运用多层建模评估印度尼西亚非理科学生的科学素养与技术课程

Bayu Setiaji[a]、Purwoko Haryadi Santoso[b,c]*、Khafidh Nur Aziz[a]、Heri Retnawati[b]和 Moh Khairudin[b]

[a]日惹国立大学物理教育系，印度尼西亚斯勒曼55281，[b]日惹国立大学教育研究与评估研究生院，印度尼西亚斯勒曼55281；[c]西苏拉威西大学物理教育系，印度尼西亚马杰内91413

* 通讯联系：bayu.setiaji@uny.ac.id

如果期刊要求，请在此处提供所有撰稿人的简短生平简介。

使用多层建模评估印度尼西亚非理科学生的科学素养与技术课程

科学教育界包括印度尼西亚教育体系正在培养科学素养。科学素养与技术课程经过设计并实施，旨在强化赋能具备科学素养的印度尼西亚社会这一国家倡议。本文旨在评估非理科专业本科生（N = 160）在考虑学生所在院系和学院环境的嵌套结构后，能够在多大程度上完成该课程。因此，采用多层建模开展分析，以处理这一特性。第一层分析涉及学生表现和情感属性，分别使用实际表现科学素养评估（SLA-D）和动机信念（SLA-MB）进行测量。后续层级的分析包括从机构档案中收集的人口统计因素。研究结果表明，人口统计因素对学生科学素养表现的影响并不显著。学生所在院系和学院层级的不同环境，显著影响了情感因素与科学素养课程学习过程之间的关联。多层方法控制了学生数据结构特性下公平的学生评估。本文提出了诸多启示，为教育数据分析拓展思路，并具体考察高等院校非理科专业学生科学素养课程的有效性。

关键词：多层建模；科学素养；非理科学生 主题分类代码：如果期刊要求，请在此处填写

引言

全球科学教育研究界已开展大量研究，致力于提升科学素养（Impey 等人, 2011; Odden 等人, 2021; Queiruga-Dios 等人, 2020; Santoso 等人, 2022）。国际学生评估项目（PISA）和国际数学与科学趋势研究（TIMSS）发布的调查报告，无疑对世界各地的教育举措产生了重大影响，其中也包括印度尼西亚科学教育

体系（Cheema, 2017; Rachman 等人, 2021; Ustun等人, 2022）。这些调查项目所揭示的印度尼西亚学生表现不尽如人意，促使政府修订课程体系，以有力应对这一挑战。自2020年起，印度尼西亚教育、文化、研究与技术部（KEMDIKBUDRISTEK）试点并推广名为“MERDEKA BELAJAR”的课程创新，旨在将原有国家评估项目的重点转向科学素养教育（Nurjati 等人, 2022）。此后，印度尼西亚更新后的评估议程旨在为中学生做好准备，使其能够应对近期PISA和TIMSS调查所衡量的科学素养领域（Salamah 等人, 2022）。在不忽视其他学科所传授的学科知识之重要性的前提下，这项改革计划旨在推动学生表现优于以往由PISA和TIMSS等权威组织开展的国际科学调查结果所呈现的水平。同时，我们相信这项政策能够持续推进，并且必须邀请全体利益相关者和政策制定者共同参与，致力于建设印度尼西亚文明的未来，其中也包括本文所探讨的高等教育。

高等教育机构必须成为开拓者，积极参与教育与研究专长领域。日惹国立大学（UNY）是印度尼西亚教师教育机构（TEIs）之一，负责为全国中学生提供科学教育。近年来，UNY更加重视强化当前培养科学素养的国家愿景。与此同时，自2020年起，科学素养与技术课程（MKU 6217）已被列为必修课，纳入各系和学院本科一年级学生的教学大纲（BAKK UNY，2019年）。该课程主要面向非理科学生，旨在提升他们的能力，使其成为未来的印度尼西亚……

教育工作者。尽管他们将从事非科学领域的职业，但已具备科学本质的若干方面知识，其中包括科学探究的原则。从本质上说，他们完成本科教育后，预计将成为社会的新一代。同时，他们仍被寄望于在各自领域成为专业的职前教师，并具备促进印度尼西亚社会提升科学素养的能力。

如前所述，这门课程对于 UNY 等教师教育机构（TEIs）尤为重要，因为这些机构肩负着培养服务于印度尼西亚社区的未来教师的责任。UNY 自2020年起才开始实施这门课程。因此，现在正是评估面向非理科学生的科学素养项目实施情况的恰当时机。诚然，有效的科学素养教学要求在实施过程中重视评估（Hobson，2003）。在本研究中，我们收集了2022/2023学年四个不同非理科专业的学生数据，分别来自市场营销、会计学、爪哇语和舞蹈教育专业。由于专业各不相同，参与本研究的学生显然具有不同的背景、先验知识、家庭资源以及其他相关变量。因此，学生在整个课程中的表现必然呈现出多样性和复杂性。由此，要更好地理解该课程对学生科学素养的影响将面临一定挑战。在此背景下，本研究需要采用一种综合性方法，以评估非理科学生在多大程度上能够完成科学素养课程。

可以参考相关研究，了解研究人员如何设计和评估科学素养教学，尤其是在高等教育背景下，更具体地说，是针对本研究所关注的非理科专业学生。我们的研究独具特色，也颇具挑战性，因为我们主张调查面向非理科专业大学生的科学素养课程，而这些学生来自不同的

院校背景。针对大学生跨学科背景下的科学素养课程，研究者采用了大量评估方法（Efthimiou & Llewellyn, 2004; Hamper & Meisel, 2020; Hobson, 2008; Impey, 2013; Parkinson & Adendorff, 2004; Ross et al., 2013; Sjöström & Eilks, 2018; Surpless et al., 2018）。综上所述，我们可以简要总结：很少有研究者考虑到由学生数据本身特性所产生的嵌套效应。事实上，学生数据大多隶属于大学行政区域的层级结构。因此，学生必然处于聚类情境中，而这种情境可能对他们产生影响。低估这一特性会导致学生表现评估出现偏差，因此应谨慎避免。

广义而言，学生表现是教育学者通常会报告的一项教育过程代理指标。人们提出了许多分析方法，以便更好地测量、考察和评估学习过程中学生的表现（Ding，2019）。为处理上述层级数据问题，人们进而提出了源自回归分析的多层建模，以进一步考量学生学习的嵌套结构（Finch 等人，2016）。在本研究中，采用多层建模方法将有助于我们深入理解学生在科学素养课程中的表现，以及该表现与若干潜在因素之间的关联，例如情感属性和人口统计变量；学者们使用国际学生评估项目和国际数学与科学趋势研究的数据时，经常会报告这些因素（Ersan和Rodriguez，2020；Mohammadpour等人，2015；Ustun等人，2022；You等人，2021）。

