

《增强型ACT链接研究报告》

董梅 李，沙里尼·卡波特，安·阿瑟，黄志宇，赵永武，仇晨，王红玲

执行摘要

从2025年4月开始，ACT将引入ACT的增强版形式。[®] 测试全国在线考试，预计到2026年春季在全国、州和地区以及国际的纸笔考试和在线考试参与者中全面推广。ACT通过改变考试长度和考试时间、每个题目提供更多时间、重新平衡报告类别、增加嵌入式实地测试题目、将科学测试设为可选以及将ACT综合分数从英语、数学、阅读和科学（EMRS）的平均分改为仅英语、数学和阅读（EMR）的平均分进行了重大更新。这些增强功能的详细信息可在ACT网站上查看（www.act.org/actenhancements）。尽管有这些变化，改进后的ACT测试旨在测量与旧版ACT相同的结构。为确保历年分数的一致性，ACT决定在改进测试中维持旧版1-36分的评分标准。

在产品发布前，进行了两项特别研究：一项是2024年6月进行的链接研究，旨在将增强版形式与旧版形式相连接；另一项是2024年10月进行的模式研究，旨在考察增强版ACT的纸质版和在线版之间的分数可比性。本文件重点关注2024年6月链接研究的分析和发现。

2024年6月的链接研究采用了随机分组设计，其中两种增强形式和一种传统形式在线进行并随机分配给参与者。最终清洗后的数据集包含6,882名学生。在开展统计链接之前，考察了这些增强对学生测量结构和学生表现的影响。随后，对增强形式的心理学特性进行了检查，并与传统形式进行了比较。

为了检验构建等价性，ACT使用不同的模型进行了确认性因子分析。增强形式的拟合指数和因子负荷量与每个模型的传统形式相似，表明测量构建的一致性。为了检验增强对学生表现的影响，ACT检查了完成率、快速响应率、跳过率和学生对调查问题的回答。对于某些测试部分，有些证据表明在测试完成率和学生对有足够时间完成测试的感知上存在一些小的积极影响。

等效方法被用于将改进形式与原有形式联系起来。链接结果的稳健性得到了子群不变性证据的支持，这种不变性与近期平行ACT形式等效中的观察结果相似。进一步的分析被进行以评估链接分数的心理测量属性，包括可靠性、标准误差等。

测量 (SEM)、条件测量 (CSEM) 以及ACT大学入学准备标准与ACT的分类一致性 ® WorkKeys ® NCRC ® 指标。

可靠性对于增强形式分别为EMRS综合指标的0.96和EMR综合指标的0.94，以及测试部分的0.83至0.88之间。正如预期，这略低于传统形式的可靠性，这是由于测试长度的减少所致。因此，测试部分与大学预备指标分类一致性之间的观察相关性也略有下降，正如预期的那样。

在SEM和CSEM方面，传统ACT分数的CSEM对于综合分数约为1，对于测试部分约为2。虽然增强形式在每个测试部分得分上显示出CSEMs约提高半分点的增长，但CSEMs在分数尺度上的变异性与传统测试相似。此外，增强形式的四舍五入SEM与传统测试的分数报告中所用SEM相似（综合分数约为1，测试部分得分约为2），除了阅读SEM略高于2.5。因此，测量精度的变化对分数解释的影响非常小。

总的来说，增强版的ACT测试测量了与原版ACT测试相同的结构，并在相同的量表上报告分数。这些增强对学生的表现影响很小，尽管有一些证据表明对测试完成率和学生对有足够时间完成测试的认识有轻微的积极影响。尽管测试长度的减少导致了预期中的测量精度下降，但对分数解释的影响微乎其微。因此，增强版的ACT测试和原版ACT测试的分数可以相互替代使用。

概览 ACT 增强功能及专项研究

从2025年4月开始，参加全国考试日期在线参加ACT考试的学生将参加增强版的ACT考试。随后，在2025年9月，增强版的ACT考试将面向所有国内外学生推出。最后，在2026年春季，增强版的ACT考试将提供给州和地区的学生。主要改进包括缩短测试长度和总体测试时间，每项测试时间增加，科学测试现在为可选项目，嵌入现场测试项目，以及ACT综合分数现在不是从英语、数学、阅读和科学成绩的平均值计算，而是从仅英语、数学和阅读成绩的平均值计算。尽管有这些变化，增强版的ACT旨在测量与旧版ACT相同的结构，并维持1-36的评分尺度。

为准备2025年4月开始的增强型ACT考试，ACT进行了两项特别研究：2024年6月的一项链接研究以及2024年10月的一项模式可比性研究。链接研究的目标是评估结构不变性和将增强型ACT考试与原有ACT考试相连接。模式可比性研究的目标是评估纸笔和在线考试的分数是否等效，并实现不同模式下的分数互换。

表1总结了增强版和传统ACT测试中各测试部分的测试长度和测试时间的变化，通过比较项目数量、测试时间和每项平均时间。在传统ACT中，所有项目均计入学生的总成绩，但在增强版ACT中，只有有效项目计入学生的总成绩。

表 1. 项目数量与测试时间在传统形式与增强形式之间的比较

测试章节	项目数量				测试时间			
	遗产	增强			总时间 分钟		平均时间每 项目秒数	
	总计	操作性的	实地测试	总计	遗产增强	遗产增强		
English	75	40	10	50	45	35	36	42
数学	60	41	4	45	60	50	60	67
阅读	40	27	9	36	35	40	53	67
科学	40	34	6	40	35	40*	53	60*

*六月份链接研究中科学测试的总时间为45分钟（平均每项68秒），但后来改为40分钟（平均每项60秒）。

该改进包括测试长度和总测试时间的减少，但每项平均时间的增加。在改进后的ACT考试中，英语、数学、阅读和科学测试的运行项目数量分别减少了35个（47%）、19个（32%）、13个（33%）和6个（15%）。传统ACT考试的EMRS综合评分基于215个项目，而改进后的ACT考试的EMR综合评分将基于108个项目。项目数量为

项目对EMR综合评分的贡献减少了107（50%），这是由于测试长度的减少和计算中排除科学导致的。

正如表1中的注释所指出的，在2024年6月的链接研究中，增强型科学测试采用了不同的时间安排，测试时间为45分钟（比传统ACT测试多10分钟）。在ACT审查了额外10分钟对科学测试表现的影响，并考虑到在科学选考的情况下需要最大化测试效率后，增强型科学测试的时间从45分钟调整为40分钟。2024年10月的模式可比性研究重新审视了时间调整，并确认增加5分钟对于科学测试来说是足够的。

本文件聚焦于2024年6月链接研究的结果。它首先描述了用于研究的资料。随后展示了关于结构等价性和心理测量特性的数据分析及结果，接着是链接研究的方法和详细结果。最后，提供了时间分析结果以及主要发现和结论的摘要。

数据收集设计和样本人口统计

在2024年6月的链接研究中，使用了其中一个传统ACT表格（以下简称“传统表格”）以及两个增强型ACT表格（以下简称“增强表格1”和“增强表格2”）。

表2. 人口统计学特征（百分比）

人口特征		研究样本	五年级ACT 测试量
性别	另一个性别	0.2	0.3
	女性	54.3	52.2
	男性	44.7	44.4
	不进行回复	0.8	0.9
	缺失	<0.1	2.3
种族/民族	美印第安人/阿拉斯加原住民	0.7	0.8
	亚洲	4.8	5.2
	黑人/非裔美国人	14.5	11.9
	夏威夷原住民/其他太平洋岛民	0.1	13.6
	西班牙裔/拉丁裔	13.4	0.2
	白色	58.0	54.2
	两个或多个种族	5.0	4.4
	不进行回复	3.4	3.8
	缺失	0.1	5.9

该研究的参与者是2024年6月ACT国家考试的网络测试者。
当学生们注册参加考试时，他们可以选择在线或纸质考试。

参与者选择在线测试的获得了额外的权限和研究信息。不同意参与研究的学生可以选择注册参加笔试。数据是在线上使用随机分组设计收集的，其中包括Legacy表格、增强表格1和增强表格2在测试中心的学生中进行循环，包括使用额外时间安排的学生。全国超过180个测试中心参与了该项研究。最终的清洗数据集包括6,882名学生：Legacy表格2,298名，增强表格1 2,280名，增强表格2 2,304名。

