



英语

听音选字母——找到错误的原因

“听音选字母”是为了考察一、二年级学生对英语字母发音的辨识，其题目指导语为：接下来你会听到一些字母，请你把听到的内容选出来，点击图片。如果需要修改，再次点击图片就能取消，确认后点击“完成”按钮。



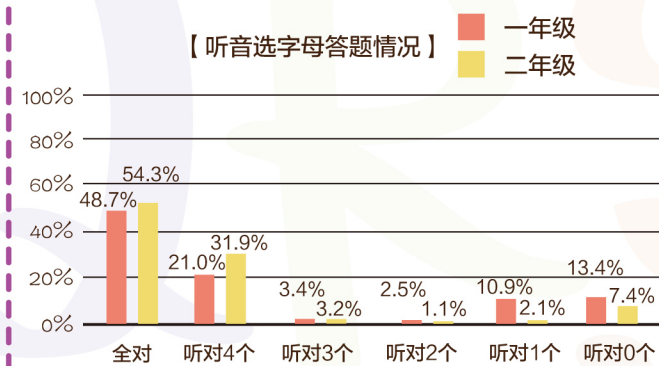
接下来你会听到一些字母，请你把听到的内容选出来，点击图片。
A s d F g H j
k L D

【计分规则】

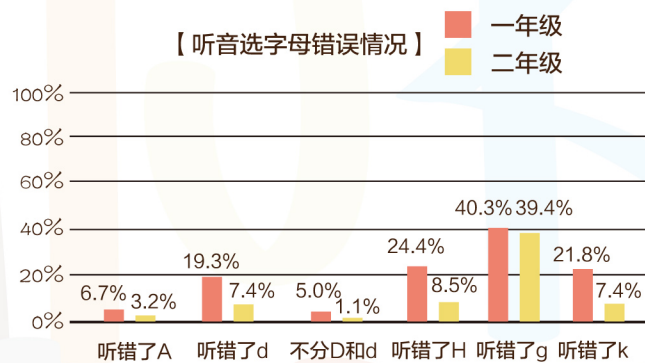
“听音选字母”需要学生根据录音选出正确的 5 个字母，每选对 1 个计 1 分，满分为 5 分。

【测评结果】

根据统计，一年级学生在“听音选字母”上的平均得分为 3.78 分（满分为 5 分），而二年级学生的平均得分为 4.31 分。



从上图可以看出，在所有参加测试的学生中，有 69.7% 的一年级学生和 86.2% 的二年级学生能够根据录音选对 4 个以上的字母，说明大部分学生良好地掌握了英文字母的发音。然而仍有 13.4% 的一年级学生和 7.4% 的二年级学生没有听对任何一个字母，需要教师关注。



从上图可以看出，在未获得满分的一年级学生（共 61 人）和二年级学生（共 43 人）中，分别有 40.3% 的一年级学生和 39.4% 的二年级学生将“g”的发音与“j”相混淆。此外，相对于二年级学生来说，一年级学生更容易在“d”、“H”和“k”的发音上出现错误。

过去的考试中，一般都是给学生 2 个选项，选其一，学生如果不会，猜中的概率也有 50%。而通过 iPad 测试后，我们可以清楚的了解到学生的答题时间，对几个，错几个，是掌握辅音较多还是掌握元音较多，是分不清大小写还是分不清字母，这些数据都可以在之后的教学中加以利用并进行针对性培训，这些都是过去的考试远远不能呈现出来的。

在本次测试中，还有许许多多诸如上面类型的新颖题目。这些题目不仅仅测试了学生的学科知识方面的能力，同时还能够看到学生在学习过程中的学习素养。更为重要的是，我们的成绩将不仅仅显示学生的结果，而是通过考试发现学生在学习过程中存在的问题，这些问题在传统的考卷中，可能是无法被发现的，这就为教师有的放矢地进行教学实践和教学改革提供了数据上的支持，更为制定学生的全面发展方针提供了依据。



扫描二维码，了解更多的信息
同时也可登录网址：www.psyllife.cc

情境任务，未来的评价形式？

文 / 徐君逸

当小车从斜坡上滑下来，当两辆列车从两个城市相向而行，当水池里一个水管在放水、一个在蓄水，当一群鸡和一群兔子关进了一个笼子……当面对这些经年累月的“题库”时，作为学生测评题目设计者的我们，不禁会自问：这些题究竟在考察学生什么样的能力？这些能力能够帮助学生适应多变的社会吗？或者说，即使这些题目学生都能正确、快速地解答，他们的能力就真的增强了吗？

会解这些题，可能仅仅只是解题的能力而已。

相比这些题目，对下面这道题的回答可能更能考察学生的综合能力：小明做了二十年的题，求小明的心理阴影面积？

不妨试一试情境任务测试

除了学科评价，我们还可以如何“另类”地、更加多元化地去评价学生？

随着近几年越来越多的研究者认为有教育意义的游戏对孩子的思维能力、空间能力、想象力、社会能力都有较为明显地促进作用。游戏化成了教育手段的新宠儿，那么不由得我们深入思考：游戏能不能作为一种评价方式？学生在真实的任务情境中所反映出来的能力，会不会比面对试卷的抽象题目所表现出来的能力更加真实？这种考试在“玩”的形态下有没有科学的讲究？能够挖掘出一些什么呢？

