

【质量与评价】

中国高考难度分析理论构建初探

于 涵

【摘 要】传统的高考难度认知停留在统计数据层面。在网络时代,高考难度被赋予情绪意义和社会意义。本研究探讨不同主体视角下高考难度问题的认知差异,构建测量难度、感受难度和印象难度三维度的高考难度分析理论,分析不同维度之间的关系与相互作用机制,并应用高考难度分析理论探究高考难度演进过程及相关问题,提出相应建议。立足中国高考实际构建难度分析理论,有助于更全面地衡量高考难度,深化高考内容改革,实现立德树人、科学选才、引导教学的有机统一,服务教育强国建设。

【关 键 词】高考;难度;测量难度;感受难度;印象难度

【作者简介】于涵,男,教育部教育考试院党委副书记,副院长,研究员(北京 100084)。

【原文出处】摘自《中国考试》(京),2024.2.1~10

一、高考难度的认知差异及其研究的必要性

在新形势下,社会对高考难度的认知更趋多元;在评价高考难度高低时,时常带有较强的个人感受色彩。这要求管理者从制度设计上更全面地衡量高考难度——不仅要深刻认识试题难度的客观性,还要关注学生的主观感受,以及社会对高考难度的普遍认知。

(一)不同群体对高考难度的认知差异

考生、教育教学从业者及各类考试相关群体是对高考难度产生不同认知的三类主体。

广大考生受高考难度的影响最为直接,也最关心难度问题。该群体能接触到高考试题,认为分数反映出对努力学习和备考的回报,希望能够在高考中取得好成绩,在竞争激烈的录取环境中脱颖而出,进入理想的高校。高中教育教学人员密切关注高考难度的变化和趋势,据此开展教学、辅导学生。媒体和社会舆论普遍关注高考难度,并通过相关新闻报道、讨论和评论等表达对高考的关切。该群体关心高考难度,主要是从社会影响、公众舆论和教育公平的角度出发。

(二)相关研究对高考难度的认识差异

近些年,随着人民群众对更加公平、更高质量

教育的期待,教育研究领域对高考难度的认识角度趋向多元,对高考难度高低的认识也存在较大差异。有观点认为,高考作为国家选拔人才的重要手段,只有考试内容及难度具有一定挑战性,方能考查出学生解决实际问题的能力^[1]。但也有观点认为,高考难度大则弊大于利^[2]。

(三)构建中国高考难度分析理论的必要性

首先,高考难度问题是办人民满意高质量高考的必过之关。当前,我国教育正加快从“有学上”向“上好学”转变,必须建立与教育目标相匹配的高考难度分析理论,充分考虑学生、家长、社会的实际情况和心理感受,既反映出对学生能力素养的客观评价,又兼顾学生的发展环境和发展需求,努力让每个学生都能得到全面而有个性的发展。在高考难度分析理论指导下,要让学生对高考有信心,激发学习动力和发展潜力;要让家长对高考有信心,努力和付出终有回报;要让社会对高考有信心,优质教育资源得到公平分配;要让党和国家对高考有信心,确保党的教育方针不折不扣地得到贯彻落实,确保一批批可堪大任的栋梁之材脱颖而出。

其次,高考难度问题是健全人才培养体系的必要环节。当前,我国高等教育已迈入普及化阶段,

高考的选拔功能也从恢复初期的英才选拔转变为目前的大众升学、多元匹配,高考要为不同类型、不同层次高校的人才选拔服务。在新时代人才培养体系下,要发挥高考选拔的复杂功能,必须全面分析高考难度的影响因素和反馈机制。要以培养德智体美劳全面发展的时代新人为目标,通过调整高考内容和难度,不断增强学生的创新能力和实践能力,提高人才培养的连贯性和有效性。

最后,高考难度问题是构建中国特色高考理论体系的必解之题。当前,以经验为主的试卷难度控制和分析方法面临严峻挑战,已无法适应改革需求和教育高质量发展的需要。这要求结合我国国情、教育环境、文化背景等探讨高考难度的认识维度,运用系统观念构建中国高考难度分析理论,以满足不同群体对高考难度的要求,服务学生健康成长、科学选拔人才和促进社会公平。

