

多邻国英语测试：安全、 监督和调整



安全、自动处理和住宿，2021年10月19日
(9页)

englishtest.duolingo.com/resources

内容

1	介绍	2
2	登机时	2
3	测试管理	4
3.1	禁止的行动.....	4
3.2	项目安全.....	4
4	测试后管理	5
4.1	人工智能辅助处理.....	5
4.2	专家人工处理.....	5
5	非认证原因	6
5.1	通过认证决策重新获取策略.....	7
6	数据保留	8
7	调整	8
	参考文献	9

通讯作者：

多邻国英语测试多邻国，股份
有限公司。5900年，美国宾夕
法尼亚州大街匹兹堡，宾夕法
尼亚州15206

电子邮件：support-englishtest@duolingo.com

1 介绍

多态英语测试是对一般英语语言能力的高风险评估。这是一种计算机自适应测试（CAT），可以在线和按需交付。从入职到测试上传，大约需要一个小时才能完成。本文档概述了测试在其管理期间和之后的安全性和监督特性和过程。

2 登机时

测试过程从注册一个多邻国英语测试帐户开始。一旦考生注册，他们必须下载多形英语测试桌面应用程序。这是一个自定义应用程序，允许考生在两个支持的平台上进行测试，即Windows和Mac。一旦考生下载并安装了这个应用程序，测试就可以开始了。考生使用他们的登录凭证登录到他们的多组态英语测试帐户，此时他们可以切换到使用该应用程序。

开机过程从检查是否有多个显示器被插入计算机开始。如果检测到多个显示设备，则会提示考生断开除主显示设备外的所有设备，以便继续进行测试。一旦通过此检查，入职将继续自动检查所需的输入和输出设备（即麦克风和网络摄像头）是否存在和可用，以及应用程序是否有访问它们的权限。

作为系统检查的一部分，考生还被要求提供任何可能来自测试过程的音频和视频记录的生物识别信号的同意。世界各地的法律都要求这种同意，考生不提供就不能进行测试。

接下来，考生将提供他们的身份证明和人口统计信息。可接受的身份证明形式包括护照、驾照和符合以下身份证质量标准的国家身份证：

1. 身份证必须由政府签发，并包括照片、姓名、生日、到期日期、持票人的签发日期和发证机构。
2. ID必须没有可见的篡改或损坏。
3. ID必须显示该类型的ID所需的安全特性。例如：全息图、凸起的密封条、条形码或可见芯片。
4. ID必须是原始文档。我们不接受影印本或任何其他形式的复制品。

除了上述可接受的身份证明形式外，联合国难民文件也被接受，尽管没有达到所有的身份证明质量标准。这些文件经过一个特殊的审查过程，可以通过联系多邻国英语测试支持来提交。

5. ID必须在测试管理的日期上有效。过期的id。
6. 身份证必须有考生当前的清晰照片。带有无法识别的照片的身份证件将不被接受。
7. 照片必须直接印在身份证上，不得通过分层或粘合剂嵌入或附着到身份证上。
8. ID上的名称必须与考生的姓名相匹配。使用缩短的名称或昵称可能会导致测试结果的延迟。分数不能以除身份证上列出的名称以外的任何名称发送给机构。

然后，测试会采集考生的照片和他们的身份证，以进行身份验证。在ID收集之后，测试者将以他们所选择的用户界面（UI）语言显示测试规则。相关规则如下：

1. 确保您始终完全可见。
2. 不要离开相机预览。
3. 不要把目光远离屏幕。
4. 保持你的耳朵可见和暴露。
5. 不要离开你的网页浏览器。如果您的光标离开窗口或您退出全屏模式，您的测试将不会被认证。
6. 除非有指示，否则不要说话。
7. 在任何时候都不要与他人交流。
8. 不要让别人和你一起在房间里。
9. 不要使用任何外部的参考材料。
10. 不要使用手机或其他设备。
11. 不要写笔记或读笔记。
12. 始终保持您的麦克风和摄像头的启用状态。

以下是，他们被要求同意以下宣誓：

- 我已阅读并同意多国式英语考试的服务条款。
- 我是为本次测试提供的身份证和个人信息的人。
- 我所提供的信息是完整和准确的。
- 我理解这个测试的规则，并且我同意遵守它们。我明白，监主管有权在他们的合理判断中对我是否遵守规则作出决定，而我的决定是最终决定。
- 我明白，不遵守规则将会阻止我的结果被认证。根据具体情况，我可能会被允许或不允许重新测试，如果出现某些违反情况，我可能会被永久禁止进入测试平台。