对比古板、非现实的传统考试，游戏化任务情境测试基于学生喜欢的情景，将任务变化为现实的挑战。学生就像电影中的英雄主角们，接受各式各样的挑战任务，依靠自己的智慧和努力克服一项项难关，最后获得成功。这个过程不仅充满趣味性，同时增强了学生的自我能动性，在游戏化任务情境中激发学生的积极潜力。

接下来，我们就具体介绍与学科评价不一样的任务测试：

1、让学生成为情境的主角

每个孩子心里都有一个主角梦。在题目里，让他变成“小明”“小红”的替身去解决问题，这该多无聊？他为什么不能成为主角，自己解决问题呢？

所以，情境任务的第一个要素，就是要让学生成为情境的主角。所有的戏剧故事，都是从“主角遇到一个困难”，“主角领到一个任务”开始。而从这样设定中起航的考试，是不是也会让学生格外兴奋呢？

其实，我们的生活，不正是一场考试？



学生们在做什么？也许是在一起讨论语数外的习题？然而我们在翻看学生手边的任务书，就能发现其中奥秘：



亲爱的同学：

你好！

我是宇宙空间站的赛博士，我和我的助手小思思在探索太空的时候已经勘测到了一些星球了，可是这样还远远不够，我们还需要勘测更多的星球！

每个星球能生产的东西也是不一样的，我们只有探索越多的星球，才能获得各种各样的产物，这样我们的宇宙空间站也能运用这些产物创造各种各样好吃或者好玩的东西啦！

记得把你和你的同伴探索到的星球和能够创造的东西记录在纸上！

最爱你们的
PSY博士



与纸笔化考试直接纯粹的试卷不同，学生需要根据任务书的要求进行活动，也就是“情境”游戏化。这种形式不仅仅充满趣味性，在“玩”中引起学生的积极兴趣，同时解放了学生的双手，让学生不再为复习而紧张焦虑，取而代之的是兴奋和跃跃欲试。在测试之前，学生只要带着一颗放松的心参与到活动中，将自己最真实的能力展现在任务中；既充分感受到快乐，也能够在完成之后获得成功的信心。考试的过程，更是一个充分展现自我的过程，学生不需要为了一分的得失而斤斤计较。为了完成任务，他们可以有千万种方法与途径去解决，每一种尝试都是能力与思路的展现。

以“星球大作战”为例，任务情境考察的不再是学生的学科知识，而是能力的应用——学习创新思维、合作交往能力的具体表现。学生需要理解任务书，尽可能地拓展想象，同时和同伴一起思维碰撞，获得“智慧果实”。可见，任务情境测试是一种能力运用的综合性测试。尝试都是能力与思路的展现。

任务情境的创设不仅限于一个年级，随着年级的提升，情境的创设也具有梯度，高年级的学生拥有较大的自主性，任务内容越发复杂，考察维度也随着年龄增长产生变化。下面具体介绍不同年级的任务情境，火眼金睛的读者可以来找一找不同之处。同时之后的文章中也会以高年级的游戏化考试为主进行介绍。

年级	任务内容	现场照片
一年级	“程老师，你体育课要用的篮球被同学们找到了！我现在就让同学们每个人拿一个过去给你！” 需要学生携带球去指定教室。	
三年级	“闯关任务共有四关，你可以自己安排闯关的顺序，同时，有些任务可以自己一个人闯关，有些任务可以和小伙伴组队闯关，你只有 30 分钟的时间完成所有任务哦！” 1、创作一个可爱的卡通图案吧 2、和小伙伴一起探索星球 3、帮助老师找到他想要的图书 4、这些运动你知道不知道	

【三年级星球大作战】



水

+



火

=



水蒸气



水

+



生命魔法

=



鱼

任务白皮书，智趣结合。
你能看出这是在考察什么吗？

游戏化考试并非低年级学生的专利。我们只要设计得当，高年级也可以享受快乐的考试过程。而从任务的描述中，我们还可以明显的发现低年级与高年级学生在任务情境上是存在着较大不同的。低年级学生的任务集中于学生的理解能力与运动能力，而高年级的任务情境更加强调了学生的适应能力、合作能力与学习能力。

那么，这些能力我们是如何从任务情境中进行挖掘的呢？

2、过程性观察，发现平时见不到的学生

实施游戏化任务的整个流程，完全以学生为主——强调了学生自己的选择。游戏化考试每批测试宜少不宜多，以 10 人一批间隔开始，确保环节实施的流畅。整个游戏过程中，在多个任务点我们都会设置游戏“NPC”——为学生提供不同的服务协助。

