为了证明按照我们建立的得分模型来分类具有其合理性，我们直接使用尝试次数作为自变量，将根据得分模型划分的难度等级作为因变量来进行分类测试，并将测试结果与我们按照得分模型所划分的结果相对比，计算其准确率、召回率、精确率以及F1值。这样做具有其内在逻辑的合理性，因为得分与难度是可以互相佐证的，因此只要我们基于尝试次数可以推导出与我们所划分的难度等级基本吻合的结果，那么我们也就证明了得分模型的合理性。

（具体的结果......）