样本的性别和种族/民族特征详见表2。为进行比较，还包括了过去5年ACT国家测试人口的统计数据。参加传统形式的样本的平均ACT综合得分为21.57，标准差为（此处缺少具体数值）。 SD 5.53。相比之下，平均综合指数和 SD 过去5年中，ACT考试人群中的比例分别为20.72和5.81。因此，该研究样本的性别、种族/民族和分数分布代表了ACT考试人群。

构建等价

有效的链接要求之一是测试特征的高度相似，包括构造、人口、推断和测量特征（Kolen & Brennan, 2014）。增强版的ACT旨在通过测量相同的构造得出相同的推断。增强版的ACT的测试规范与原始ACT的测试规范相似，但测试部分的长度和/或时间不同。

表3. 一因子模型的模型拟合统计量

主题	表格	χ^2	df	p - 值的 χ^2 test	RMSEA	CFI	平均 标准化 加载中
English	遗产	7,008.428	2,700	.00	.03	.95	0.55
	增强 1	1,749.684	740	.00	.02	.97	0.52
	增强 2	1,710.667	740	.00	.02	.97	0.51
数学	遗产	4,150.435	1,710	.00	.03	.97	0.55
	增强 1	1,739.304	779	.00	.02	.97	0.50
	增强 2	2,197.379	779	.00	.03	.95	0.49
阅读	遗产	1,771.695	740	.00	.03	.97	0.55
	增强 1	594.894	324	.00	.02	.99	0.52
	增强 2	514.656	324	.00	.02	.99	0.51
科学	遗产	2,096.352	740	.00	.03	.97	0.56
	增强 1	1,528.193	527	.00	.03	.95	0.50
	增强 2	1,193.889	527	.00	.02	.97	0.51

为了评估结构等价性，ACT的心理计量学家为三种形式估计了确认性因子模型。使用操作评分项和加权拟合了一系列模型。



最小二乘估计量。这些模型都是单因素模型，其中所有条目都加载在每个测试部分的单个潜在因素上，该因素代表每个测试部分的（例如，英语）知识或能力。表3报告了每个测试部分的单因素模型的模型拟合统计量。Hu和Bentler（1999）的可接受拟合指南包括非显著性的²均方根误差（RMSEA）小于0.06，比较拟合指数（CFI）值大于0.95。基于RMSEA和CFI，该模型对所有□□□□□□形式和□□部分均符合可接受□准。□果□²测试结果已知会受到样本大小的影响，因此即使模型拟合良好，当样本量大时，也可能得到具有统计学意义的显著结果。表3中还展示了单因素模型的平均标准化因子负荷。平均标准化负荷值介于0.49到0.56之间，表明项目与遗留和增强版ACT测试的潜在因素之间具有中等相关性。

此外，在每个测试部分内的报告类别都拟合了确认性因素分析模型，且在传统和增强形式下，报告类别得到了相似的因素负荷。

心理学特性

ACT调查了测试变更对测量精度、测试分数可靠性和ACT测试间观察分数相关性的影响。还对分数解释的实际影响以及个体分数的决定一致性进行了考察。

条件测量标准误、测量标准误和信度

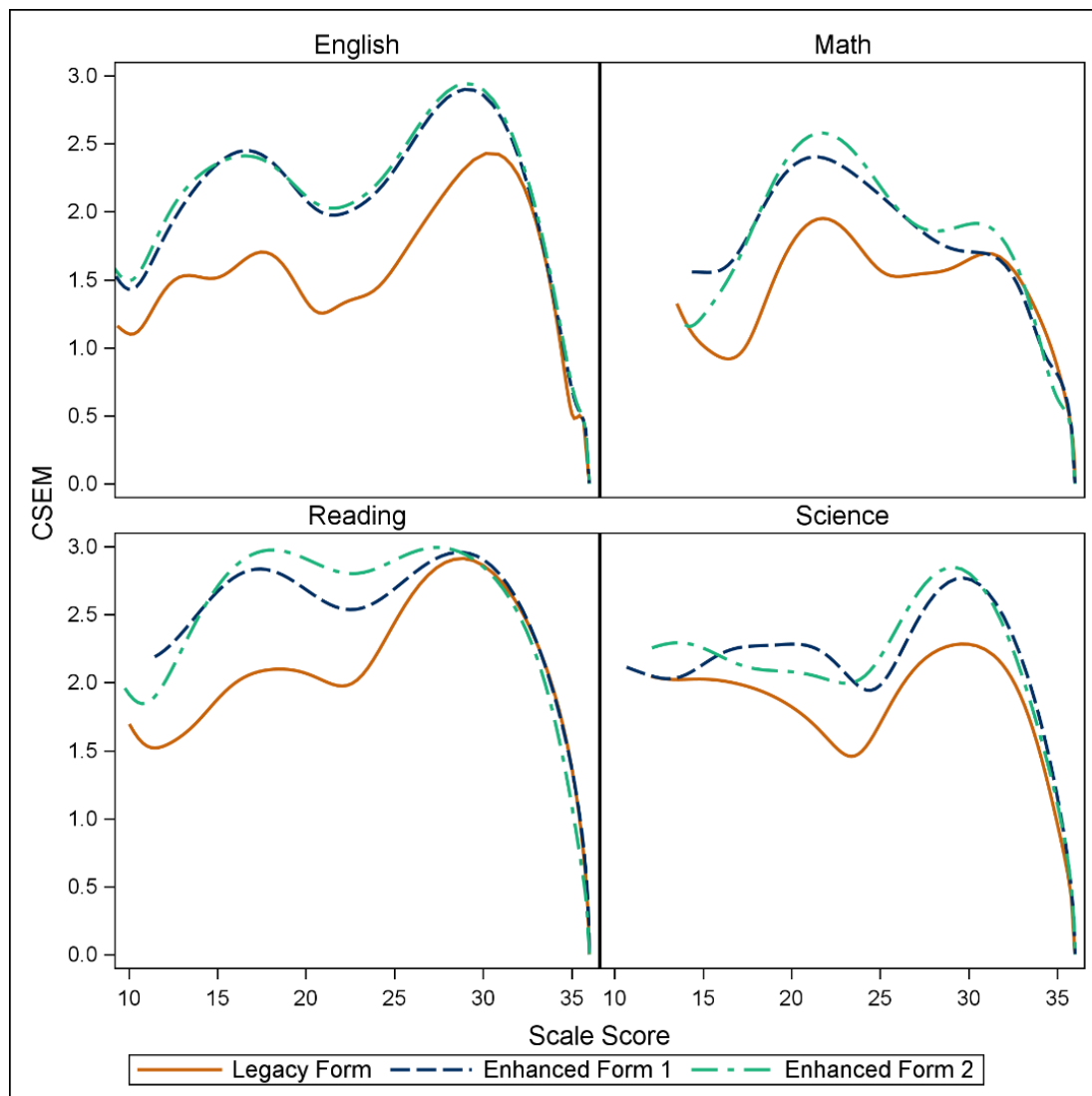
在1989年，ACT测试的量表调整为沿分数尺度具有大致相等的条件标准误差测量（CSEMs）。后续的测试形式通过一系列的等值链与基础形式进行等值，对所有ACT形式的CSEMs进行了持续监测。预计新测试形式在统计和内容方面尽可能接近基础形式时，将最好地保持恒定的CSEM属性。

规格说明。然而，由于测试规格和时间推移中的形式难度逐渐变化，传统形式之间存在差异。正如预期的那样，随着年份的增加，评分尺度上的CSEMs（连续学生等效模型）表现出增加的变异性，但传统的ACT多项选择题部分仍然大致保持约2的CSEM（连续学生等效模型）（ACT，2024）。