在游戏起点，会由任务 NPC 将任务书交到学生手中，并进行整体任务的讲解。在开始时，还会为每个学生分配二维码，只要用相应的软件进行简单扫描和登录即可，随后任务则给予学生最大的自由程度，让学生来选择要先完成什么，最后完成什么。

前面已经介绍过在每次任务开始之前，学生都会收到一份独一无二属于自己的二维码。二维码中包含了学生的姓名、班级甚至学号。当使用 IPAD 的应用进行二维码扫描后，学生的信息就会被记录到后台的数据中。这些信息不仅仅包括了学生的基本信息，同时包括了扫描的时间，这个数据又具有什么意义呢？



当学生结束任务进行再次扫码，这个数据就不再是一个时间点。通过两次扫码的时间差就可以知道该学生完成任务所花的时间。游戏化考试大大拓展了学生活动的空间，时间由学生自行把握。因此，游戏化考试随着空间的拓展与时间的自由，学生在进行任务时行为表现都可以作为数据，最终作为学生评价的依据。

另外，低年级学生所参与的任务同样设置了提示“NPC”，在学生无法完成任务或需要提示时进行适当地提醒，帮助学生最终完成任务。

游戏化考试并不将结果作为唯一的标准，而是多种因素所呈现出的学生行为的综合体现。有时候，过程性记录能解释学生行为结果的原因，更能帮助教师找到学生能力不足的问题所在。游戏化考试使得考试不再局限于对错决定结果，通过信息化数据记录、客观考察学生的能力水平。

测试分配了多位 NPC 在各个任务区域，为学生发放相应的任务书，观察学生完成任务的情况。除了时间之外，根据任务要求的不同，观察员会记录相应的学生行为、学习结果，并根据指标进行打分。

每项任务的指标以水平为基准，分为 Lv1、Lv3 和 Lv5，并附以具体的描述说明。观察员根据学生的表现为学生划分水平。指标中并没有所谓的“零分”，这同样也是游戏化考试“玩”的本质体现。

任务名称	Lv1	Lv3	Lv5
星球大作战	无合作，或产生矛盾时并不能解决。	偶尔出现交流与合作，或以一方为主，非共同探讨。	良性合作，出现争执时能尊重对方的意见。
电脑大作战	无法完成，只能大体模仿视频中图形创作，细节上并未做到一致。例如：树干下两层树叶并没有置于下一层；3 层树叶大小一样。	能基本模仿视频中的图形，注意到细节，有简单的自我创新，例如颜色的变化或有别于视频图形的添画。	能自主创新画一些与视频里不一样的卡通形象，且美观漂亮。

这些指标的呈现以结果为主，却离不开观察。同时，观察的过程也将会作为实例说明数字背后所隐藏的“秘密”。

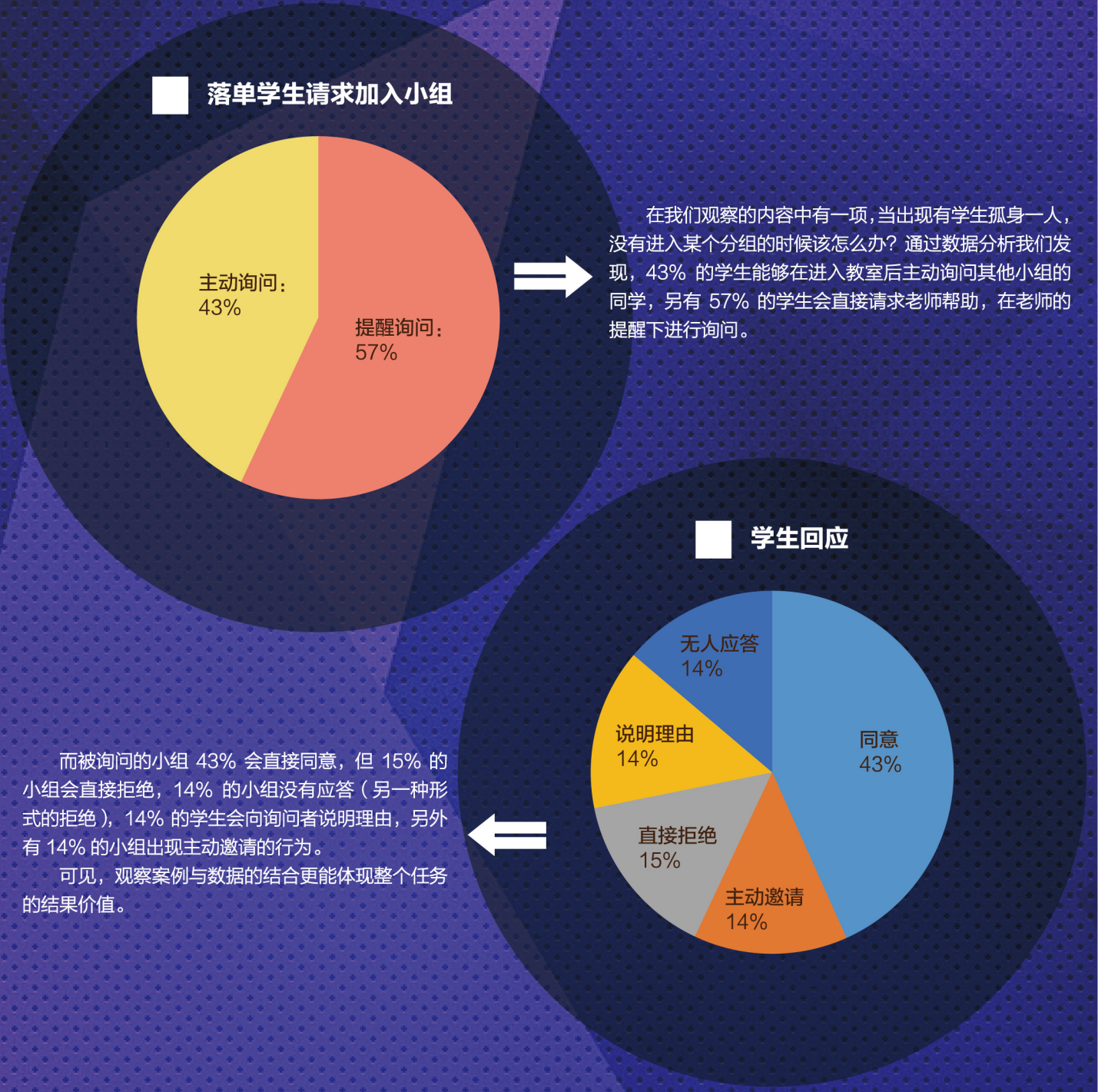
例如，星球大作战的任务内容需要考察小组中学生们合作沟通能力，因此 NPC 们需要通过照片、文字记录学生的合作行为，随后进行过程性整理，根据指标打分。然而在分数之外，我们却意外发现了一些小事件。