采用多层建模进行数据分析时，必须考虑数据点所体现的数据聚类层级。在本研究中，第一层数据点是学生。他们能够产生教育数据及相关因素，这些因素可能对学生的学习产生影响。除了

认知方面外，学生在学习过程中还具有情感属性。这也构成了我们数据的第一层，因为这些数据由学生产生。随后，院系因素会构成学生数据的下一层级。我们可以根据机构信息系统记录的人口统计变量来理解这一点。学生由特定的系和学院管理。由学生所在院系和学院的嵌套结构产生的人口统计变量，可能与学生表现的潜在差异相关（Kanim & Cid, 2020年；Salehi等人, 2019年；Simmons & Heckler, 2020年）。因此，必须采用多层建模，使分析更加全面。

以往研究已广泛记录了学生表现与科学素养学习中的情感属性（Bellová 等人，2021；Fives 等（2014）；Fortus 等人，2022；Rudolph，2020）和人口统计变量（Ustun等人，2022；You等人，2021）之间的关联。这些研究指出，情感结果会显著影响学生的科学素养。然而，不同研究者采用了不同的情感测量构建方法。因此，这些研究结果对其他情境的启示仍需进一步考察。例如，Fortus（2022）将情感属性定义为四种构念，即兴趣、态度、自我效能或自我概念，以及动机。Fives 等人（2014）则提出了另一种观点，为其科学素养评估（SLA）开发了动机信念测量。

此外，正如 Ustun 等（2022）和 You 等（2021）在国际学生评估项目（PISA）数据背景下所报告的，人口统计变量可能会影响科学素养评估的方差。他们发现，经济/社会/文化地位（ESCS）等人口统计变量可能会对科学素养产生显著影响。然而，已有多位学者指出，人口统计效应的结果并不一致，而且

仍未有定论（Simmons & Heckler，2020年）。还必须考虑其他更重要的因素，以便更全面地开展调查。

为弥补以往研究中的不足，本研究旨在对 UNY 课程体系近期实施的科学素养与技术课程进行评估。采用多层数据建模技术，对学生表现、情感属性和人口统计变量的嵌套数据结构进行测量和分析，数据分为三个层级（学生、系、学院）。为指导本研究，提出以下两个研究问题。

研究问题1：学生在科学素养课程中的表现和情感属性的均值差异，在院系层级、学院层级和人口统计变量方面的变异程度如何?

RQ2. 系层级和学院层级在多大程度上会影响学生在科学素养课程中的表现、情感属性与人口统计变量之间的关联?

探究影响学生学习的因素，有助于学者、教育工作者和实践者设计公平的科学学习。本研究基于 UNY 2022/2023学年的数据，对近期实施的科学素养与技术课程进行了评估。分析依据多层建模结果，考察学生表现、情感属性和人口统计变量之间的关系，并按照大学生所在院系的性质及学院环境进行聚类。本文报告的证据有助于开启有关教育数据分析的讨论，尤其是如何处理必须审慎对待的复杂学生数据结构。

方法

本文旨在评估印度尼西亚非理科学生在科学素养课程中的表现，并考察其与情感属性和人口统计变量的关联；这些变量按层级从研究涉及的四个不同院系和两个不同学院收集。为此，以下各节将详细介绍研究方法。

背景

如前一节所简要介绍，本研究以日惹国立大学（UNY）在2022/2023学年第一学期开设的科学素养与技术课程（MKU 6217）为背景。该课程是非理科专业学生的必修课程。课程内容包括科学本质、科学方法、人类文明的发展、作为系统的宇宙、科学技术及其应用，最后以科学中的大观念为主题。科学素养课程的教学大纲是根据大学模式和上述国家运动制定的。

由于疫情仍处于持续过渡阶段，远程学习仍是学习方式之一。日惹国立大学（UNY）在疫情前开发的基于 Moodle 的学习管理系统（LMS）“BESMART”（Priyambodo，2016年；Surjono 等，2017）被用于向全体学生授课，并对课程进行管理和组织。图1展示了 BESMART 为科学素养课程提供的学习仪表板。

>>> 图 1 <<<

第一作者（B.S.）和第三作者（K.N.A.）被安排担任来自两个学院的四个非理科本科院系所开设课程的讲师。研究参与者以自愿方式确定，同时

不忽视 UNY 在读非理科学生的代表性。参与者招募自市场营销系（n = 38）、会计学（n = 39）、爪哇语（n = 51）和舞蹈教育（n = 50）。市场营销系和会计教育专业课程隶属于“经济学院”，其余两个专业课程则由“语言与艺术学院”管理。总计有 178 名学生参与了本研究。被记录为缺少某些数据点的学生，其部分或某个变量的数据安排在第二次调查阶段收集。尽管如此，由于有 18 名学生未意识到我们提供了再次参与的机会，因此在后续数据分析中应将其删除。

测量

学生数据的第一层级通过课堂评估收集，包括学生在科学素养课程和期中考试中的表现，以及他们的情感属性。学生表现将作为因变量。在同一层级，同时测量期中考试成绩和情感属性，它们将作为可能影响学生表现的预测变量。随后，各系和学院往往具有相似特征，这些特征对应于学生数据中出现的人口统计变量。本研究的人口统计数据汇编自机构档案。系是数据的第二层级，学院是第三层级。

本研究采用 Fives 等（2014）发布的一项科学素养评估（SLA）来考察学生表现。SLA 涵盖“实际表现”科学素养评估（SLA-D）和“动机信念”科学素养评估（SLA-MB）。研究使用 SLA-D 中的 26 道选择题，作为学生在本课程中表现的代理指标。SLA-MB 则用于调查学生的情感属性。

首先，需要将 SLA 的英文版本翻译并调整为符合印度尼西亚社区语境的版本。两位在科学素养教育方面拥有十年以上教学和研究经验的专家对内容效度进行了评估。他们是本文的第四作者和第五作者。

期中考试还设置了五道开放式题目。我们从科学素养的多个方面对学生进行了考查，同时也借鉴了 Fives 等人（2014）所阐述的构念，但采用了其他形式的开放式问题，而非上述 SLA-D 所测量的二分式回答。一些证据表明，与封闭式回答相比，采用开放式格式测量科学素养可能更为合适（Miller，1998）。此外，本课程的讲师（本文的第一作者和第三作者）制定了期中考试的评分量规。与 SLA 印度尼西亚语版本此前的验证步骤类似，我们也邀请相关专家对其进行了表面效度检验。