二、高考难度分析的三元结构

要重新审视对高考难度的界定,就要从横向上关注社会各群体、教育各领域的感受和期待,从纵向上关注不同教育发展阶段对高考难度、区分度的需求。不同主体对高考难度的认知差异表明,可以将高考难度划分为测量难度、感受难度和印象难度三个维度,并以此为基础构建高考难度分析框架。下面系统分析三个维度各自的内涵与影响因素。

(一)测量难度及其影响因素

一般情况下,高考难度问题是从考试测量理论和命题实践的角度提出的。测量难度反映的是特定考生群体在高考试卷作答中的总体表现,也反映了试题试卷与相应考生群体学业能力之间的匹配程度,是引起难度讨论的、不以人的主观意志为转移的、内在的、本质的因素,是高考难度的基本维度。在命题实践的不同阶段,测量难度表现为不同的形式——蓝图计划阶段表现为设计难度,试题命制阶段表现为预估难度,考后报告与分析阶段表现为实测难度,均可以通过得分率、合格率、难度值等方式呈现。在测量难度指标设定上,理想的考试命题要符合教育测量学的流程与要求。首先,在对特定考生群体进行全面、充分了解的基础上,根据考试目的作出整体难度设计和各科难度设计;其次,

通过基于题库的选题组卷或新题命制来实现这一难度设计,通过高质量的难度预估及调控来校准实现的精度;最后,依据实考统计数据核验难度设计在各题各卷各科上的达成情况,并结合改革与试验的正面反馈和问题与教训的负面反馈,进行全流程数据分析,形成完整的难度分析报告。

决定测量难度的三个要素分别是任务内容、任务完成者和任务设计。任务内容包括考试科目、考查内容、题型题量等。任务内容体现的难度是考试所固有的难度,又称绝对难度。任务完成者(即考生群体)的学业水平不同,在完成同一任务时的表现就有差异,体现出考试难度的相对性。任务设计体现为高考试卷的命制标准与命制水平。能否实现绝对难度和考生群体水平的匹配,取决于能否在对不同考生群体深度把握的基础上设置与之相适应的任务内容,也就是命制出更适切的高考试卷。

(二)感受难度及其影响因素

感受难度是通过直接感知和观察获得的对高考难度的主观感受和情绪反应,基于个人切身体验,比较具体和个性化,是连接测量难度和印象难度的显性维度。学生、家长、教师等直接关系者最能体现考试过程中所面临的压力、焦虑、期望等心理状态。分析高考感受难度,就是要更好地理解 and 评估高考对学生身心健康和学习成效的影响,为他们提供支持和帮助,减少由于对高考难度的错误认知而导致的压力和焦虑过大等问题,提升学生的成就感,提升教师的价值感,提升家长的获得感。

主观感受既受个人经验、情绪状态、价值观和信念的影响,也受文化环境、外部信息等因素的影响。由于个人禀赋、学习能力、基础条件有差别,不同的学生群体和个体对同一套试卷的感受也会有所不同。

考生家长是高考的直接关系者,也会产生有关高考难度的感受。家长群体对高考难度的感受主要来自子女,包括家长通过观察得到的子女的学习情况、准备程度和考后情绪。

教师获得高考难度感受的方式是多样的。一是纵向比较。教师深入研究过去几年的高考试卷,

与当年高考试卷作详细对比,分析考查内容、试题形式的变化和难度的调整,从而产生对当年高考难度的直观感受。二是教学反馈。在日常教学过程中,教师与学生互动,对学生的学习表现进行评价。通过观察学生的学习动态、应试训练和模拟考试结果,教师可以对学生的学习情况有所了解,从而产生对高考难度的期望。三是政策预估。教师关注政策变化和考试改革动向,预判改革对命题的影响,这也会影响教师对当年高考难度的直观感受。这些方式相互作用,形成教师对当年高考难度的整体感知。相比于学生和家长,教师对高考难度的感受更加全面,也更为深刻。