ACT对增强型ACT测试形式的规模分数CSEMs（条件标准误差测量）和测量标准误差（SEMs）进行了考察，以评估测量误差是否在不同分数间仍然大致相等，以及误差的大小是否超过了在传统ACT分数报告中所报告的误差量（即，每个选择题部分为2，综合分数为1）。CSEMs、SEMs和可靠性是基于Kolen等人在1992年所描述的四参数贝塔复合二项模型进行估计的。图1展示了每个测试部分的CSEM图。尽管传统形式、增强形式1和增强形式2的CSEMs在规模分数上具有相似的模式，但由于测试长度的缩短，两种增强形式都增加了CSEMs。尽管如此，增强形式1的CSEMs

并且2仍然大约位于分数尺度的多数部分，除了它们在阅读方面正接近3。

图1. 各部分测试的测量条件标准误



可靠性估计和SEM估计在表4中报告。尽管测试长度减少导致SEM略有增加和量表的可靠性得分降低，但增强形式仍然表现出高度的可靠性：EMRS综合评分为0.96，EMR综合评分为0.94，测试部分的可靠性介于0.83至0.88之间。SEM的小幅增加——EMR和EMRS综合评分约为0.3，测试部分约为0.5——对分数解释的影响最小。与旧形式相似，增强形式在EMR和EMRS综合评分的SEM约为1，测试部分的SEM约为2，但增强阅读测试的SEM略高于2.5。总的来说，增强形式的可靠性依然稳健，确保了分数解释的一致性和可靠性。

表 4. 可靠性估计和每一部分测试及综合分数的标准误差

测试/分数	遗产 表格	可靠性		遗产 表格	半导体设备制造 (Sem	
		增强 表单 1	增强 表2		增强 表单 1	增强 表2
English	.94	.88	.88	1.58	2.23	2.26
数学	.93	.88	.88	1.43	1.92	1.94
阅读	.89	.84	.83	2.22	2.66	2.75
科学	.89	.85	.85	1.88	2.22	2.25
STEM	.95	.93	.92	1.18	1.47	1.49
电子医疗记录 (Electronic Medical Records)	.97	.94	.94	1.03	1.32	1.35
复合材料	.97	.96	.96	0.90	1.14	1.16

注意。STEM是学生数学和科学成绩的组合。

相关性与去校正相关性

观察到的相关性以及去除了测量误差的影响后的去相关性（无衰减相关性）在表5中呈现。可能由于可靠性的降低，增强形式1和2的测试部分之间的所有观察相关性都低于传统形式的观察相关性。传统形式与增强形式1和2之间的无衰减相关性相似。

表 5. 相关系数（粗体表示上对角线）和非去偏相关系数（斜体表示下对角线）

形状和截面		English	数学	阅读	科学
遗产	English	—	.77	.83	.79
	数学	.82	—	.71	.82
	阅读	.91	.78	—	.79
	科学	.87	.90	.89	—
增强 1	English	—	.69	.79	.74
	数学	.78	—	.63	.78
	阅读	.91	.73	—	.72
	科学	.86	.91	.85	—
增强 2	English	—	.70	.79	.76
	数学	.80	—	.63	.80
	阅读	.92	.74	—	.74
	科学	.88	.92	.87	—

有效权重

ACT综合分数是各部分测试分数的简单平均值，表明所有组成部分拥有相等的名义权重。然而，每个组成部分测试对综合分数方差的实际贡献可能并不相等。有效权重，衡量可以归因于每个组成部分测试的综合分数变异性的比例，分别显示在表6中的EMRS综合分数和表7中的EMR综合分数。Legacy表格和增强表格1和2的有效权重是相同或非常相似的。

表6. EMRS综合评分中各组成部分测试的有效权重

表格	English	数学	阅读	科学
遗产	.27	.22	.27	.24
增强 1	.27	.22	.27	.24
增强 2	.27	.22	.27	.24

表7. EMR综合评分的组成部分测试的有效权重

表格	English	数学	阅读
遗产	.35	.29	.36
增强 1	.35	.28	.36
增强 2	.36	.28	.36

分类准确度：ACT大学准备度基准和WorkKeys NCRC 指标

表8展示了ACT大学准备标准及ACT进步情况的分类一致性。® WorkKeys® 国家职业准备证书® (NCRC®) 指标。分类一致性指的是学生在重复测量过程中的分类保持一致性的频率。然而，由于学生只能参加一次考试，分类一致性通常是根据单一考试行政来估计的，并假定测量误差和真实分数的分布。使用Livingston和Lewis (1995) 提出的方法，ACT估计了ACT大学入学准备基准 (分别为英语、数学、阅读和科学18、22、22和23分) 的分类一致性。分类一致性的估计值相对较高，两种增强形式的值在0.83到0.89之间，虽然略低于传统形式的值，后者在0.87到0.92之间。

表 8. ACT大学准备标准和工作关键NCRC指标的分类一致性指数

		遗产 表格	增强表单 1	2	增强
学院 准备就绪 基准指数	English	.90	.85	.85	
	数学	.92	.89	.89	
	阅读	.87	.84	.83	
	科学	.87	.85	.85	
国家认证认可监督 管理委员会 指标	NCRC 综合指数	.80	.75	.74	
	EMR 综合指数	.77	.71	.70	

使用相同的方法，ACT 还估计了 NCRC 指标的分类一致性。根据他们的 ACT 综合分数，学生被分为五个类别之一，对应于 WorkKeys NCRC 级别，从不太可能获得 NCRC 到最有可能获得铜、银、金或白金证书。这些 NCRC 级别预测的 ACT 综合分数的分数线分别为 13、17、22 和 27。同时获得了 EMR 综合分数和 EMRS 综合分数的这些分数线的分类。增强形式的分类一致性估计值在两个综合分数上介于 .70 和 .75 之间。尽管这些估计值低于传统形式的估计值，其范围在 .77 到 .80 之间，但考虑到有五个 NCRC 类别，增强形式的协议率仍然相当高。

连接方法、结果与评估

ACT使用IRT真分数等值方法将增强版1和2的分数与Legacy版本的分数联系起来。在本研究中使用的Legacy表格之前已经进行过等值处理，并用作连接增强版1和2的基准。首先，对所有三个表格的项目参数进行了校准。然后，将项目参数转换为与Legacy ACT项目库相同的规模；ACT使用了Stocking和Lord (1983) 的程序，根据Legacy表格的参数得到比例转换的截距和斜率。最后，ACT进行了IRT真分数链接，以获得新ACT表格的原始到规模分数转换。