虽然从整体的合作能力上看，大部分学生都表现出良好的合作行为，但仍然在任务过程中会发现一些“不和谐现象”，我们来看看老师们的观察感受：

有单独进入教室的学生在提出想要入队时，还是会发生没有人回应或是直接拒绝的情况。

不仅在选择队友时，学生更倾向于同班同学，即使是一组内（人数大于 2 人）学生也更倾向于与同班同学进行交流。

在完成作品后，提醒学生写名字，大部分学生只写自己的名字就放下笔离开教室，留下组员自己填写姓名。



在我们观察的内容中有一项,当出现有学生孤身一人,没有进入某个分组的时候该怎么办? 通过数据分析我们发现, 43% 的学生能够在进入教室后主动询问其他小组的同学, 另有 57% 的学生会直接请求老师帮助, 在老师的提醒下进行询问。

而被询问的小组 43% 会直接同意, 但 15% 的小组会直接拒绝, 14% 的小组没有应答（另一种形式的拒绝), 14% 的学生会向询问者说明理由, 另外有 14% 的小组出现主动邀请的行为。

可见，观察案例与数据的结合更能体现整个任务的结果价值。

3、学生综合能力的反映

进行过程性观察的目的是为了综合性评价，那么我们在游戏化任务情境测试中的综合能力，到底体现在何处？

如果说游戏化学科测试是通过学科相关的题目考察学生对学科知识的掌握，那么游戏化任务情境测试注重的是学生综合素质的体现。测试过程中并不需要考验学生任何的学科知识，而是将学生综合能力作为评价内容，例如：学生是否和同伴合作完成指定任务——包含了对学生合作能力的考察；学生是否能够学习视频中的方法自主地使用电脑进行创作——包含了对学习创新能力的考察……

游戏化任务情境测试以快乐的情境形式，通过实际的任务环境引发学生不断尝试解决问题的行为，挖掘学生真实的反应与能力,比单独依靠纸笔测试所获得的成绩高低更加具有启发性。