(三) 印象难度及其影响因素

印象难度是基于对高考的认知、传闻、经验以及与其他考试的难度对比等形成的一种集体性认知,主要反映间接关系者对高考难度的理解、评价和期待。相比于感受难度,印象难度较为稳定持久,是构成高考难度的隐性维度。合理的印象难度能够从学习动力、信心、成就感等诸多方面整体赋能基础教育,减轻学生、家长及社会的焦虑,减轻不必要的内卷,提升幸福感和获得感。然而,一旦超出偏难阈值,印象难度就会对社会信心值、满意度和幸福感形成负反馈。因此,合理调控高考印象难度,有助于塑造社会对高考的正确认知和评价,有助于提升高考选才促学的公平性和有效性。

高考印象难度本质上是由测量难度和感受难度长期积累而逐渐形成的社会认知。除此之外,还有多种因素会对其产生影响。一是文化传统和社会价值观。二是教育资源分配和教育机会公平。三是教育发展和高考改革。随着教育发展和高考改革的推进,考试的内容和形式必然面临渐进、持续的调整。每当要引入开放性更强、更具灵活性的题型,就会引发公众的关注和讨论。

综合来看,能够提高印象难度的影响因素明显多于能够降低印象难度的影响因素。同时,偏难印象更容易被感知、传播和累积,但过高的印象难度无论是对于学生健康成长、基础教育导向、家庭教育负担还是对于社会心理反应,均会形成持续的负反馈效应,这是当前考试工作者研究解决高考难度

问题时要面对的首要问题。

三、高考难度各维度之间的耦合关系

(一) 考生主体视角下的决定机制与共振效应

测量难度、感受难度、印象难度三者紧密相关。从考生的视角来看,三者之间相互交织,形成不可分割的高考难度整体。

一是三者之间存在决定机制。测量难度直接影响感受难度。一般情况下,如果试题试卷的绝对难度较高,考生就会感受到更大的困难和挑战,进而产生较高的感受难度,测量难度和感受难度呈正相关。例如,就理科科目而言,其客观性较强,假如考生能力达不到甚至一个字都写不出来,则相关考试的测量难度与感受难度具有较强的一致性。但是,不同的科目、考试要求和题型,不同的任务复杂性、挑战性,不同的情境熟悉度,也会使得不同学生群体对高考难度的感受有所不同。例如,一些文科科目主观性较强,不管是否理解试题,也不管能否解决问题,考生均能有话可说,此时就可能出现测量难度与感受难度不一致的情况。总的来说,感受难度是对测量难度的传导与放大。感受难度在时间、空间(特别是网络空间)上累积,会形成印象难度的变化。因此,印象难度也可以理解为感受难度在外圈群体中的积分。

二是三者之间存在共振效应。考生即时的感受难度和所处环境内的印象难度也可能对最终测定的测量难度产生共振式影响。若考生在考试中产生的感受难度或所处环境内的印象难度较高,其对试卷的评价就会偏向较困难的方向,就可能增加其抵触情绪和压力,进而影响其临场发挥;相反,若考生在考试中产生的感受难度或所处环境内的印象难度较低,其对试卷的评价就会偏向较容易的方向,就会促使其更加积极、自信地面对考试,有利于其在考试中取得更好的成绩。

(二) 大众视野下的偏差放大机制与纠偏机制

高考难度是以测量难度为基础,感受难度和印象难度相互影响、相互转化并共同制约测量难度而形成的动态平衡的体系,三者之间存在相互作用机制。

一是三者之间存在偏差放大机制。感受难度

与设计难度之间的差值,特别是偏难负差,会通过网络放大和累积。“我们总是在亲眼观察世界之前就被预先告知世界是个什么模样;我们总是先对某一事件进行想象,然后再去切实地经历它。”^[3]当刻板印象形成后,强化和放大更易发生。正因为如此,每年高考季偏难方向的影响更易发生、更易累积,以致印象难度提高易、降低难。