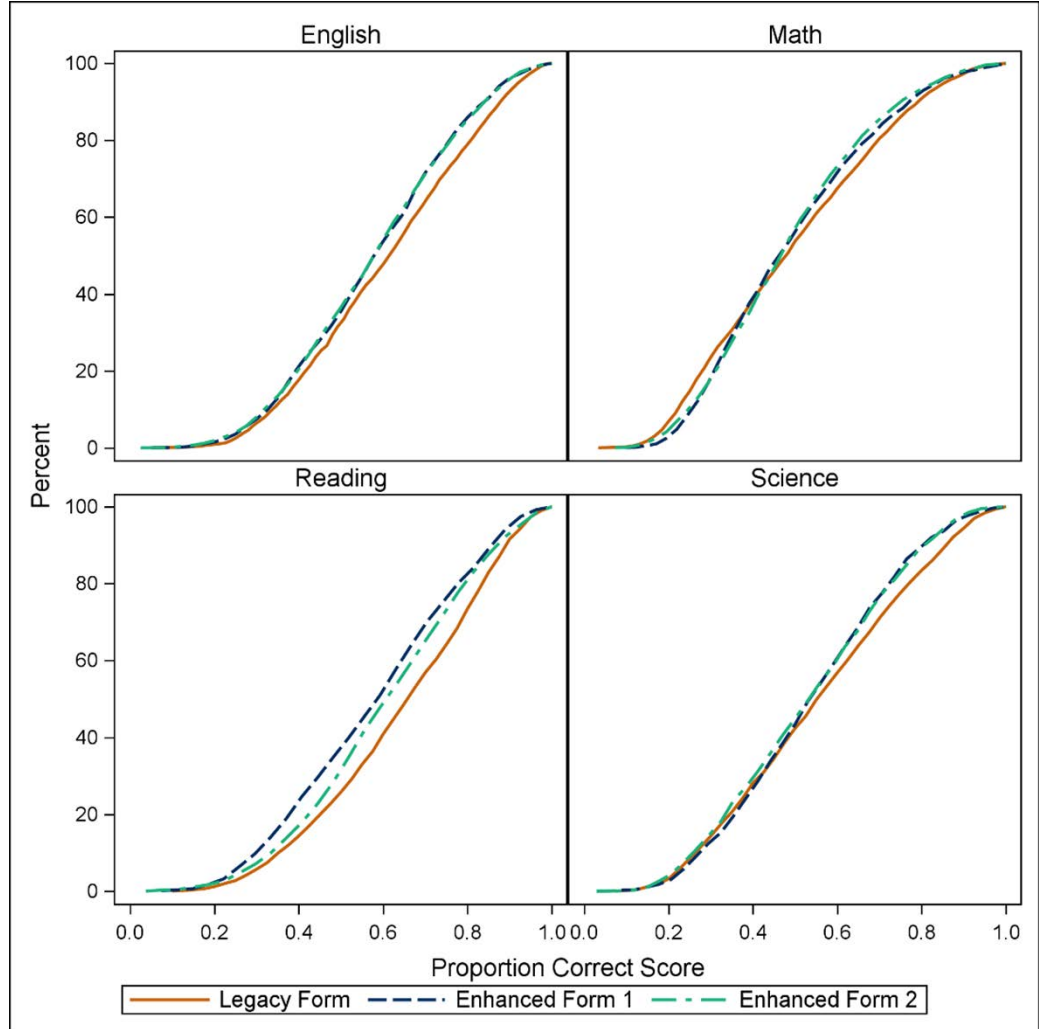
描述性统计与分布

正确比例得分描述性统计提供在表9中，它们的相对累计频率分布展示在图2中。增强形式的难度与旧测试在所有四个测试部分中相似。

表9. 正确比例原始分数描述性统计

测试章节	表格	平均数	SD	偏度	偏度
English	遗产	0.62	0.20	-0.08	-0.92
	增强 1	0.59	0.19	-0.04	-0.76
	增强 2	0.59	0.19	-0.05	-0.74
数学	遗产	0.50	0.21	0.24	-0.93
	增强 1	0.50	0.19	0.45	-0.61
	增强 2	0.49	0.19	0.36	-0.50
阅读	遗产	0.65	0.20	-0.34	-0.77
	增强 1	0.60	0.21	-0.08	-0.90
	增强 2	0.62	0.20	-0.18	-0.65
科学	遗产	0.57	0.22	0.02	-0.99
	增强 1	0.55	0.20	0.04	-0.75
	增强 2	0.55	0.20	0.01	-0.93

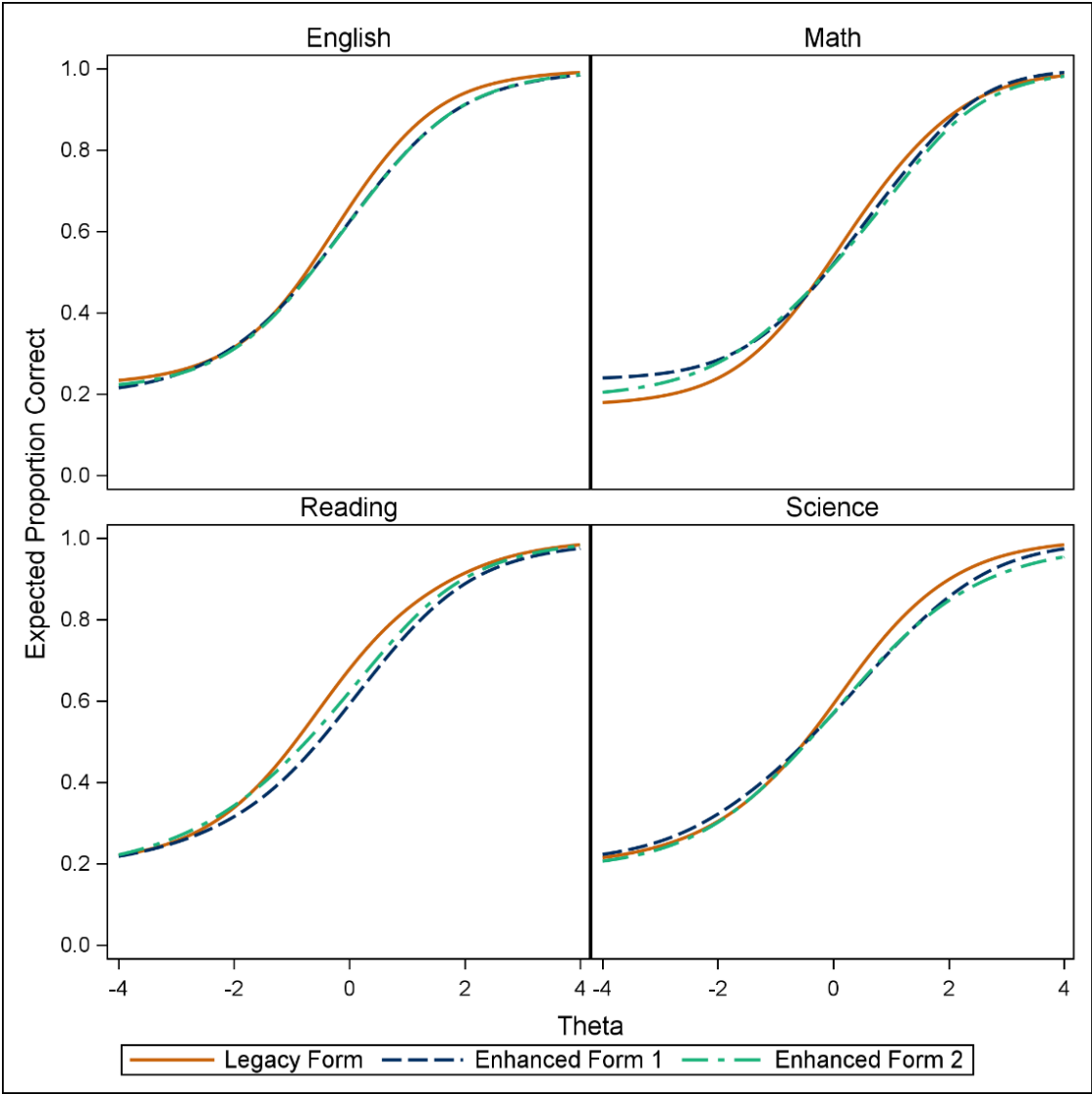
图2. 正确得分比例的相对累积分布



测试特性曲线

ACT使用了三参数逻辑回归项目反应理论模型来拟合项目分数数据。对每个测试版块的项目参数进行了估计，然后根据Stocking-Lord方法，使用旧版形式中的项目将参数转换到ACT旧版项目库中。图3显示了每个测试版块三个研究形式的测试特征曲线（TCC）。实线代表旧版形式，虚线代表两个增强形式。 x 横轴表示theta得分，并且 y - 轴表示正确比例。在这些图表中，向左移动的曲线表示更简单的形式。这与正确比例原始分数分布是一致的。尽管增强形式相对于传统形式（尤其是阅读）略难，但差异很小。