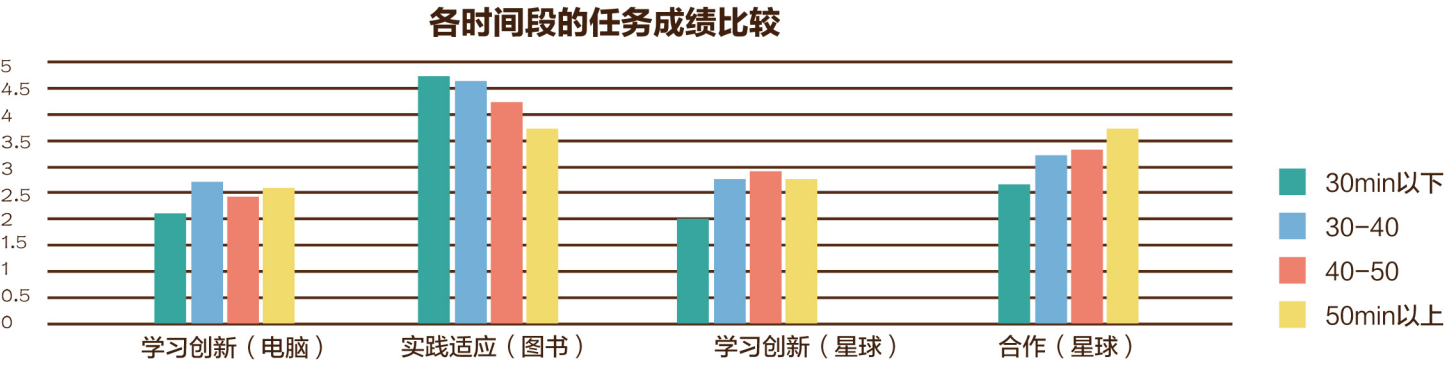
学生综合能力主要包括合作责任、学习创新、自信自主及实践适应。有些任务只考察一种能力，有些任务考察了两种能力。维度根据水平指标进行加权计分，并根据整体水平获得学生四个维度的相对位置，是不是比学科成绩更加丰富，又代表了学生学习、生活的基本能力呢？

我们便来具体看看学生能力在游戏化考试中所产生的化学反应！



【时间真的过去太久了】

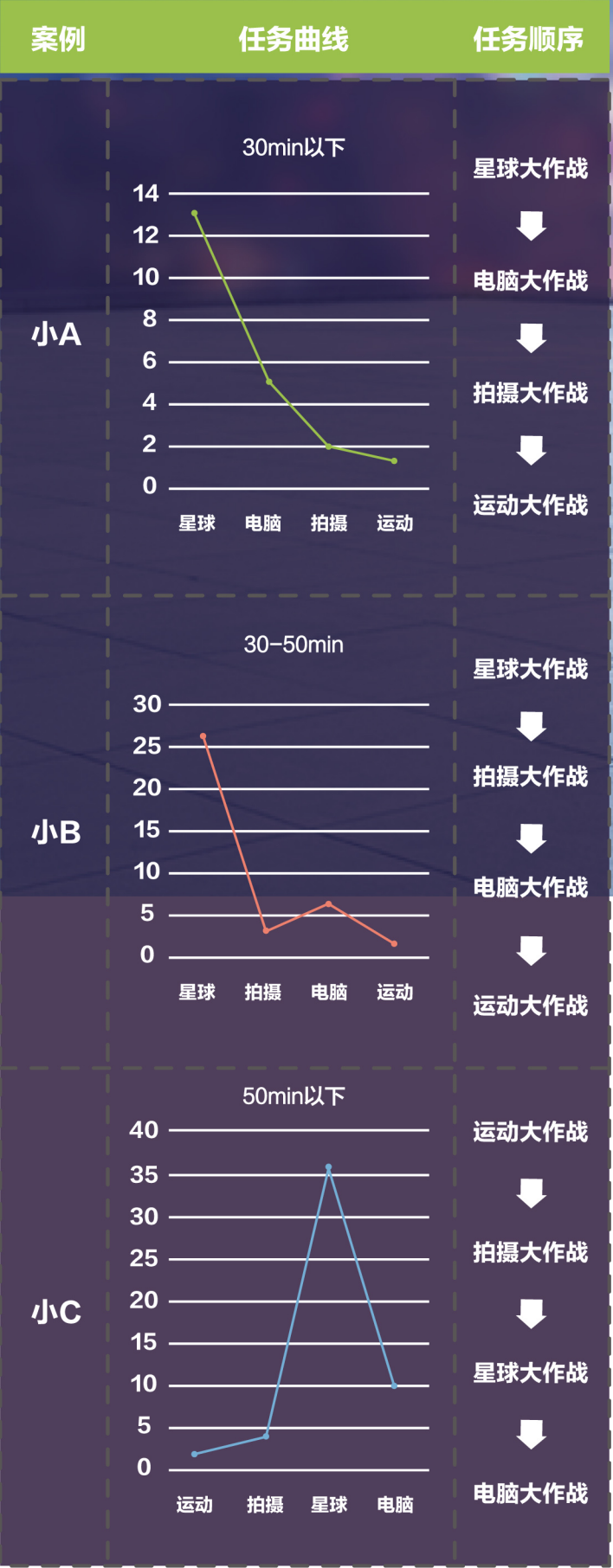
在游戏化考试中，如果没有了时间限制，如何一番景象？



将每个学生完成时间分为四档:30 分钟以下、30-40 分钟、40-50 分钟及 50 分钟以上，并比较各个任务（除运动大作战外）的成绩，发现学生花费的时间越长，其在星球大作战中的合作表现越好，但图书大作战的实践适应效果却反而下降了。同时，学生学习创新表现并不会因为时间的延长而变得更好或更差。