理论上，SLA-MB 所呈现的三个构念构成了科学素养情感属性的测量基础，分别是科学价值（VOS）、我能在科学领域做什么（DIS）以及我对科学的看法（BAS）（Fives 等人，2014）。这三个因素下共设置了25个观测项目，采用五点李克特量表。利用学生的回答数据（N = 160），我们发现其信度表现良好，与 Fives 等人的研究结果一致（$\alpha_{all} = 0.781$，$\alpha_{VOS} = 0.868$，$\alpha_{DIS} = 0.786$，$\alpha_{BIS} = 0.903$）。此外，我们还采用探索性因子分析（EFA），验证 Fives 等人提出的构念与学生认知之间的一致性。我们发现了三个定义明确且彼此独特的成分。SLA-MB 的特征值均大于1。因此，这与先前的研究结果一致：SLA-MB 的三个构念可以彼此独立地使用（Fives 等人，2014）。

人口统计变量取自学校注册记录。本研究收集了十项人口统计变量，并在后续分析中加以考察，分别为性别（1 = 男性，2 = 女性）、入学途径（1 = “SNMPTN”，即全国作品集入学制度；2 = “SBMPTN”，即全国笔试入学制度；3 = “SM Prestasi”，即校内作品集入学制度；4 = “SM Utul”，即校内笔试入学制度）、学费资助（1 = 有补贴的，2 = 无补贴的）、奖学金获得者（1 = 是，2 = 否）、高中背景（1 = 科学类，2 = 非科学类）、居住地（1 = 北苏门答腊，2 = 西苏门答腊，3 = 廖内，4 = 占碑，5 =明古鲁，6 = 邦加-勿里洞，7 = 雅加达，8 = 西爪哇，9 = 中爪哇，10 =日惹，11 = 东爪哇，12 = 东加里曼丹）、父母受教育程度（1 =硕士，2 = 学士，3 = 文凭，4 = 高中，5 = 初中，6 =小学，7 = 未受教育）以及父母月收入（1 = 超过4,000,000印尼盾， 2 = 3,000,000–4,000,000印尼盾，3 = 2,000,000–3,000,000印尼盾， 4 =1,000,000–2,000,000印尼盾， 5 = 低于1,000,000印尼盾）。表1概述了参与者在各班级各属性类别中的分布情况。

>>> 表 1 <<<

我们发现，大多数学生为女性（76%），通过校内招生凭作品集或“SM Prestasi”录取（48%），未获得经费补贴（74%），未获奖学金（84%），来自非理科高中背景（67%），来自中爪哇（39%），与高中毕业的父亲（44%）和母亲（43%）同住；父亲和母亲的月收入分别在100万至200万印尼盾之间和低于100万印尼盾。

总之，我们从课堂评估中收集了三个变量，并从大学教务处记录中提取了十个变量。对于……的要求

后续说明中，表2总结了本研究所用各变量的描述，包括每个对应变量采用的代码（用于下文的方程说明）、测量工具和数据类型。

>>> 表 2 <<<

分析

本研究采用 Fives 等人（2014年）提出的正确选项评分量规，对通过 SLA-D 测量的学生表现进行评分。在 26 道选择题中，根据学生的作答情况进行评分，并将得分转换为百分制。随后，在汇总通过 SLA-MB（Aff）测量的情感属性之前，由于 BAS 因子中包含负向题项，因此需要对学生在这些题项上的作答进行反向计分。期中考试（MidTerm）也采用与学生表现相同的评分标准进行评分。

在进行多层建模之前，研究问题1（RQ1）采用方差分析（ANOVA），以检验各系、学院及人口统计因素下不同班级之间学生表现、期中考试成绩和情感属性的均值差异。通过方差分析，验证系和学院层级是否会影响学生表现的方差。此外，方差分析还用于确定应纳入多层模型方程的潜在因素（RQ2）。只有在各因素下不同班级之间存在显著差异（$\alpha < 0.05$）的因素，才会被纳入多层模型方程，以分析其对作为目标变量的科学素养方面学生表现的影响。

在 RQ2 中，采用多层建模方法处理学生所在院系和学院层级的嵌套数据结构（Finch 等人，2016年）。多层建模技术适用于处理分布在四个不同院系且隶属于两个不同学院的学生数据。首先对两层模型进行分析，随后进一步开展纳入学院层级的三层建模。图2

展示了一个包含三个层级的数据结构，其中 i、j 和 k 分别代表学生、系和学院层级。

>>> 图 2 <<<

分析的第一步是构建空模型或基线模型，其中方程中不包含任何预测变量。空模型被用作模型构建和比较的基线。空模型的构建如下。

$$\text{SPerf}_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij} \quad \text{(Two-level)} \quad (1)$$

$$\text{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \varepsilon_{ijk} \quad \text{(Three-level)} \quad (2)$$

作为参考，我们借鉴了 Finch 等人（2016）提出的多层建模数学符号。方程（1）依次对应两层建模，方程（2）则对应三层建模。SPerf_{ij} 指第 j 个系中的第 i 个学生的学生表现，γ_{00} 指第 j 个系的总体截距均值，U_{0j} 表示第 j 个系的随机效应，ε_{ij} 项表示模型无法解释的学生层面随机误差。方程（2）与前者略有相似之处。然而，该方程是为更高层级构建的，因此我们发现其中使用了 k 下标，以表示学生所属的学院层级。如上文所述，δ_{000} 是嵌套在第 k 个学院中的第 j 个系的总体截距均值。相应地，V_{00k} 表示第 k 个学院的总体截距均值。最后，U_{0jk} 指随机效应，而ε_{ijk} 则是模型无法解释的随机误差。

之后，根据研究问题1中采用方差分析得出的初步发现，并结合空模型的组内相关系数结果，构建了后续模型。多层建模使用 R 编程环境中 lme4 软件包提供的 lmer 函数进行拟合（Bates 等人，2015年）。由于限制性最大似然（REML）在估计方差参数方面已被证明比最大似然估计（MLE）更准确，因此选择了这种估计方法

（Luo 等人，2021）。总体而言，随后使用赤池信息准则（AIC）（Matuschek 等人，2017）对所构建的模型进行比较。根据经验法则，我们应将结果中赤池信息准则最低的模型视为最佳模型。

结果

院系层级、学院层级与人口统计变量之间学生表现与情感属性的均值差异（研究问题1）

研究问题1的初步信息证实了基于统计分析所发现的特征之间可能存在关联。方差分析将用于检验不同聚类或组别之间特定变量的均值差异。在本研究中，聚类必须是系、学院和人口统计变量，具体如表2所述。随后，组内相关系数（ICC）用于衡量分类变量（如人口统计因素）与特定变量（例如学生表现、期中考试和情感属性）之间的相关程度。因此，方差分析的显著结果以及预测变量合理的ICC值，将作为决定是否将这些变量纳入多层模型的考量依据。研究问题1的结果列于下方的表3中。