图3. 测试特性曲线



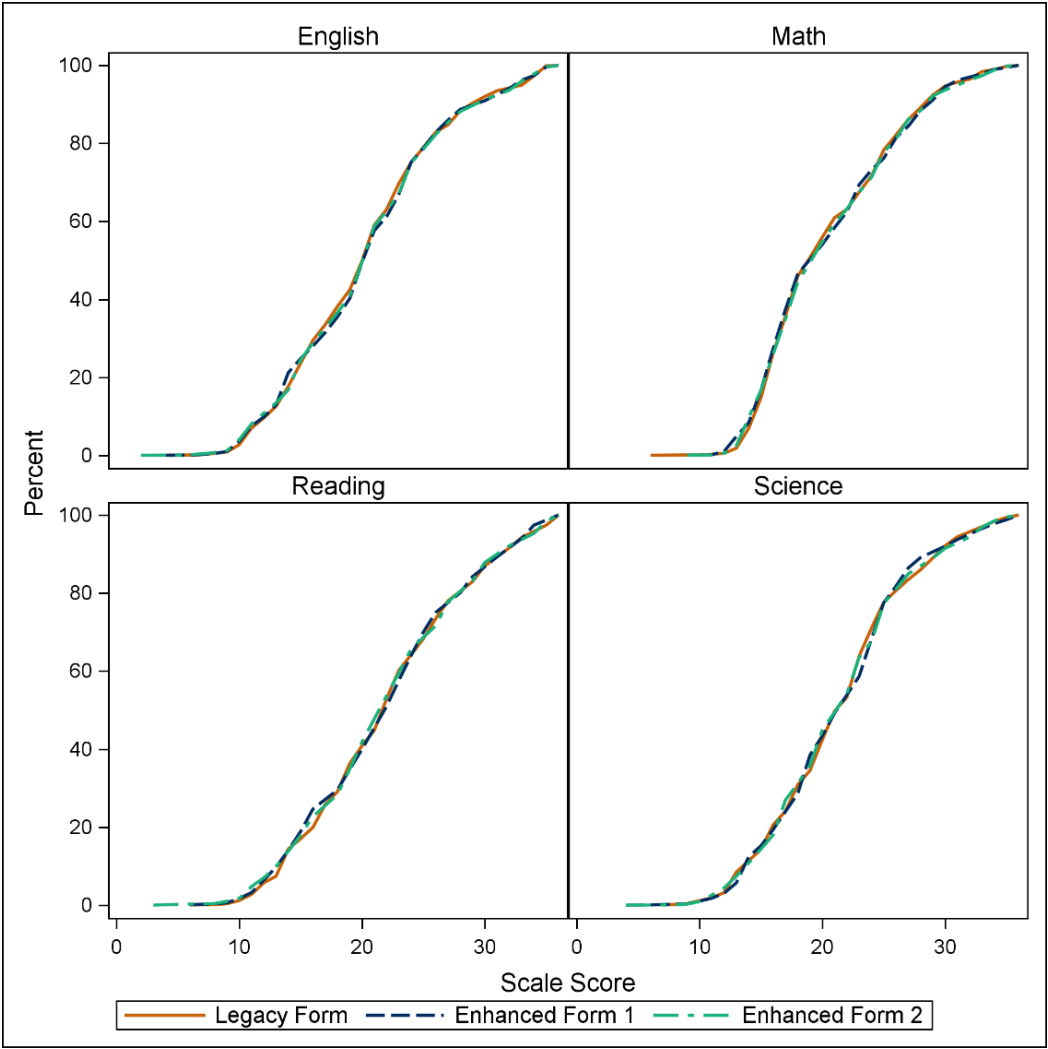
规模得分描述性统计和分布

使用IRT真分数链接，ACT得到了两个改进的ACT形式的原始分数到标准分数的转换。章节、STEM、EMR综合和EMRS综合分数的描述性统计在表10中提供。传统形式和改进形式1和2的平均值几乎相同。请注意，EMR综合分数的平均值比EMRS综合分数的平均值低约0.2。图4展示了相对累积标准分数分布。正如预期的那样，等值后，标准分数的描述性统计和分布非常相似。

表10. 规模得分描述性统计

得分	表格	平均数	SD	偏度	偏度
English	遗产	20.68	6.31	0.36	-0.38
	增强 1	20.78	6.40	0.29	-0.45
	增强 2	20.76	6.44	0.28	-0.43
数学	遗产	20.85	5.44	0.63	-0.50
	增强 1	20.84	5.60	0.59	-0.59
	增强 2	20.93	5.57	0.61	-0.52
阅读	遗产	22.44	6.59	0.22	-0.71
	增强 1	22.37	6.67	0.14	-0.75
	增强 2	22.41	6.71	0.10	-0.70
科学	遗产	21.76	5.74	0.23	-0.38
	增强 1	21.79	5.73	0.26	-0.27
	增强 2	21.78	5.83	0.26	-0.42
STEM	遗产	21.55	5.33	0.47	-0.51
	增强 1	21.56	5.35	0.47	-0.44
	增强 2	21.61	5.39	0.48	-0.51
电子医疗记录 (Electronic Medical Records) 复合材料	遗产	21.33	5.63	0.36	-0.59
	增强 1	21.33	5.58	0.28	-0.69
	增强 2	21.37	5.62	0.29	-0.61
EMRS 复合材料	遗产	21.57	5.53	0.34	-0.59
	增强 1	21.57	5.44	0.30	-0.62
	增强 2	21.59	5.50	0.30	-0.62

图4. 规模得分相对累积分布



人口不变性评估

等值函数预计将具有人群不变性。也就是说，等值函数对于不同的子人群和整个人群应该是相同的。尽管这一属性在实际情况下从未完美地保持，但如果要等值的各种试卷之间有更多差异，预计等值函数将更多地依赖于人群（Dorans & Holland，2000）。由于增强形式1和2中测试规格的变化，ACT通过比较使用等值样本中的不同子组和使用整个等值样本所获得的转换表，来检验人群不变性属性。

人口不变性评估仅适用于具有足够样本大小的较大子群体。因此，在本研究中，对男性和女性子群体以及总群体进行了链接。将每个子群体的转换表与总群体的转换表进行了比较。为了总结转换表差异，对每个比较计算了加权均方根差异（WRMSD）统计量。

每个比较的WRMSD统计量按测试形式呈现于表11中。所有比较的WRMSD值范围从0.1到0.49。这些统计数据都小于半分值，表明了子组中链接结果的不稳定性。此外，还对11份最近被等同化的传统ACT格式进行了类似分析，以作为比较的基准。在表11中呈现了这11份传统格式中的WRMSD统计数据范围。几乎所有增强形式WRMSD统计数据都落在此范围内，表明增强形式的人群不变性与传统形式相似。这些发现为链接结果的稳健性提供了额外证据。

表11. 子组与总组比较的均方根误差 (WRMSD) 统计

章节	表格	女性 versus 总计	与男性总数
English	增强 1	0.24	0.36
	增强 2	0.13	0.17
	之前的遗产形式	0.10–0.43	0.12–0.55
数学	增强 1	0.30*	0.20
	增强 2	0.23	0.21
	先前遗留形式	0.08–0.27	0.10–0.30
阅读	增强 1	0.31	0.49
	增强 2	0.23	0.32
	先前遗留形式	0.16–0.44	0.17–0.55
科学	增强 1	0.34*	0.19
	增强 2	0.11	0.11
	先前遗留形式	0.12–0.24	0.14–0.31

*此值不在范围内。

时序分析及结果

学生对于传统和增强版测验的反应及反应时间均进行了分析。此次评估基于快速响应率、完成率、跳过率以及关于学生是否觉得有足够时间的调查反馈。结果在下文部分呈现。

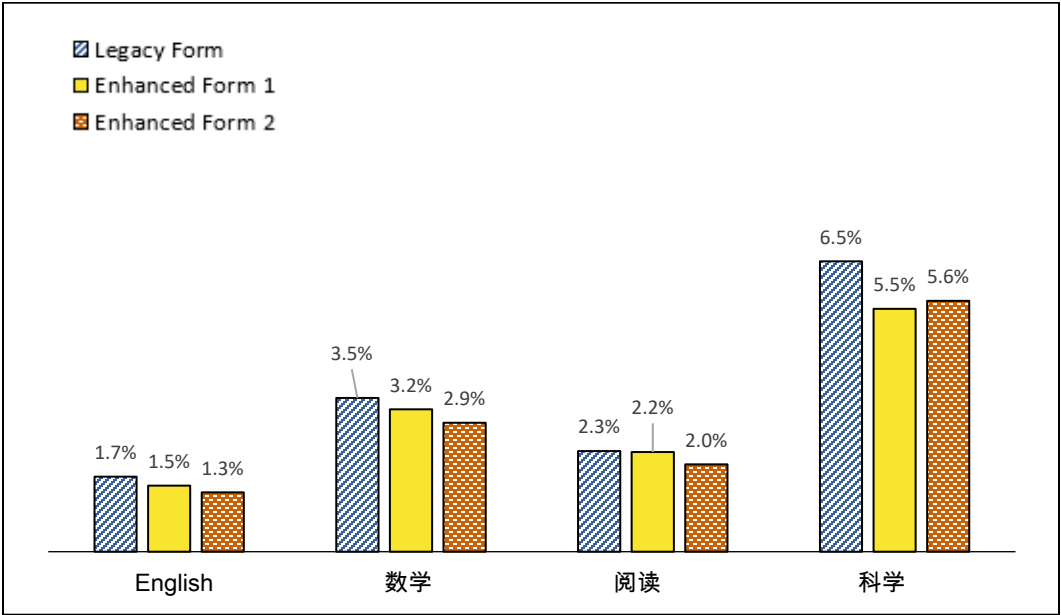
快速响应率

对快速响应率的分析可以揭示延长每个项目分配时间对被试测试行为的影响。在本研究中，如果一个项目被快速响应，那么其响应时间应少于5秒。

图5显示了在所有三种表格中，受试者在5秒内回答的平均项目百分比。如图所示，科学类目在传统和增强表格中都有最高的快速响应率。平均而言，Legacy表格中的6.5%的项目在5秒内被回答，而Enhanced Forms 1和2中的项目分别有5.5%和5.6%在5秒内被回答。对于所有