二是三者之间存在纠偏机制,包括主动引导和考后回归两种纠偏。主动的难度预期引导对修正感受难度与设计难度的偏差会起到一定的纠偏作用,如考前发布难度保持稳定的信息,强调难度与考情学情相适应等。此外,感受难度主要体现于面对高考时的困难和压力,印象难度则体现出对高考成绩的评价和回顾——学生在面对高考时感受到较大困难,而在获得实考成绩后,通过比较,他们向社会释放和传播的印象难度也会回调。

(三) 考试机构视域下的校准机制和赋能机制

测量难度、感受难度和印象难度都是动态变化的。考试工作者要充分重视高考难度的校准机制和赋能机制,合理控制高考难度。

一是校准机制,包括循环校准和迭代校准。循环校准指的是在测量难度概念内设计难度、预估难度、实测难度之间的校准,根据实测结果和反馈,与命题人员进行交流和讨论,分析实际执行过程中的问题和困难,采取可能的改进措施,对设计和预估的难度进行校准和调整,以降低对经验的依赖。重复设计难度、预估难度和实测难度的循环校准,有助于不断优化测量任务的准确性和可靠性。迭代校准指的是测量难度和感受难度之间的校准,综合统计数据和主观感受调查结果,分析测量难度和感受难度的匹配度,对不匹配的方面或问题进行相应的调整和优化。在下一轮高考中,再次评估测量难度和感受难度的匹配度,验证任务调整的效果。这种迭代校准需要多方参与和反馈,要依靠大量的实践数据和专家评估,可使任务难度与个体感受更加匹配,也可从中获取更准确、更全面的任务难度信息,为考试命题提供更科学、有效的参考依据。

二是赋能机制。首先,深入了解学生的感受难度和当下的印象难度,可以更好地帮助识别学生遇到的困难和障碍。如果学生感受到较高的难度和

负面的印象,可以向他们提供额外信息和任务支持,有针对性地调整难度设计策略,以避免产生显著高于设计难度、预估难度的感受难度,增加学生的学习自信心,帮助他们克服挑战。其次,应积极引导社会对高考的评价,避免产生超出偏难阈值的印象难度,提升高考的正面形象,减轻考生的心理压力,为考生、教师、家庭、基础教育及社会赋能。最后,通过应用教育测量技术和主观感受评价相结合的新模式,改革考试内容和考查方式,合理设置难度,助推高中新课程改革,促进高中育人方式改革,更好地评价学生应对未来学习、生活和工作中不确定性的能力素养。

四、高考难度分析理论视角下的高考演进过程及问题分析

(一) 演进过程

1977年恢复高考以来,高考难度被普遍认为是评价教育质量和学生能力的重要指标。然而,随着社会的不断变化,人们对高考难度的理解也在不断发生变化。

恢复高考后的10余年间,全国高校招生录取比例较低。20世纪80年代初期,随着教育与心理测量学的引进,有关研究人员和教育考试部门开始利用测量理论分析试题试卷的难度^[4-5]。由于这一时期高考考生在同龄人中占比较小,上大学并非主要发展途径,感受难度并不为社会所重视,印象难度亦尚未形成。

20世纪80年代中期,高考探索推行标准化考试。首先,随着我国经济发展和全球化进程的推进,社会对人才的需求度越来越高,高考更注重能力考查,增加了一些挑战性问题,以选拔出更有潜力的学生,测量难度略有提升。其次,参加高考的学生越来越多,学习水平差异较大,高考偏难的感受迅速形成。最后,对高考难度的讨论开始跳出试题对当届考生的影响,更多地关注对教育长远发展的影响,较难的高考印象逐渐形成。

1999年,高校大扩招启动,我国高等教育向大众化加速迈进。高校招生开始探索分类考试、综合评价、多元录取的新模式,教育考试改革旨在全面评价学生,减轻高考压力。测量难度有所降低,参加高考的学生逐年增加,社会对高考难度的关注度

也越来越高,致使高考难度极易被个别科目偶尔的感受难度波动所引爆,印象难度持续增大。

当前,公众对高考难度的关注有增无减,测量难度的调整与社会反馈不相匹配,考试分数的提升与考生家长的感受不相匹配,考查内容和形式的变化同公众的印象不相匹配。高考难度各维度的总体情况是印象难度大于稳定的感受难度,而反映出的感受难度又往往大于测量难度。印象难度、感受难度与测量难度的负向偏差仍在随着每年高考举办不断累积,只有程度大小的区别。