>>> 表 3 <<<

首先，我们发现学生的科学素养和期中考试成绩在院系层级和学院层级之间存在显著的均值差异（$p < 0.05$）。此外，院系层级和学院层级与学生的科学素养之间可能存在相关性。期中考试成绩也与院系层级相关，但我们未发现其与学院层级相关。这些结果表明，学生在科学素养课程中的表现可能会受到院系和学院之间差异的实际影响。因此，方程（3）和（4）将期中考试（MidTerm）纳入模型是有依据的。

$$\text{SPerf}_{ij} = \gamma_{00} + U_{0j} + \gamma_{10}\text{MidTerm}_{ij} + \varepsilon_{ij} \quad \text{（两层）（3）}$$

$\text{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \delta_{100}\text{MidTerm}_{ijk} + \varepsilon_{ijk}$ （三层）模型3和模型 (4) 4加入了γ_{10} 和δ_{100} 项。它们列出了预测变量与结果变量之间的回归系数。

其次，我们发现不同系和学院之间学生的情意属性不存在显著的均值差异。相反，Fives 等（2014）认为，情感测量与学生的科学素养之间存在较强的相关性。然而，可以参考图2来支持 Fives 等（2014）的观点，该图展示了 SPerf、Aff 和 MidTerm 皮尔逊相关值的散点图矩阵。可以看出，SPerf 与 Aff 之间存在相关性。因此，在理解情感属性与学生科学素养之间的关联时，不应忽视情感属性。基于此，第三个模型将情感属性纳入其中，如下文方程（5）和（6）所示。

$$\text{SPerf}_{ij} = \gamma_{00} + U_{0j} + \gamma_{10}\text{MidTerm}_{ij} + \gamma_{20}\text{Aff}_{ij} + \varepsilon_{ij} \quad \text{（两层）（5）}$$

$\text{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \delta_{100}\text{期中考试}_{ijk} + \delta_{200}\text{Aff}_{ijk} + \varepsilon_{ijk}$ (三层) (6) 显然，γ_{20} 和δ_{200} 解释了情感属性对学生科学素养的系数。

>>> 图 2 <<<

第三，在学生高中背景和社会地位（以母亲受教育程度衡量）的班级之间，情感测量存在显著的均值差异（$p < 0.05$）。令人意外的是，其他人口统计变量，如性别、入学途径、经费资助、奖学金获得者身份、居住地、父亲受教育程度和父母收入，均未产生显著影响。因此，应剔除这些不显著的人口统计因素。下一个模型应纳入这些结果，并按照方程（7）所示的二层模型和方程（8）所示的三层模型，加入两个显著的人口统计变量（HS 和 MEdu），具体如下。

$$\text{SPerf}_{ij} = \gamma_{00} + U_{0j} + \gamma_{10}\text{期中考试}_{ij} + \gamma_{20}\text{Aff}_{ij} +$$

$$\gamma_{30}\text{HS}_{ij} + \gamma_{40}\text{MEdu}_{ij} + \varepsilon_{ij} \quad \text{(两层) (7)}$$

$$\text{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \delta_{100}\text{期中考试}_{ijk} + \delta_{200}\text{Aff}_{ijk} +$$

$$\delta_{300}\text{HS}_{ij} + \delta_{400}\text{MEdu}_{ij} + \varepsilon_{ijk} \text{ (三层) (8)}$$

其中，γ_{30} 和δ_{300} 表示不同学生的高中背景对其科学素养课程学生表现的影响。相应地，γ_{40} 和δ_{400} 表示学生的科学素养课程表现与母亲受教育程度之间的依赖关系。

系和学院层级对学生表现、情感属性和人口统计因素之间依赖关系的影响（RQ2）本节将展示拟合这些方程后得到的多层建模结果。结果见表4。为了便于读者理解描述内容，我们首先介绍两层建模结果（模型1、3、5和7），随后介绍三层建模结果（模型2、4、6和8）。>>> 表 4 <<<

两层模型的结果

表4第二列中的模型1旨在考察院系层级与学生科学素养表现之间的关联。在固定效应方面，我们发现截距显著（γ_{00}，$p < 0.05$），表4中括号内的数值为相应的标准误。截距显著表明，院系层级可能会影响学生的科学素养表现，而这种影响可能与随机效应相关。表4随机效应部分报告的数值为方差分量，括号内为相应的标准差。直观而言，方差和标准差

结果可解释为截距（U_{0j}）随院系层级变化的程度。随后，表4报告的残差成分即为方程（1）中的ε_{ij} 项。

相应地，模型3将期中考试变量加入了方程。我们还发现，结果具有显著性，表明期中考试可能会影响学生的科学素养。在模型3中，我们还发现截距显著（γ_{00}，$p < 0.05$），且期中考试系数显著（γ_{10}，$p < 0.05$）。与空模型相比，模型3的截距标准误更大。相反，我们发现截距（U_{0j}）和残差（ε_{ij}）的随机效应模式在方差和标准成分上都有所减弱。这表明，多层模型能够更精确地估计我们的参数的标准误。

在模型5中，我们将情感属性纳入模型。然而，截距并不显著（γ_{00}，$p > 0.05$），但其标准误大于前两个模型。期中考试的系数（γ_{10}）仍然显著（$p < 0.05$），结果相似，且与模型3相比略有下降。研究发现，情感属性（γ_{20}，$p < 0.05$）对学生的科学素养具有显著影响，这与 Fives 等（2014）的报告一致。与前一个模型相同，也发现了截距（U_{0j}）和残差（ε_{ij}）的递减趋势。

关于模型7，在学生的高中背景（γ_{30}，$p > 0.05$）和母亲受教育程度变量（γ_{40}，$p > 0.05$）中，均发现系数不显著。因此，无法得出这些变量是预测学生科学素养的影响因素这一结论。这一结果与上述引言中提及的先前文献（Ersan和Rodriguez，2020年；Mohammadpour等人，2015年；Ustun等人，2022年；You等人，2021）的结论存在明显差异。正如情感影响的模型5所报告的那样，固定效应的截距（γ_{00}，$p > 0.05$）也不显著。

最后，总体模型已在表4的最后一部分报告。这些结果有助于描述和比较两层模型。报告自由度、组数（N）和观测值，以描述这些模型。在四个模型中，模型5的AIC最低。因此，模型5是最适合本研究所分析数据的两层模型。

三层模型的结果

与两层建模相同，表4第三列也列出了三层模型的空模型。结果还发现截距显著（δ_{00}，$p < 0.05$），且其标准误大于上述模型1。因此，将学院层级纳入模型，能够更准确地理解学生在科学素养方面的表现。随后发现，系层级的截距方差和标准差（U_{0jk}）低于模型1，但残差值相同（ε_{ijk}）。这一剩余值可归因于学院层级所提供的新增信息（V_{00k}）。