- 我明白，我被禁止以任何方式分享或重新发布任何测试材料，并且这样做违反了服务条款，并可能导致我的测试结果被取消认证，并且我被禁止进入测试平台。

3 测试管理

3.1 禁止的行动

多邻国英语测试采用了一种“威慑和探测”的方法来确保安全。为了防止作弊，测试会了可能提供不公平优势的键盘操作（例如截图）。它还限制了可疑的动作（例如，将鼠标移出屏幕或切换窗口），如果这些动作发生的次数过多或持续的时间过长，则结束测试会话。如果测试因为这些行为而结束，考生在必须购买新的测试学分之前有三次免费的重新录取尝试，并且他们可以在30天的滚动窗口中购买不超过两个测试学分。如果考生由于非恶意原因（如技术错误）而无法考或测试学分，他们可以联系客服支持以获得额外的测试学分。

3.2 项目安全

多态英语测试通过一个大的物品池来实现“被动”的物品安全。这个池包含了成千上万的项目。多组态英语测试团队使用公开可用的数据源来训练机器学习（ML）模型，并使用这些模型来生成项目。然后，人类审查这些项目以确保公平和偏见（Settles等人，2020年）。

系统将定期刷新项目池。多邻国英语测试团队还不断监测项目曝光率和测试重叠率。项目的暴露是该项目出现的测试部分。测试重叠是在任意两个测试管理机构之间共享的项目的平均数量（Way, 1998）。因为多异态英语考试是用于大学录取的，所以它的风险很高。由于它是远程管理的，考生非法获得物品的可能性也很高。因此，暴露率和重叠率以及项目特征被持续监测，当项目超过期望的阈值时就被淘汰。

由于测试是完全在线的，攻击者可能会试图通过访问项目数据库或通过协调的测试获取来破坏项目池。多邻国英语测试团队通过维护一个安全的API、赞助定期的安全审查和规范对项目数据库的访问来防止API/数据库攻击。服务器和客户端（浏览器）之间的通信是加密的，因此第三方不可能尝试“嗅探攻击”。此外，所有项的分级都是在服务器端完成的。考生从来无法获得正确的答案或项目成绩。多邻国英语测试团队通过监控异常测试模式、扫描互联网上的物品存储库和提供举报人报告来检测协调的测试进行。

导致共享测试项目的项目违规不太可能对测试的安全构成很大的风险；由于物品库的规模很大，任何看到违规物品的考生都不太可能在测试中收到相同的物品。但是，如果违规影响了考试成绩，这些项目将从付费测试中退出到练习测试中。

4 测试后管理

监监在测试管理后的48小时内进行，构成了审查时间的大部分。由于学习器的时间和注意力是有限的资源，该测试用自动化工具补充他们的专家判断，以提高效率。在测试管理之后，一个人工智能（AI）算法立即进行了一轮自动监督。然后，人类监考人员考虑这个人工智能监考的结果，以确定规则是否被打破。

4.1 人工智能辅助处理

第一轮监考工作由人工智能算法进行。这些算法的目的是评估测试性能的特征，以识别违反规则和恶意行为的潜在指标。这个过程利用计算机视觉和生物识别过程来评估考生的行为，自然语言处理算法来检查对生产问题的反应，以及统计程序来分析反应模式。这些算法“升起旗帜”，并附上他们对这些旗帜的信心的指标；然后由人工检查员进行审查，以做出最终的验证决定。

4.2 专家人工处理

多邻国英语考试是由一个盲、匿名、多轮的审查过程。每一轮的监考都由精通英语的人进行；接下来的几轮由认证英语作为第二语言（ESL）教育者进行。在每一轮监考中，测试被随机分配给监考员，而监考员缺乏寻找特定考生或考生组的能力（例如，来自特定地区的考生、测试完成时间或与特定机构相关的测试）。此外，进程器彼此之间是匿名的，从而消除了进程器在前几轮中更信任某些进程器的可能性。如果在任何时候不同一轮的监考不同意，一个特殊的高级监考小组做出最终决定。

使用提供的人工智能工具，检查器检查测试的每个部分中超过80种不同的行为，这些行为被分组到几个不同的严重程度类别中。处理器所寻找的第一类行为是ID验证问题。检查姓名、生日、过期日期、ID类型和ID照片，并注意任何不一致的地方。