测试部分，受访者在增强版形式上的快速响应项目平均百分比略低于传统形式，相差0.1%至1%。

图5. 具有快速响应的平均百分比



完成率

谁完成测试，可以...
考试完成率，定义为以不同方式计算的已通过考试的考生的百分比。图6显示了完成传统形式和增强形式的考生的百分比，其中完成率定义为回答了所有考题的考生的百分比。所有 在过去的五项中，无论受试者是否有足够的时间回答问题或快速反应，都不考虑。如图6所示，阅读和科学加强表单的完成率略有上升（1%–2%），但英语（1%–3%）和数学（3%）加强表单的完成率略有下降。

图6. 完全回答最后五道题的考生的百分比

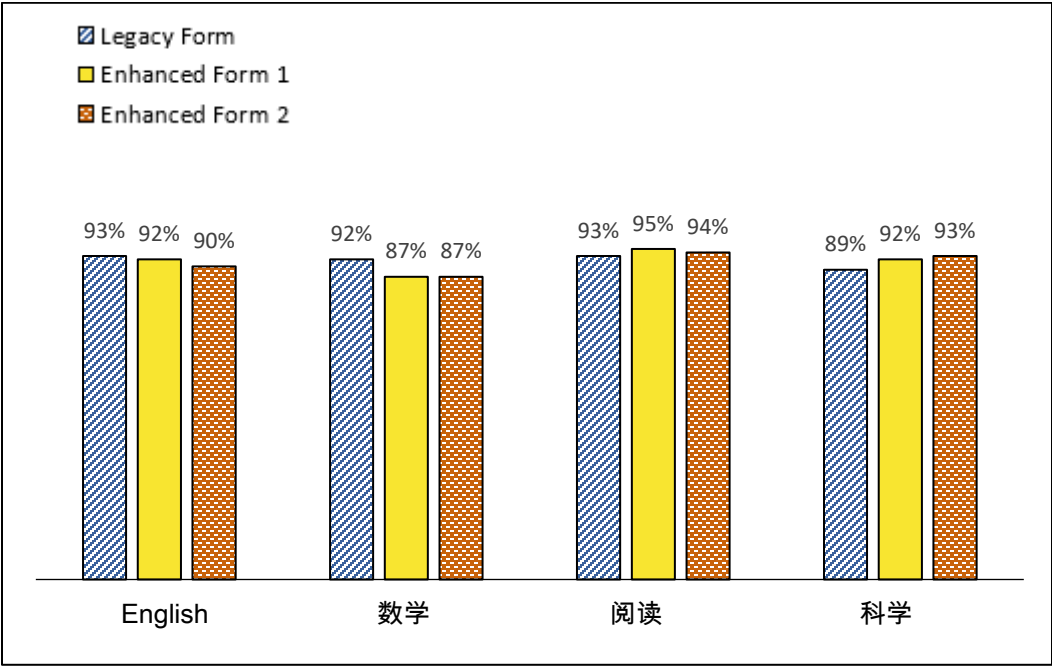
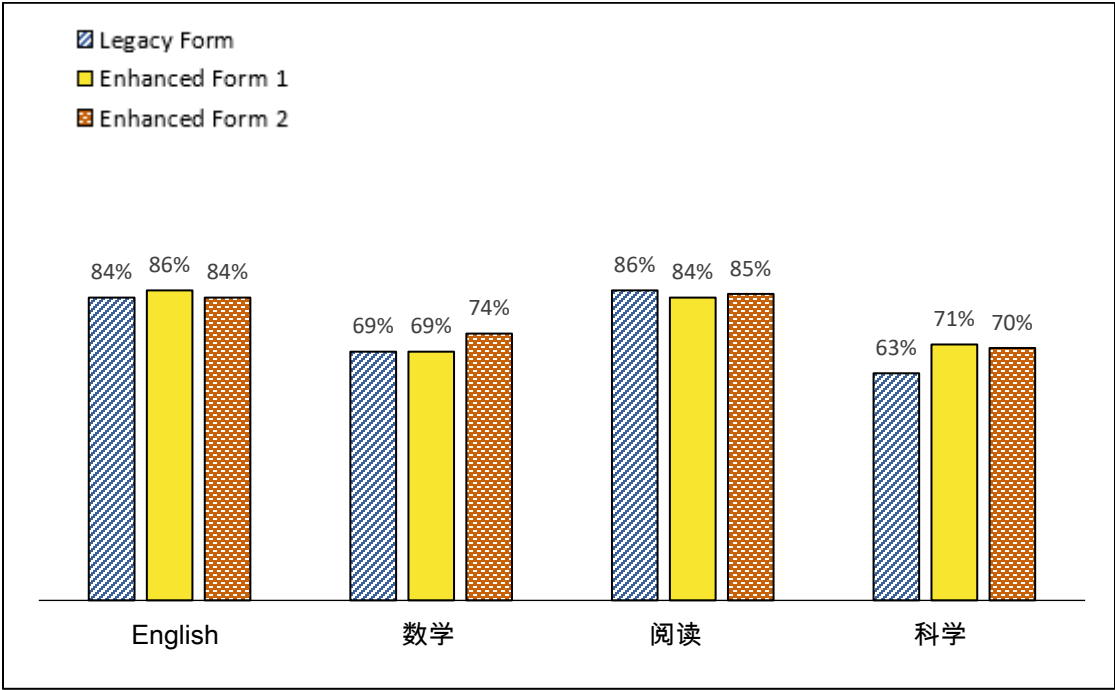


图7显示了在排除答题者在最后五道题中任何一道题用时少于5秒的受试者后的完成率。根据这一计算，与旧版表格相比，英语增强版1的完成率略高2%。数学增强版2的完成率提高5%。科学增强版1和2的完成率分别提高8%和7%。然而，阅读增强版的完成率比旧版低1%至2%。

图7. 移除快速响应后的完成率 (最后五项)



两种方法均显示出在增强形式上的科学完成率的提升。对于其他测试，完成率根据所使用的方法略有增加或减少。

省略利率

被省略的项目数量（即未回答的项目）也用作衡量给定时间是否足够完成测试的指标。图8显示了所有受试者平均省略项目的百分比。对于英语，在传统形式和增强形式之间没有观察到明显的模式，因为一种形式的省略百分比略有增加，而另一种形式略有减少。在数学的情况下，增强形式中省略项目的百分比略有增加。科学和阅读的图案，增强形式的省略率低于传统形式，比英语和数学的图案更为明显。