期中考试是预测学生科学素养课程的显著因素（δ_{100}，$p < 0.05$）。这与上述模型3报告的结果一致。显然，显著截距的标准误也更大（δ_{000}，$p < 0.05$）。对于随机效应，我们还发现院系层级的截距（U_{0jk}）和模型残差（ε_{ijk}）呈现下降趋势。

在模型6中，将情感属性纳入模型。与模型5相关的是，研究发现截距不显著（$\delta_{000}, p > 0.05$）。与此同时，其标准误仍然大于模型4。期中考试系数显著（$\delta_{100}, p < 0.05$）。随后，研究发现情感属性对学生的科学素养具有显著影响（$\delta_{200}, p < 0.05$），这与两层建模结果中模型5的报告一致。

关于模型8，高中变量（δ_{300}，$p > 0.05$）和母亲受教育程度变量（δ_{400}，$p > 0.05$）的系数均未达到显著水平。因此，与模型7报告的两层建模结果相同，这些变量无法作为影响因素来解释和预测学生的科学素养。最终，在四个三层建模结果中，AIC最低的是模型6，该模型仅将学生层级变量（期中考试和情感属性）纳入模型，与上述模型5的做法相同。因此，基于本研究分析的数据，模型6是拟合效果最佳的三层模型。

讨论

本研究旨在回答两个研究问题。在研究问题1中，我们探讨学生的科学素养均值差异及相关因素在多大程度上会因不同的院系层级、学院层级和人口统计变量而有所不同。在 RQ2 中，我们研究研究问题1中发现的院系效应和学院效应在多大程度上有助于探究情感属性与人口统计变量同学生在科学素养课程中的表现之间的关联。

我们可以为研究问题1的答案归纳出四项主要发现。首先，学生所在院系和学院层级显然是影响因素，会造成其科学素养的显著差异。这与引言中提及的既有研究结果一致（Kanim & Cid，2020年；Salehi等人，2019年；Simmons & Heckler，2020年）。学生在学习过程中所处的教育环境会塑造其学习氛围。学生学习过程的不同情境涉及许多因素，其中可能包括教师素质、班级规模、同伴动机、硬件设施，以及其他难以识别的因素。因此，对教育工作者而言，采用层级数据控制学生表现的测量，仍值得坚持。

其次，学习过程中的形成性评价可能是评估学生科学素养课程表现峰值的最佳预测变量。这与 Hobson（2003）所强调的内容相关。他指出，有效的科学素养教学要求在实施过程中关注评价。包括期中考试在内的形成性评价，是在学习过程中把握学生学习积极性的一种评价方式。在本研究中，保持学习意愿直至最后阶段，可以预测学生学习科学素养的成效。对于许多地方的大多数教育工作者而言，这一发现或许并不令人意外，因为它可能属于常见的教学知识。

除了直接补充形成性评价的影响外，本研究还证实了情感因素对学生科学素养的影响具有显著性。这与引言中提及的先前研究成果一致（Bellová 等人，2021；Fives 等人，2014；Fortus 等人，2022；Rudolph，2020）。对于科学素养教育工作者而言，这一发现可以作为一项建议：在评价学生表现时，充分考虑学生的态度方面，对于实现更加公平的评价至关重要。将课程评价与学生相关的情感因素相结合，是科学教育中真实性评价的基本原则（Nurjati 等人，2022；Salamah 等人，2022）。有效的科学素养课程应更加注重识别这一因素，并将其纳入班级评价标准。

在本研究收集的十个人口统计变量中，我们只发现两个变量会导致学生的科学素养出现均值差异。令人意外的是，在大多数研究中被预测会对科学学习产生重大影响的性别（Cheema，2017；Ustun等人，2022；You等人，2021），在本研究中并未对学生科学素养的方差产生显著影响。与此同时，高中教育准备和母亲受教育程度已被证实会影响学生科学素养的达成情况

。因此，有必要开展进一步研究，以便更全面地阐述这一结果，从而使其能够推广至更广泛的人群。

此外，我们可以结合多层建模结果的背景，讨论为回答研究问题2（RQ2）而阐述的两个主要发现。首先，我们可以总结上述双层建模所确定的最佳多层模型及其估计结果所呈现的模式。双层建模和三层建模的结果均表明，在本研究中，加入期中考试和情感变量是评估学生科学素养的最佳模型。与认知方面相邻，情感属性是真实性评价中的常见做法，这也得到了科学素养教育相关文献的推荐（Nurjati 等，2022；Salamah 等，2022）。这一点与上述研究问题1（研究问题1）的回答相关。另一方面，在模型5中，我们发现由系和学院环境驱动的随机效应截距均不显著。这可以解释为学生评价中公平评价标准的控制因素。然而，将情感测量纳入评价维度的考量时，应谨慎构建相关测量，因为其潜在因素难以测量。

其次，我们发现，在采用多层次分析对模型进行控制后，截距的标准误呈现出减小的趋势。这可以理解为，来自系和学院层级的方差来源确实会影响因变量，因此在进行更完善的分析时，我们应当将其纳入考虑。具体而言，正如先前研究所建议的（Ersan和Rodriguez，2020年；Mohammadpour等人，2015年），采用基于层级的分析方法显然能够更准确地捕捉学生的科学素养。学习过程中的层级差异是大学行政区域所产生的结果。直观而言，这一结果可以为政策制定者提供建议，促使其确保全体学生享有公平的教育机会。不可否认，平衡各学校之间的教育质量能够让学生感受到更多的认可。

尽管如此，我们认为，本研究可能受到三个因素所导致的若干不确定性来源的影响。首先，我们的数据收集方法可能引入错误和不确定性。本研究旨在评估 UNY 设计的科学素养课程的实施情况。因此，本文报告的研究结果可能与印度尼西亚教育体系中其他大学的情况有所不同。其次，情感测量是一种潜在构念。许多研究提出了不同的观点，以界定科学素养方面的情感测量。本研究的重点并非旨在描述非理科专业学生与理科专业学生之间的差异。这可能是科学领域相关论文所报告的偏差来源。因此，强烈建议通过进一步研究，将当前讨论与科学专业学生进行比较，从而加以完善。最后但同样重要的是，由于本研究在获取.....方面存在局限性，因此采用了随机截距模型。

t

由大学信息系统提供更高层级的预测变量。强烈建议采用另一种多层模型，例如随机斜率模型。

结论

印度尼西亚教育体系一直在推行渐进式政策，以提升社会整体的科学素养。本文通过为非理工科印度尼西亚本科生开设科学素养课程，落实了通过高等教育领域支持这一愿景的举措。在大学行政管理体系的框架下，学生隶属于不同的系和学院环境。采用多层模型对科学素养课程进行评估，揭示了系和学院环境对学生整个课程期间表现的影响。本研究发现，大多数人口统计学因素无法对学生科学素养的均值差异产生显著影响。相反，形成性评价和情感测量是两个应当受到认真关注的重要因素。