(二) 问题分析与建议

高考难度在偏难方向上超出阈值,导致考生不满、人民群众不解、教育本身不认可,考试功能难以发挥。对命题而言,主要是因为对高考难度的认识不够清晰,包括难度设计目标、存在的问题、改进措施及检验方法等,主要表现为:(1)难度控制观念较为陈旧,对高考难度不重视、不正视、不愿动、不敢动;(2)难度质量控制体系不健全,存在习惯闭门造车、习惯按部就班、习惯因循守旧、习惯单打独斗等问题;(3)命题制度有欠缺,难度控制流程不够严密;(4)难度控制能力有短板,相关研究投入严重不足等。

作为一个全新的理论框架,高考难度分析理论可以解释不同群体对高考难度的认识。高考难度分析理论更是一个高考内容改革及命题评价的分析框架,可以用于指导高考命题的难度控制工作。考试工作者要深刻把握高考“立德树人、服务选才、引导教学”的核心功能^[6],基于大规模高利害选拔性考试的特点,应用高考难度分析理论指导命题工作实践,不断优化试卷难度结构和试题设计,形成科学的高考难度质控体系。

一是应用高考难度分析理论构建试题试卷测量难度的校准机制。难度的校准机制应包括如下三个环节:(1)在每年高考命题前,全面分析当年报考人数、录取比例等方面的变化,深入开展教情学情考情调研,全面了解当年学生与往年学生的水平差异,应用大数据进行模拟分析,确定高考试题试卷的设计难度;(2)在命题过程中,以设计难度为目标,从考查内容、考查要求、考查载体等方面与往年试题进行分析对比,不断优化试题,形成试题试卷

的预估难度;(3)在考试结束之后,及时统计实测难度,将设计难度、预估难度与实测难度进行分析对比,校准预估难度,形成科学合理的测量难度调控与校准机制。

二是应用高考难度分析理论指导试题命制,优化感受难度,提升测量难度与感受难度的一致性,赋能基础教育。在保证区分效果、不降低考查要求的前提下,优化试题的考查内容和设问角度,为学生和教师提供良好的答题体验。通过优化感受难度,一方面整体赋能基础教育的教与学,激发学生的学习兴趣,培养学生的自信心,提高获得感;另一方面进一步减少不必要的内卷和超前超量学习的负担,减轻学生、家长乃至社会的焦虑,提升幸福感。

三是应用高考难度分析理论引导各类主体科学认识高考的功能,不断改善高考的印象难度。首先,要持之以恒地合理控制测量难度,认真处理可能对印象难度造成不良倾向的各种因素,谨慎对待每次考试、每个学科、每套试卷、每道试题的测量难度,为优化感受难度做好基本功。其次,要针对印象难度形成的特点,畅通宣传舆论渠道,加强教育信息的传递与沟通,更加精准释放高考内容改革信号,培养和提升不同群体的考试评价素养,让社会更加了解高考测量难度的科学性、公正性和合理性,引导社会对高考难度分析理论的认同,重塑高考的印象难度。

参考文献:

- [1] 陈志文. “高分通胀”加剧分数竞争:再议中高考试题的难与易[N]. 光明日报, 2019-08-06(14).
- [2] 王书金. 大幅度降低高考难度 为实施素质教育创造有利条件[J]. 现代教育科学, 2009(3): 106-109.
- [3] 李普曼. 舆论[M]. 常江, 肖寒, 译. 北京: 北京大学出版社, 2018: 73.
- [4] 心理系测验研究小组. 对高考试题的统计分析[J]. 北京师范大学学报, 1981(5): 65-73.
- [5] 张厚粲. 教育测量学: 高考科学化的技术保障[J]. 中国考试, 2017(8): 4-7.
- [6] 教育部考试中心. 中国高考评价体系[M]. 北京: 人民教育出版社, 2019: 9-12.