从任务的难度来看，星球和电脑的难度要高于拍摄及运动，因此导致完成这两项任务的时间要高于后两项。高时间占比预示了每项任务学生的轮换率相对较低，就有可能出现这样的情况：在某学生完成了拍摄和运动任务之后，他将因为还未完成的星球及电脑任务的学生而无法直接开始任务，产生了等待的时间“浪费”。

因此，我们再来看看完成任务时间不同的 3 个学生：小 A、小 B 和小 C，分解他们的任务曲线（完成任务顺序及所花费时间），不难发现三个同学之间选择的任务顺序不尽相同。整体花费时间较短的小 A 首先选择了星球和电脑这两个时间占比较长的任务，而时间最长的小 C 则将星球和电脑任务放在最后，并且花费时间要远远高于小 A 与小 B。



学生在完成任务时对于时间的管理实际上是一种自信自主的表现，是否能够准确把握任务完成时间的进度，合理安排每项任务的时间配比才真正具有较好的综合能力水平。对于有些学生而言，时间真的过去太久。

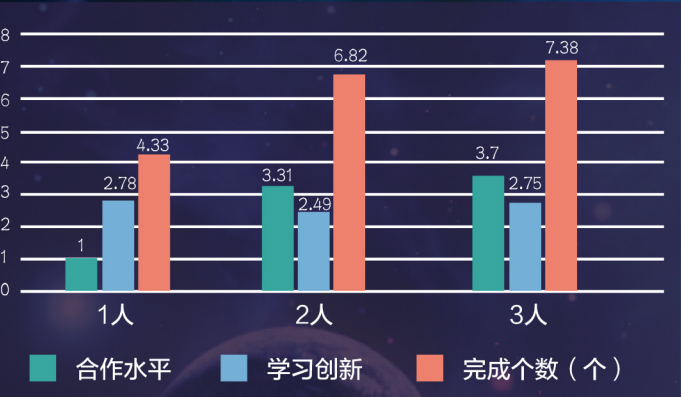
试想一下，在生活中，你是不是有这样的朋友：有人一天什么也没做，却感觉自己一直在忙活；有人一天的工作有条不紊，效率极高。这在孩子们任务的排布中，是不是就能窥见端倪呢？

【集体力量大？】

星球大作战是一项合作性的活动，观察员还会记录下合作小组的人数、作品数量作为依据，考察团队合作的效果。在数据分析时，我们便发现了不少有趣的现象。

在星球大作战的任务要求中需要学生任意组队（1~3 人），若学生不愿组队也可以单独完成。我们将不同人数小组的合作水平、作品结果（科学创新性）以及个数进行了均值比较，发现除 2~3 人小组完成的作品个数会远高于单独行动的学生，3 人小组的合作水平相对较高于 2 人小组。

然而，与合作水平相对的“科学创新性”——作品的质量却并没有因为小组人数增长而有任何的起色，反而单独完成任务的学生，作品质量相对较高。



进一步从相关性来看，合作水平与小组人数、完成个数正相关性达到了极其显著的结果，而与科学创新性几乎不存在相关性。同样验证了集体的力量虽大，但效率并不是由“集体”说了算的。这与心理学上群体的“责任扩散”不谋而合。



【源于题目？源于生活！】

在星球大作战中要求学生根据任务所给的提示，进行创意性的元素或星球合成，既考察了学生创新能力，也考察了学生在创新的同时是否能够兼顾科学性。什么样的作品是有创新性与逻辑性的？我们可以来看看学生的作品：

星球大作战				
年级	某三年级学生		某五年级学生	
要求	根据任务书和同伴一起完成元素的创意星球组成。（给了参考元素）		根据任务书和同伴一起完成星球产物的创意组成。（给了参考星球与产物）	
作品1	风+石头=沙尘暴	Lv5	蜜糖苹果+西瓜糖+阿尔卑斯糖=Candy Lab	Lv5
作品2	恐龙+森林=侏罗纪公园	Lv5	草莓星球+冰雪星球=草莓冰沙	Lv3
作品3	火+生命魔法=凤凰	Lv3	宝石盒+结婚戒指+博物馆=钻戒	Lv1

在众多学生的作品中科学的合成公式不少，但却创意平平。综合了各组学生作品，大多与样例都非常相近，而合成的元素和物品大多也取自任务书中给到的提示参考。而在 Lv5 的几个等式中，学生并没有拘泥于任务书的内容，而是利用任务书里的参考合成了自己身边常见的东西——常见的沙尘暴、小孩子都喜欢的 Candy Lab。可见，游戏化考试中的“创新”并不一定是异想天开的过程，生活的经验非常重要。因为，奇妙妙想同样也需要源自于学生的生活，关键在于是否能把“日常”变为“创意”。

【二次创造or自我创新？】

电脑大作战		
年级	作品	
水平	Lv1	Lv5
某三年级学生		
某五年级学生		

以上 Lv5 的两幅作品是否既可爱又漂亮？电脑大作战同样要求学生能够进行创意作画，在大部分学生仿造视频的圣诞树和小猪进行二次创作时，有的学生已然读懂了任务书中表明“画你喜欢的图案”，尽情表现自己的“艺术细胞”。游戏化考试也是一个平台，让学生自主地表现表达，反应最真实的艺术创新能力。