图8. 考试参与者平均省略项目的百分比

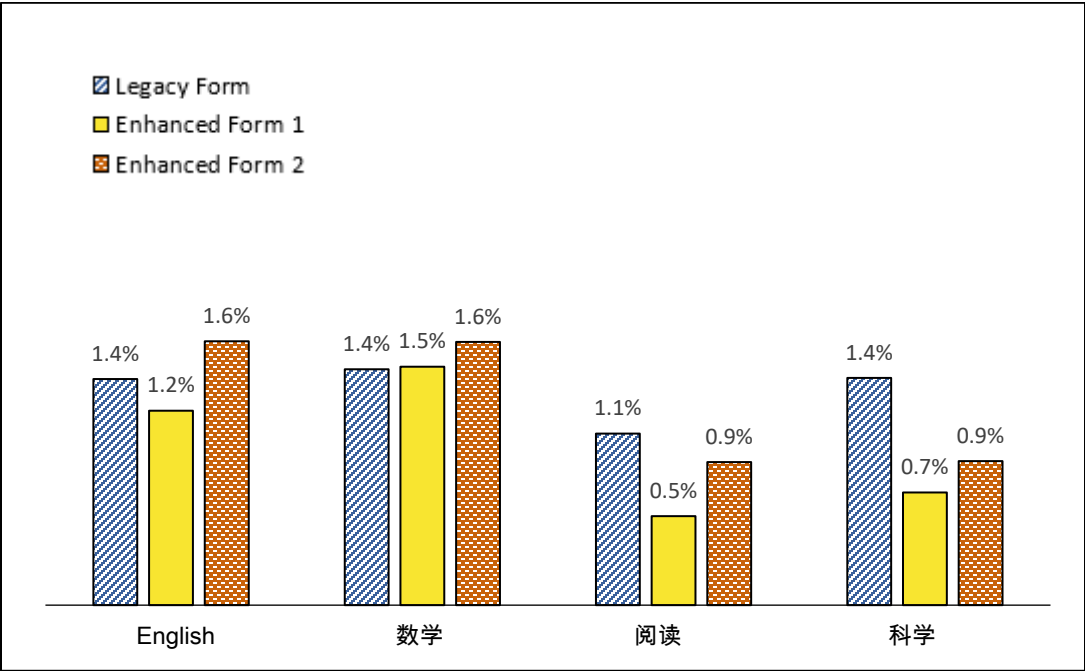
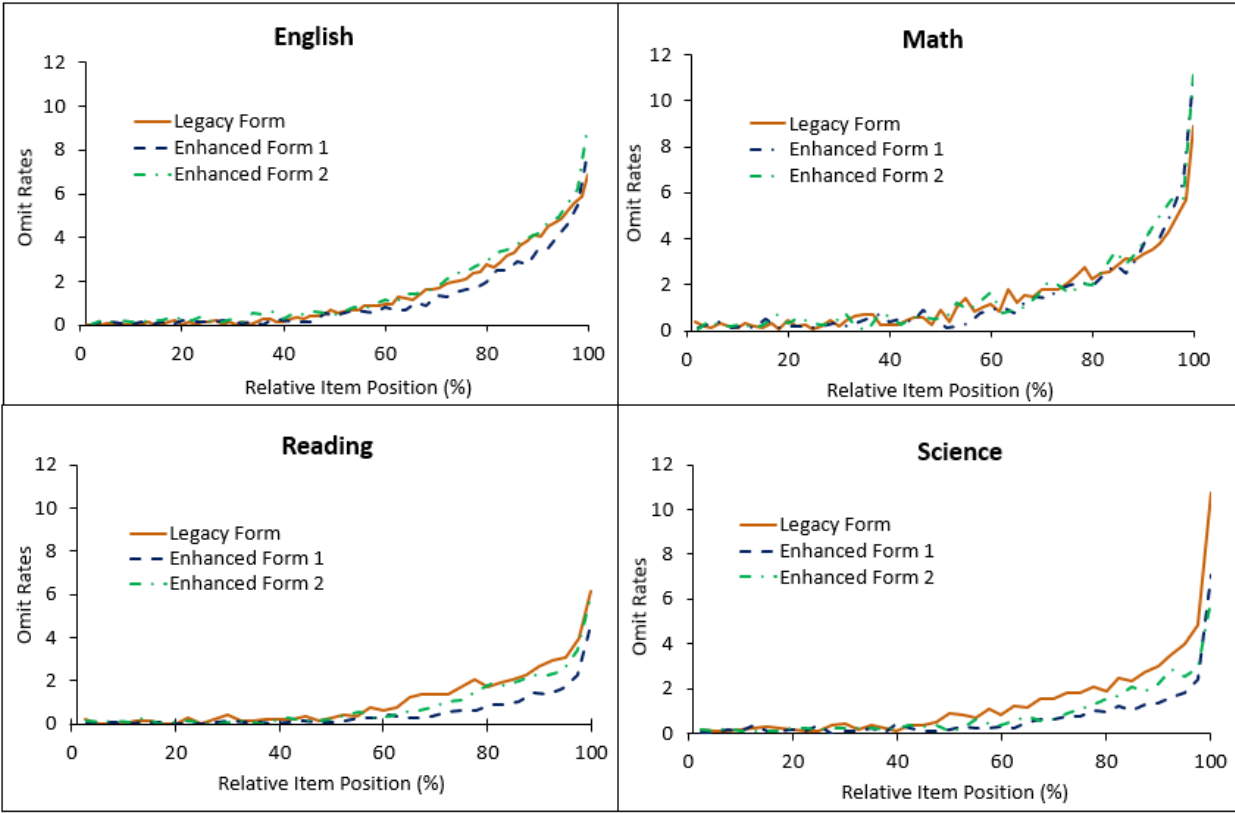


图9展示了所有项目的遗漏率，通过显示遗漏率开始增加的具体项目，提供了一个更全面的视角。为了比较传统形式和增强形式的遗漏率，这两种形式具有不同的测试长度，ACT将项目顺序放在一个共同尺度上，如图所示为相对项目位置百分比。 x -轴在图9中。图中的每个方块代表一个测试区域。 y - 轴显示省略特定项的考生的百分比。曲线越高表示省略该项的考生百分比越大。图9表明，在所有科目和形式中，省略率在测试大约进行到一半时开始增加。对于传统和增强形式，省略率最高的部分是最后几个项目。

忽略英语和数学的常规和增强形式的比例，并没有显示出明显的模式。在测试的第二半部分，对于阅读和科学而言，增强形式的比例低于常规形式。

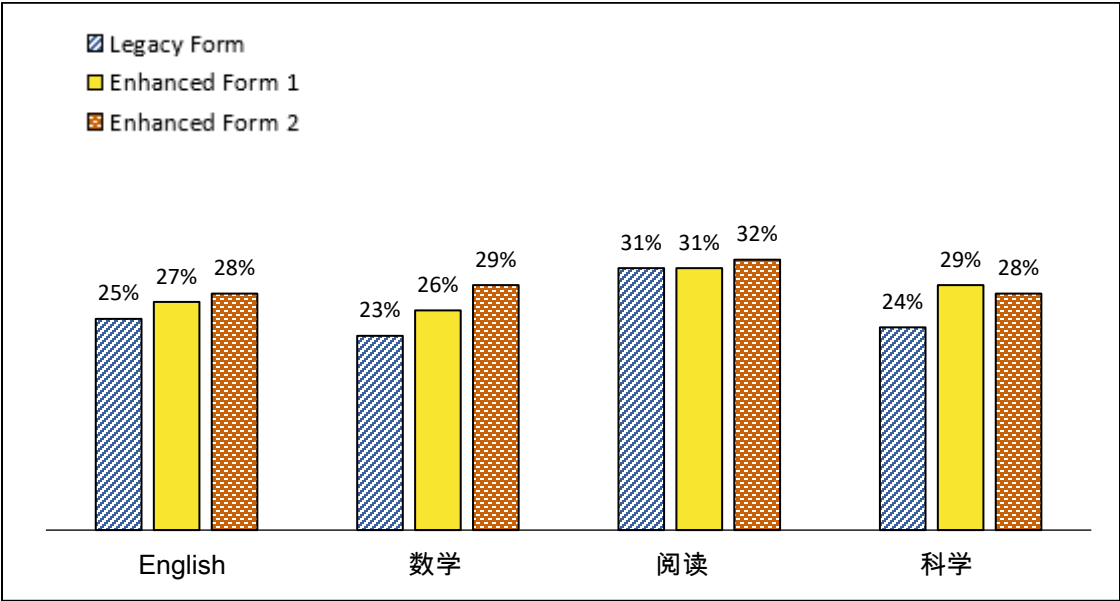
图9. 每项内容的省略率



访问次数超过一次的物品百分比

如图10所示，在所有测试中，访问超过一次的项目在增强型表格中比在传统表格中更高。传统型和增强型表格之间的最大差异在于科学和数学，分别为4%至5%和3%至6%。假设考生仅在完成所有项目一次并剩下足够时间后重新访问项目，这些结果表明考生在增强型表格上有更多时间审阅问题。

