型、多版本、多维度、多数据源的精细化学业质量分析。在反馈内容上,提供基于多数据源的多维度、多指标测评结果;在反馈分析技术上,基于认知诊断的评估在评价被试心理特质方面可以提供更精细的评估方法和更细致的评估结果。2.建立反馈机制,以促进学习为目标,在不同层面有不同类型、不同版本的报告,实现一班一报告、一师一报告、一校一报告的呈现方式。3.建立长效改进机制,发挥教学诊断和学习诊断的功能,通过反馈分析,制订跟进措施,提供针对性指导和改进,促进学业质量提升,最终形成学习—评价—分析—决策—改进的价值链,对于改进教学、促进学习具有更适切的意义。

朱华山在《综合评价体系构建中的高考改革探究——基于“云海工程”的实证研究》一文中介绍:1.云海工程评价体系的设计思想,即针对中国高考题库建设、考试命题、适用范围、公平度、问卷设计、评价使用、客户服务、信度、量度、等值等层面,预设一系列严格自我评价环节,提高考试评价服务水平。2.成绩报告单涵盖了35项基本信息与10项报录信息,“普通高考考生成绩报告”包括考试成绩、分项表现、升学指导三项内容。3.数据资源的利用上,项目省各州市成绩进行比较分析,并基于问卷调查比较高考成绩影响因素的重要程度等。4.人本评价彰显社会价值。试点目标就是通过逐步建立全面、综合、多元化的考试评价体系,使高校招生制度更有利于拔尖创新人才脱颖而出。文章认为,普通高考考生个案分析具有三个特点:一是通过附有全省位次、学科内容分项表现、能力结构分项表现等的考试成绩报告单,不仅让考生知分知位,合理人生规划,还能为学校评价教师教学

水平、社会评价学校教育质量提供量化参数;二是基于问卷测试的升学指导归纳了考生的兴趣、能力、优势,并简要判定考生在七种类型中的表现,给予考生填报志愿、选择专业的帮助;三是深入数据挖掘、梳理分析,揭示高考数据资源中考生成长潜力和专业发展潜力、中学教育资源均衡状况、教师整体性职业素质、社会背景公平程度、个人成长环境重要性、高校录取新生校际偏好和专业发展权重设置等。

刘海龙在《大规模教育考试诊断性分数报告的研发与应用》一文中,以青岛市初中毕业生学业水平考试成绩报告单为例,介绍了:1.诊断性分数报告的研发路径,包括“建构分数报告的诊断性测评框架、依据框架命制试题并标定试题属性、考试数据分析建模与测评框架效度验证”三项核心任务。2.诊断性分数报告的设计由整体报告(两个维度:学科学业发展评价和综合能力与素养发展评价)和学科报告(考生的9个学科报告各包含两个维度:知识领域评价和能力层级评价)两部分组成,以雷达图和描述性评价等方式评价考生在综合能力与素养、学科各知识领域与能力层级的表现。3.诊断性分数报告的发布与使用。发布前需广泛验证分数报告的实际展示效果,重点评估包含极低分、极高分、数据部分缺失等极端情况的分数报告效果,同时通过多种媒介解读发布分数报告的目的与意义,合理引导社会预期,并编制简洁易懂的分数报告使用手册以帮助考生正确理解与使用。4.进一步完善思考。一是描述性评价缺乏可比较的参照标准,导致诊断信息不够直接和明确,可考虑建立百分等级常模,使用百分等级分数报告便于考生与家长获取清晰的参照信息;二是基于学科的测评框架在初期便于理解和操作,未来应以基于素养的测评框架构建为目标;三是应进一步开发在过程性评价中同样可操作的互动式考试分数报告和评价系统,既可实时监控与反馈,还可对学生学习行为与考试作答情况进行相关性分析,真正实现数据驱动下的学习改进。