【能力与任务结果】

对学生而言，学科知识的掌握只是本身知识的积累与应用程度的体现，而掌握学习知识是否能够转化为适应社会、激发潜能的能力，并不能完全依靠学科的成绩体现。学生基础素养测试结果表明：不少低年龄阶段的学生数学成绩良好，但逻辑能力却低于平均水平，而理解能力却较高，这反映出在低年级时数学成绩往往是取决于学生对题目的理解程度——也就是学生的理解能力。而到了高年级，题目所引起的理解差异越来越小，逻辑能力上的缺陷则越来越凸显。因而我们会发现，有些同学在低年级数学成绩良好，但升到高年级之后却逐渐跟不上班级的进度。

将综合能力的结果与学科成绩、基础素养进行相关分析，以二年级为例可以发现，学生的综合能力与他们的记忆力具有较高的相关性，其相关系数达到了 0.213。这说明在二年级学生中，记忆力较好的学生能够更好地完成任务，获得较高的综合能力得分。同时我们也发现，只有在英语学科中呈现正相关性。

二年级	注意	逻辑	记忆	理解	表达	英语
综合能力			0.213*			0.235*

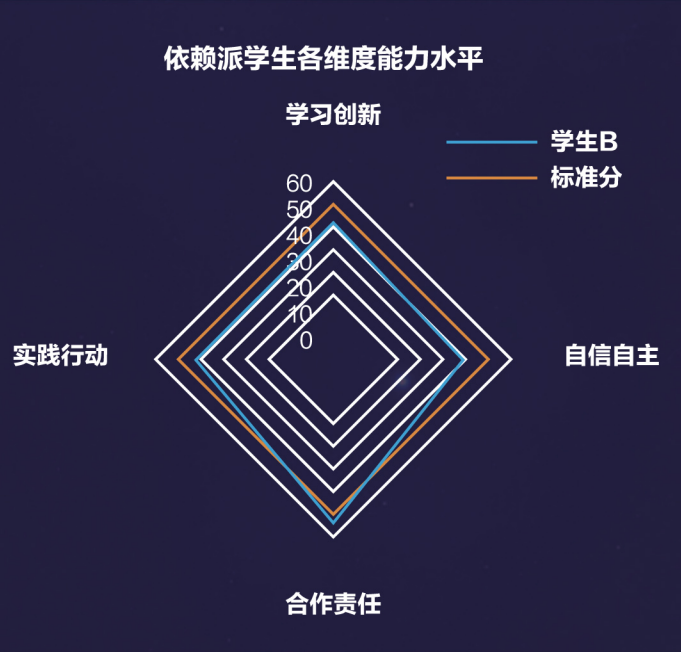
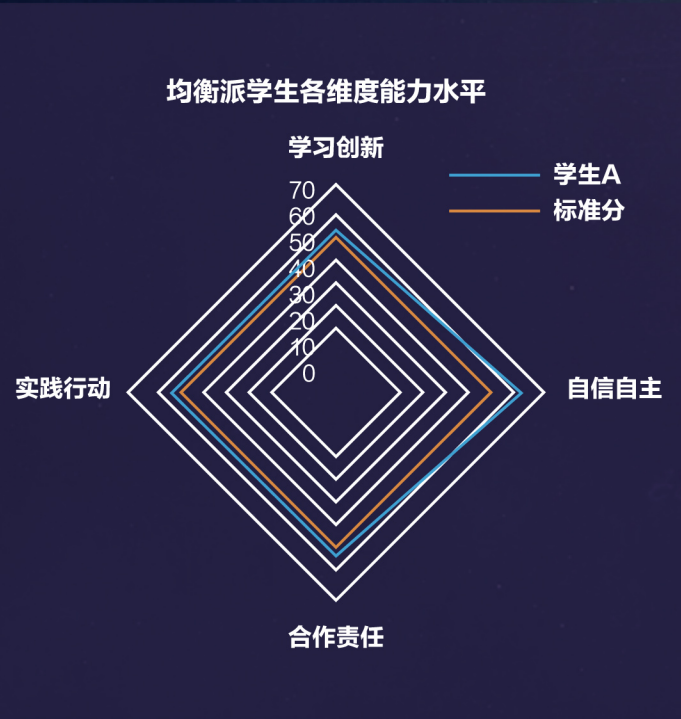
注：*在 0.05 水平（双侧）上显著相关。

由此可见，学习基础素养尽管对于学科发展有很大影响，但并不能很好地预测学生的综合能力。因此对于学校来说，进行综合能力的测试是非常有必要的，综合能力的测试有助于我们得出学科素养和学习基础素养测试无法得出的结果（比如阅读能力、合作能力、创新能力等等），可以与学科素养测试和学习基础素养测试结合起来，帮助我们更为全面地评价一个学生的发展水平。