在评估学生的科学素养时，应考虑这一点，以确保对全体学生更加公平。本文提供的证据应作为一项建议，供高等院校参考，以便为未来从事印度尼西亚科学教育的职前教师开展有效的科学素养教育。

参考文献

BAKK UNY。（2019年）。课程结构。UNY。
http://bakk.uny.ac.id/informasi/peraturan-akademik/struktur-kurikulum

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C.（2015年）。使用 lme4 拟合线性混合效应模型。《统计软件杂志》，67(1)。
https://doi.org/10.18637/jss.v067.i01

Bellová, R., Balážová, M., & Tomčík, P.（2021）。对科学与技术的态度是否与科学教育中的关键领域相关？《科学与技术教育研究》。
https://doi.org/10.1080/02635143.2021.1991298

Cheema, J. R.（2017）。国际学生评估项目科学素养题目中的跨国性别差异项功能。《欧洲发展心理学杂志》，16(2)，152–166。
https://doi.org/10.1080/17405629.2017.1358607

Ding, L.（2019年）。定量物理教育研究的理论视角。《物理评论：物理教育研究》，15(2)。
https://doi.org/10.1103/PhysRevPhysEducRes.15.020101

Efthimiou, C. J., & Llewellyn, R. A.（2004年）。电影作为科学素养的工具。arXiv 预印本 Physics/0404078。

Ersan, O., & Rodriguez, M. C. (2020年). 社会经济地位及其他因素：结合土耳其学生与学校背景对国际数学与科学趋势研究数学成绩进行的多层次分析. 《教育中的大规模评估》, 8(1), 1–32.
https://doi.org/10.1186/S40536-020-00093-Y/TABLES/7

Finch, W. H., Bolin, J. E., & Kelley, K. (2016年). 使用 R 进行多层次建模. In 使用 R 进行多层次建模. https://doi.org/10.1080/00949655.2020.1805564

Fives, H., Huebner, W., Birnbaum, A. S., & Nicolich, M.（2014年）。为初中生开发科学素养测量工具。《科学教育》，98(4)。
https://doi.org/10.1002/sce.21115

Fortus, D., Lin, J., Neumann, K., & Sadler, T. D. (2022年). 情感在全民科学素养中的作用。《国际科学教育杂志》，44(4), 535–555. https://doi.org/10.1080/09500693.2022.2036384

Hamper, B. C., & Meisel, J. W.（2020年）。通过实验与讲授相结合的啤酒酿造课程向非理科专业学生介绍科学素养。《化学教育杂志》，97(5)，1289–1294。https://doi.org/10.1021/acs.jchemed.9b00870

Hobson, A.（2003）。设计科学素养课程。载于 T. M. Cooney（编），《两年制大学科学教学》（第45–46页）。NSTA出版社。

Hobson, A. (2008). 大学科学素养课程出乎意料的成效。《物理教师》，46(7)。https://doi.org/10.1119/1.2981285

Impey, C.（2013）。美国本科生的科学素养。《天文学中的组织、人物与策略》，2(2)，353–364。

Impey, C., Buxner, S., Antonellis, J., Johnson, E., & King, C. (2011). 大学生科学素养的二十年调查。《大学科学教学杂志》，40(4)，31–37。

Kanim, S., & Cid, X. C. (2020年). 物理教育研究的人口统计特征. 《物理评论：物理教育研究》, 16(2). https://doi.org/10.1103/PhysRevPhysEducRes.16.020106

Luo, W., Li, H., Baek, E., Chen, S., Lam, K. H., & Semma, B. (2021). 多层建模中的报告实践：十年后的重新审视. 《教育研究评论》, 91(3). https://doi.org/10.3102/0034654321991229

Matuschek, H.、Kliegl, R.、Vasishth, S.、Baayen, H. 和 Bates, D.（2017）。在线性混合模型中平衡第一类错误与检验效能。《记忆与语言杂志》，94。https://doi.org/10.1016/j.jml.2017.01.001

Miller, J. D.（1998年）。公民科学素养的测量。《公众对科学的理解》，7(3)。https://doi.org/10.1088/0963-6625/7/3/001

Mohammadpour, E., Shekarchizadeh, A., & Kalantarrashidi, S. A.（2015年）。参与国际数学与科学趋势研究的国家科学成就的多层建模。《教育研究杂志》，108(6)。https://doi.org/10.1080/00220671.2014.917254

Nurjati, N., Rahayu, E. Y., & Khabib, S.（2022年）。面向泗水公立高中教师的基于AKM的英语学习国家评估编制技术指导。《Kanigara》，2(1), 216–224。https://doi.org/10.36456/KANIGARA.V2I1.5083

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). 过去100年来科学教育发生了怎样的变化？一项使用自然语言处理的分析。《科学教育》，105(4)。https://doi.org/10.1002/sce.21623

Parkinson, J., & Adendorff, R. (2004年). 科普文章在科学素养教学中的应用. 专门用途英语, 23(4). https://doi.org/10.1016/j.esp.2003.11.005

Priyambodo, E.（2016年）。通过学生的自我调节学习提升协作式学术在线学习的有效性。《教育与学习杂志》（EduLearn），10(4)。https://doi.org/10.11591/edulearn.v10i4.4268

Queiruga-Dios, M. Á., López-Iñesta, E., Diez-Ojeda, M., Sáiz-Manzanares, M. C., & Dorrío, J. B. V. (2020年). 正规教育中公民科学对科学素养和可持续发展目标实现的促进作用。《可持续性》, 12(10), 4283. https://doi.org/10.3390/SU12104283

Rachman, T. A., Latipah, E., Zaqiah, Q. Y., & Erihadiana, M. (2021). 在颠覆性教育时代，通过国际学生评估项目国际评估推进印度尼西亚教育的课程创新。

ACM国际会议论文集系列。
https://doi.org/10.1145/3516875.3516899

Ross, K., Ann Hooten, M. 和 Cohen, G.（2013）。通过跨学科方法提升科学素养。Bioscene，39(1)。

Rudolph, J. L.（2020年）。科学教育失落的道德宗旨。科学教育，104(5), 895–906。https://doi.org/10.1002/SCE.21590

Salamah, A., Prastowo, T., & Hariyono, E.（2022年）。开发研究：基于最低能力评估（AKM）第4级的太阳系概念科学素养评估工具。《物理教育研究与发展期刊》，8(1)，115–124。https://doi.org/10.21009/1.08111