在ID验证之后，检查人员将全面检查测试是否存在轻微、重大或恶意违反测试规则。一个轻微违规的例子是背景噪声或设备警报噪声的存在，这在本质上可能是无意的。这些轻微的违规行为将在随后的几轮调查中进行审查，以确定它们是否曾经获得了不公平的优势。主要违规行为的例子包括考生戴耳机或看太久。重大的违规行为会导致“违反规则”的认证决定。最严重的违规行为，即恶意违规行为，包括考生接受外部帮助，会导致“恶意行为”裁决和禁止多态英语考试。在每一种情况下，考生都会得到缺乏认证结果的原因，这样他们就可以在不再次违反规则的情况下获得关于如何重新测试的明确和可操作的反馈。

5 非认证原因

在监督过程结束时，一个测试要么被认证或未被认证。在前一种情况下，考生会收到他们的分数，他们的分数会被发送到考生选择的任何机构。在后一种情况下，该测试因以下原因之一而被拒绝：

- 技术错误
 - 网络摄像头、麦克风或扬声器出现了故障。
 - 出现了一个上传问题，导致了严重的视频或音频延迟。
 - 考生的网络连接有问题。
- ID问题
 - 标识被裁剪、模糊、模糊、难以辨认或完全丢失。
 - 身份证已经过期了或者是复印件。
 - 该ID不是在特定国家接受的ID。
- 规则被打破
 - 考生都戴着耳机。
 - 考生的耳朵被遮住了。
 - 这个房间太黑了。
 - 这个房间不是私人的。
 - 这个房间并不安静。
 - 考生反复或延长一段时间。
 - 还有其他人在和考生说话。
 - 考生在开放式演讲部分说话不够多（必须按照指示说话至少30秒）。
- 恶意行为
 - 考生不是身份证图片中的人，也不是身份证被篡改的人。

- 考生在过去的30天里已经通过另一个账户完成了两项有效的测试。
- 考生正在做假唱。
- 考生正在使用外部设备或工具。
- 考生要在纸上写字/用钢笔或铅笔写字。
- 考生正在接受另一个人的帮助。
- 考生将测试屏幕截图。

5.1 通过认证决策重新获取策略

- 总体政策：
 - 一个信用，一个信用=，一个测试结果。
 - 考生有三次尝试使用学分来完成一个完整的测试。
 - 例如：考生开始考试两次，互联网中断，考试失败；考试参加者还有一次尝试通过该学分的完整考试。
- 技术错误：
 - 考生可以不收取额外费用（如果信用仍然可用）；如果信用尝试过多（尝试启动>3次），支持可以手动添加信用。
- ID问题：
 - 必须重新提交一个有效的分数的ID；将不需要重新参加完整的测试
- 违反规则：
 - 考生可以不收取额外费用（如果信用仍然可用）；如果信用尝试过多（尝试启动>3次），支持可以手动添加信用。
 - 如果考生有两个考试违反了同一信用的规则，该信用自动到期；考生可以对这个结果提出上诉。
- 恶意行为：
 - 考生被阻止了；他们不被允许重新参加考试。
- 上诉程序：
 - 如果考生认为在认证过程中存在错误，他们可以通过上诉程序免费申请。这将导致测试通过监督团队进行额外一轮的审查。然而，经过一轮上诉后的决定是最终的，不能再被提出质疑。

6 数据保留

为了防止考试中作弊和违规，从考生那里收集个人信息（PII）。这包括：

- 护照和/或政府签发的身份证的照片
- 法定名称
- 出生日期
- IP地址和位置数据
- 考生参加考试时的屏幕和面孔的视频

多态英语测试完全符合欧盟的通用数据保护条例（GDPR）。在考生的要求下，多态英语测试团队将与他们共享或删除已经收集到的任何数据，除非共享/删除数据会危及测试的安全性（例如，考生屏幕上的视频，因为这样会泄露物品）。

在收集到的PII中，ID照片是最敏感的。所有ID照片在测试管理后只在加密存储中保存4天。在这四天之后，他们会被转移到一个更安全的地方，在那里只有一组选定的员工被允许访问他们。

信用卡信息作为支付过程的一部分被收集起来了，但没有一个被保留下来。它被直接转发给信用卡处理器，他们也通过了gdpr的认证。

7 调整

目前，多邻国英语测试提供了以下可访问性调整：允许耳机、屏幕放大镜、助听器/设备，以及完成测试所需的任何类型的人体工程学鼠标或键盘。这些都是根据要求而允许的。任何觉得需要调整的考生都可以联系客户支持团队，逐个处理他们的案件。

参考文献

- Settles, B., LaFlair, G. T., & M. (2020)。机器学习驱动的语言评估。《*计算语言学协会学报*, 8, 247-263。 https://doi.org/10.1162/tacl/_a/_00310
- 《方法》(1998)。保护经计算机化的测试项池的完整性。
教育测量: 问题与实践, 17(4), 17-27。