【玩出“性格”】

游戏化考试通过相对分数比较学生在年级中的位置：以标准分 50 分基线，高于 50 分的维度处于较为优秀的位置，低于 50 分的维度是特别需要关注的维度。因此维度分数不单单只是一个数值，教师和学生都可以根据分值所显示的相对位置了解自身的能力趋势。

学生的优劣势各不相同，各有所长，根据最终综合能力的评估结果发现学生的学习表现往往可以分为以下几个类型：



“隐形学霸”——均衡派

代表门派：武当

乾坤八卦，五行平衡。武当武功刚柔并济，攻守兼备，温和缺不失力量。均衡派学生讲究“大音希声，大象无形”，能力维度相对发展均衡，没有特别突出的维度但在整体的位置却处于中上的水平，不可小觑。

“团队合作的积极拥护者”——依赖派

代表门派：丐帮

丐帮弟子遍布天下，虽个体武功不济、穷困潦倒但善于传递消息，合作无间。可见丐帮人多势众，通过团队的力量展现了武林第一大帮派的实力。因此，依赖派学生以合作能力见长，其他维度都较弱，相对缺乏自信自主。

“平易近人的学霸典范”——合作派

代表门派：少林

少林作为武学门派的泰山北斗，不仅具有较高的声望，同时又保有许多武功秘籍立于嵩山之颠。因此，以少林为代表，合作派学生具有较高的合作能力和实践适应性，学习创新能力、自信自主处于中上水平，相对优势较为显性。

“默默耕耘的平凡人”——沉默派

代表门派：衡山派、青城派

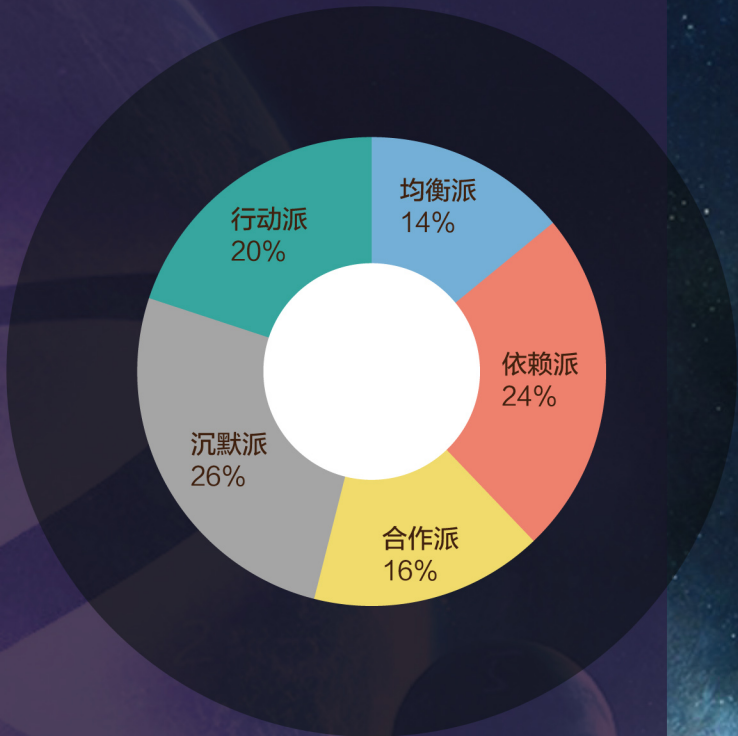
武侠小说中我们很容易发现一些背景门派，武功不高，门派声望没有大门派响亮。然而恰恰这些门派作为平凡的大多数所存在着。而我们的大部分学生，同样作为平凡人，各个维度能力水平处于中等偏上水平，但相对自信自立能力较弱，在群体中并不突出。

“一切行动高于嘴炮”——行动派

代表门派：明教

明教被误以为是“魔教”，但大部分明教教众不喜辩解，而以武功论长，显示出相当的“行动力”。行动派的学生虽然在各维度能力水平相对处于中下位置，但和明教一样，实践适应能力相对突出，是对自身而言是真正意义上的“行动派”。

学生类型比例



将“性格各异”的学生人数进行统计,发现“沉默派”与“依赖派”是学生中所占比例较高的,多达 26% 及 24%,也许“沉默是大多数”并非一句戏言。与此同时,我们也不禁提出这样的疑问:团队合作对个体的影响是否如同我们想象中的这般好?也许不久的将来在更多的研究探索中,能够获得答案。

情境游戏是游戏化考试的载体,我们通过这些有趣、客观的考试收获的是对学生的尊重、对学生发展性成长的关注。学生评估游戏化的过程,实质上是评价观由教师转移到学生身上的一个过程。当成绩不再是成绩,将“游戏化”赋以科学的属性,那么快乐地“玩”也不再是小孩子“不务正业”的表现,而是成为一种标准化的教育评价手段,根据评价结果来帮助学生,提高他们的综合能力,正确指引孩子们将来更好地立足于社会的大环境。