Salehi, S.、Burkholder, E.、Lepage, G. P.、Pollock, S. 和 Wieman, C.（2019年）。人口统计差距还是准备程度差距？入学前准备程度对基础物理学学生学业表现的重大影响。《物理评论：物理教育研究》，15(2)。https://doi.org/10.1103/PhysRevPhysEducRes.15.020114

Santoso, P. H., Istiyono, E., Haryanto, & Hidayatulloh, W.（2022年）。使用机器学习对印度尼西亚物理教育研究文献进行主题分析。《Data》，7(11), 147。https://doi.org/10.3390/data7110147

Simmons, A. B., & Heckler, A. F. (2020年). 基础物理学中的成绩、成绩构成权重与人口统计差异. 《物理评论：物理教育研究》, 16(2). https://doi.org/10.1103/PhysRevPhysEducRes.16.020125

Sjöström, J., & Eilks, I. (2018). 基于教化概念重新审视科学素养与科学教育的不同愿景。载于 Y. J. Dori, Z. R.

Mevarech 与 D. R. Baker（编），《STEM教育中的认知、元认知与文化：学习、教学与评估》（第65–88页）。施普林格国际出版。https://doi.org/10.1007/978-3-319-66659-4_4

Surjono, H. D., Muhtadi, A., & Wahyuningsih, D.（2017）。印度尼西亚本科生多媒体课程中混合式学习的实施。《国际信息与教育技术期刊》，7(10), 783–786。

Surpless, B., Bushey, M., & Halx, M.（2018）。在入门实验课程中培养科学素养：课程设计与评估模型。《地球科学教育杂志》，62(2), 244–263。https://doi.org/10.5408/13- 073.1

Ustun, U., Cansiz, M., Ozdemir, E., & Cansiz, N.（2022年）。用于预测2015年国际学生评估项目中两个表现优异国家（芬兰和新加坡）科学素养的学生层面和学校层面因素。《国际科学教育杂志》，44(4), 579–603。https://doi.org/10.1080/09500693.2022.2037167

You, H. S., Park, S., & Delgado, C.（2021）。深入审视美国学校：哪些特征与科学素养相关？基于2015年国际学生评估项目的多变量多层分析。《科学教育》，105(2)。https://doi.org/10.1002/sce.21609

表1. 基于人口统计属性和院系分布的参与者描述

属性	班级	经济学院				学院				总计	
		市场营销		会计学		爪哇舞蹈		语言与艺术学院			
		n	%	*n*	%	*n*	%	n	%	n	%
性别	Male	5	.03	12	.08	18	.11	4	.03	39	.24
	女性	33	.21	24	.15	29	.19	5	.22	.76	
录取途径	SNMPTN	0	0	0	0	13	.08	0	0	13	.08
	SBMPTN	0	0	0	0	14	.09	14	.09	28	.18
	SM 成绩	35	.22	24	.15	5	.03	12	.08	76	.
	SM Utul	3	.02	12	.08	15	.09	3	.08	43	.27
经费资助	有补贴的	0	0	0	0	27	.17	4	.09	41	.26
	Non-有补贴的	38	.24	36	.23	20	.13	25	.16	.74	
奖学金 是		8	.05	1	.01	6	.04	10	.06	25	.16
	No	30	.19	35	.22	41	.26	9	.18	.84	
High 学校专业	科学	9	.06	12	.08	15	.09	17	.11	53	.33
	Non-科学	29	.18	24	.15	32	.20	2	.14	.67	
居住地	北部苏门答腊	0	0	2	.01	0	0	0	0	2	.01
	West 苏门答腊	0	0	0	0	0	0	1	.01	1	.01
	Riau	0	0	0	0	0	0	1	.01	1	.01
	占碑	1	.01	0	0	0	0	0	0	1	.01
	明古鲁	0	0	0	0	0	0	1	.01	1	.01
	雅加达	0	0	2	.01	1	.01	0	0	3	.02
	西爪哇	1	.01	5	.03	0	0	4	.03	10	.06
	中部 Java	23	.14	13	.08	20	.13	7	.04	63	.
	日惹	10	.06	14	.09	17	.11	14	.09	55	.34
	东爪哇	3	.02	0	0	19	.12	9	.06	21	.13
	East 加里曼丹	0	0	0	0	0	0	1	.01	1	.01
	西努沙登加拉	0	0	0	0	0	0	1	.01	1	.01
父亲教育	硕士	0	0	0	0	0	0	1	.01	1	.01
	学士	11	.07	4	.03	11	.07	10	.06	36	.23
	文凭	3	.02	0	0	2	.01	1	.01	6	.04
	高级人员	18	.11	25	.16	13	.08	4	.09	70	.
	大三学生	0	0	3	.02	10	.06	7	.04	20	.13
	小学	5	.03	4	.03	9	.06	5	.03	23	.14
	未受过正规教育	1	.01	0	0	2	.01	1	.01	4	.03
母亲教育	硕士	2	.01	0	0	0	0	0	0	2	.01
	学士	9	.06	6	.04	4	.03	9	.06	28	.18
	文凭	4	.03	3	.02	3	.02	2	.01	12	.08

属性	班级	经济学院				学院				总计	
		市场营销		会计学		爪哇舞蹈					
		n	**%**	***n***	**%**	***n***	%	n	%	n	**%**
	高级	16	.10	13	.08	21		3	.11	68	
	初级	3	.02	9	.06	14	.09	4	.03	30	.19
	小学	4	.03	5	.03	4	.03	4	.03	17	.11
	未受过正规教育	0	0	0	0	1	.01	2	.01	3	.02
父亲收入	> 4 百万	7	.04	7	.04	2	.01	4	.03	20	.13
	3–4 百万	2	.01	2	.01	7	.04	1	.01	12	.08
	2–3 百万	9	.06	6	.04	5	.03	6	.04	26	.16
	100万至200万	15	.09	12	.08	17		3	.08	57	
	< 1 mio	5	.03	9	.06	16	0.1	15	.09	45	.28
母亲收入	> 4 百万	3	.02	1	.01	1	.01	3	.02	8	.05
	3–4 百万	7	.04	1	.01	4	.03	2	.01	14	.09
	2–3 百万	3	.02	4	.03	2	.01	2	.01	11	.07
	1–2 百万	6	.04	4	.03	4	.03	6	.04	20	.13
	< 1 毫秒	19	.12	26	.16	36		6	.16	.67	