图10. 被访问超过一次的项目百分比

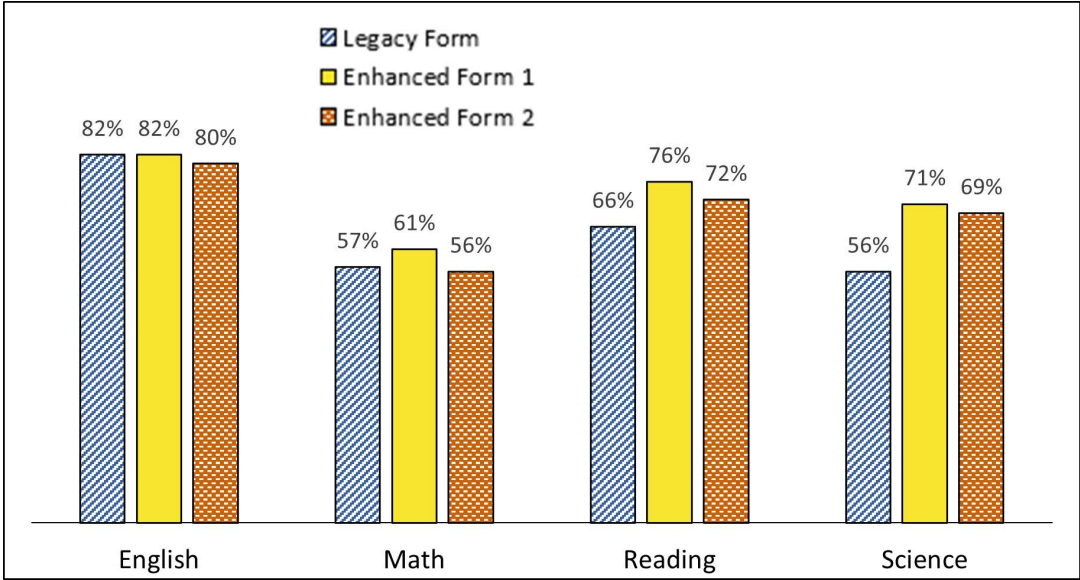


调查结果：感知时间充足程度

受试者被问及他们是否有足够的时间完成测试的每一部分。选择“是的，我有很多我需要的时间”、“是的，我稍微有点多余的时间”或“我刚好有足够的时间”的受试者被认为有足够的时间完成测试。

图11显示了按测试形式划分的回应百分比。对于英语，表示他们有足够或更多时间完成测试的考生的百分比在所有三种形式中相似。对于数学，传统形式和增强形式之间的百分比不一致（对于一种增强形式，它们高于传统形式，而对于另一种，则低于）。对于阅读，表示他们在增强形式1和2上有足够时间的考生的百分比分别比传统形式高10%和6%。对于科学，增强形式1和2的百分比分别高15%和13%。

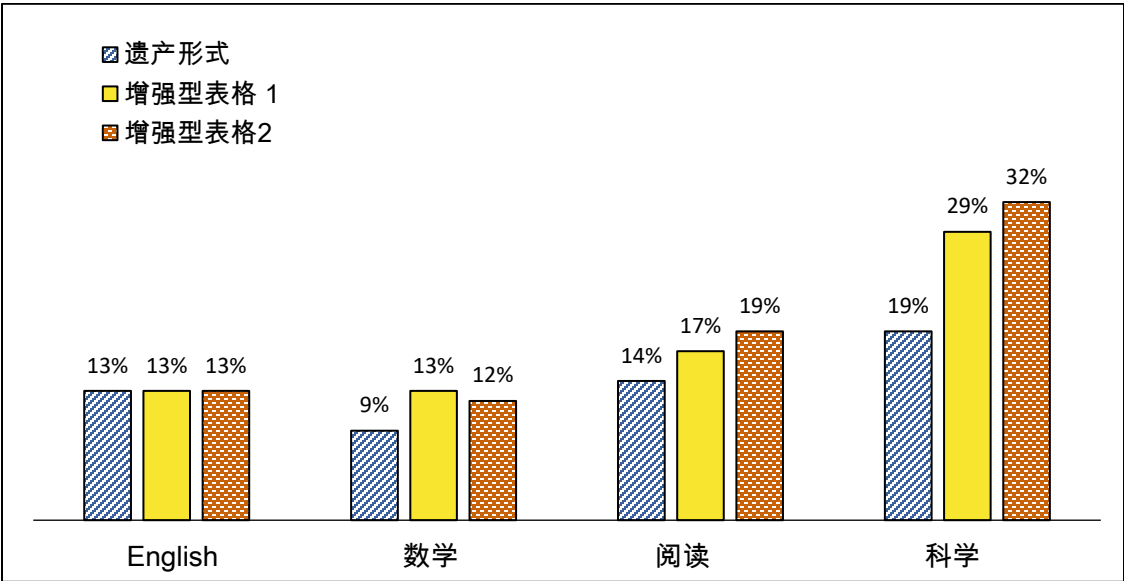
图11. 对感知时间充足度的调查回应



延长考试时间考生

为了观察时间加倍安排对延长时间学生的影响，ACT将时间加倍的学生纳入了研究中。图12显示了在标准时间内提交考试的长时学生比例。在数学（3%–4%）、阅读（3%–5%）和科学（10%–13%）方面，在标准时间内提交考试的长时学生比例有所增加。

图12. 学生提交考试的时间超出标准测试时间所占的比例



摘要

总结来说，改进后的ACT考试测量相同的结构，并在与旧版ACT考试相同的量表上报告分数。虽然改进对学生的表现影响微乎其微，但有一些证据表明，对测试完成率和学生对完成测试所需时间充足的感知产生了一定的小影响。尽管测试长度的缩短导致测量精度预期下降，但对分数解释的影响很小。因此，改进后的ACT考试和旧版ACT考试的分数可以互换使用。

致谢

作者们想要感谢辛·李 (Xin Li)、克里斯蒂娜·安古亚诺-卡拉斯科 (Cristina Anguiano-Carrasco)、凯特·沃尔顿 (Kate Walton) 和达娜·穆拉诺 (Dana Murano) 对报告的审阅和贡献。

参考文献

2024. *ACT 技术手册* .

https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *《教育测量杂志》* , 37 (4), 281–306.

[-3984.2000.tb01088.x](https://doi.org/10.1111/j.1745-3984.2000.tb01088.x)

<https://doi.org/10.1111/j.1745>

胡, 李. T., 与本特勒, P. M. (1999). 协方差结构分析中拟合指数的截断标准：传统标准与新替代方案. *结构方程模型* , 6 (1), 1–55.

[0705519909540118](https://doi.org/10.1080/0705519909540118)

<https://doi.org/10.1080/1>

科伦, M. J., 布伦南, R. L. (2014). *测试等值、缩放和链接：方法和实践* (第3版) 施普林格出版社。

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). 条件测量误差的标准误差用于量表得分。 *《教育测量杂志》* , 29 (4), 285–307.

<https://doi.org/10.1111/j.1745-3984.1992.tb00378.x>

Livingston, S. A., & Lewis, C. (1995). 评估基于测试分数的分类一致性及准确性. *《教育测量杂志》* , 32 (2), 179–197.

[.1745-3984.1995.tb00462.x](https://doi.org/10.1111/j.1745-3984.1995.tb00462.x)

<https://doi.org/10.1111/j>

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory

应用心理学测量 , 7 (2), 201–210.

<https://doi.org/10.1177/014662168300700208>



关于 ACT

美国ACT组织正在改革大学和职业准备路径，以便每个人都能发现自己的潜力并实现之。ACT的资源 and 评估工具基于65多年的研究，这些资源、评估、研究和就业资格证书在美国及全球范围内得到学生、求职者、教育工作者、学校、政府机构和雇主的信赖，以帮助人们在生命各个阶段实现其教育和职业目标。欢迎访问我们的网站：www.act.org。