注：着色单元格为每个属性的多数类

表2. 本研究汇总的调查变量。

否	变量	**Code**	来源	量表	**Type**
1	学生表现	SPerf	26 道选择题 SLA-D 的	100 分制	连续/间隔
2	期中考试	期中考试	5道开放式题目	100分制	持续/区间
3	情感的属性	Aff	25 问卷条目 SLA-MB	5点李克特量表	分类/有序
4	性别	Gend	机构	2分	名义的
5	录取途径	Adm	注册员记录	4 点	名义
6	经费资助	Fund		2 点	名义
7	奖学金	Sch		2 点	名义的
8	高中主要的	HS		2 分值	名义的
9	居住地	Res		12 点	名义尺度
10	父亲教育	FEdu		7点评分尺度	名义的
11	母亲教育	MEdu		7 点量表	名义值
12	父亲收入	FInc		5点	名义
13	母亲收入	MInc		5点	名义

表3. 学生表现（SPerf ），

情感属性（Aff）和期中考试（MidTerm）均值差异的方差分析结果（按各班级划分）

按学生所在院系、学院和人口统计因素划分

变量	SPerf			Aff			MidTerm		
	F	***p***	**ICC**	***F***	***p***	**ICC**	***F***	***p***	**ICC**
系	11.21	0.000	0.209	0.162	0.922	0.000	6.902	0.(	0.129
学院	19.04		.185	0.002	0.964	0.000	0.130	0.719	0.000
性别	0.116	0.734	0.000	0.000	0.988	0.000	1.405	0.238	0.007
入学途径	1.968	0.121	0.026	1.360	0.257	0.000	0.719	0.542	0.000
经费资助	0.617	0.433	0.000	0.573	0.450	0.000	0.265	0.607	0.000
奖学金获得者	0.600	0.440	0.000	0.008	0.929	0.000	0.010	0.920	0.000
高中专业	0.746	0.389	0.000	5.182	0.024	0.056	0.945	0.333	0.000
居住地	1.056	0.401	0.004	0.888	0.553	0.000	0.707	0.730	0.011
父亲受教育程度	0.867	0.521	0.000	1.329	0.247	0.030	0.613	0.720	0.000
母亲教育程度	1.162	0.330	0.000	2.245	0.042	0.055	0.228	0.967	0.000
父亲收入	0.793	0.532	0.000	0.557	0.694	0.000	0.619	0.650	0.000
母亲收入	0.511	0.728	0.000	1.185	0.320	0.020	1.059	0.379	0.000

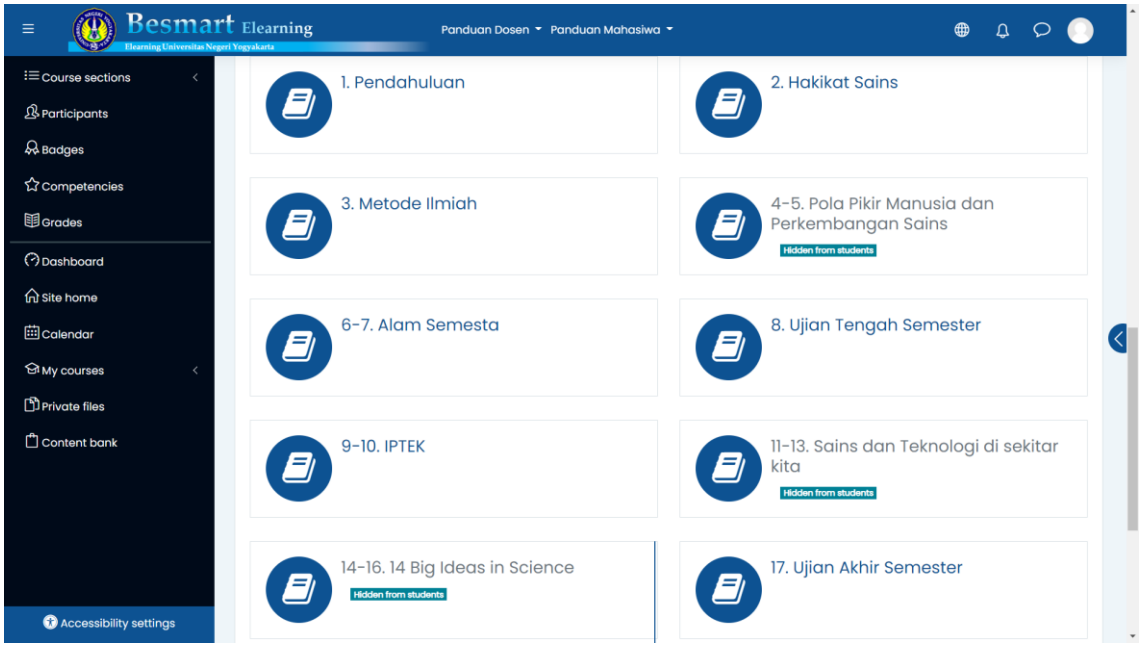
注：有颜色的单元格表示学生所在院系、学院和人口统计因素中各班级之间存在显著均值差异（$\alpha < 0.05$）。

表4. 多层建模结果

参数	模型1	模型2	模型3	模型4	模型5	模型6	模型7	模型8
固定效应								
截距	221.46*(9.545)	221.54*(13.18)	119.86*(26.47)	116.86*(27.85)	21.10(28.12)	18.26(29.44)	-4.00（36.36）	-8.96（37.49）
期中考试			1.27*(0.3132)	1.31*(0.31)	1.28*(0.28)	1.32*(0.27)	1.34*(0.27)	1.39*(0.27)
情感的					1.18*(0.18)	1.19*(0.18)	1.18*(0.19)	1.18*(0.19)
高中							-5.72(5.20)	-5.55(5.19)
母亲教育程度							16.96（22.68）	16.84（22.59）
随机效应								
院系层面								
截距	332.7(18.24)	168.1(12.96)	248.9(15.78)	63.58(7.974)	240.7(15.51)	52.47(7.24)	238.5(15.44)	35.06(5.92)
院系级								
截距		247.4(15.73)		274.31(16.56)		279.03(16.70)		299.21(17.30)
残差	1257.3(35.46)	1257.2(35.46)	1150.9(33.92)	1150.75(33.92)	905.2(30.09)	905.06(30.08)	890.0(29.83)	889.90(29.83)
总体模型								
df	3	4	4	5	5	6	12	13
组数	抑郁 (4)	因素 (2), 抑郁(4)	抑郁 (4)	因素 (2)，抑郁(4)	抑郁 (4)	因素 (2)，抑郁(4)	抑郁 (4)	因素 (2)，抑郁(4)
N（观测值）160		160	160	160	160	160	160	160
AIC	1610.6	1612.4	1596.68	1597.90	1560.00	1560.99	1564.06	1564.63

注：* $p < 0.05$

图 1 BESMART 科学素养与技术课程的电子学习仪表板



F图2. 展示学生（第1级）、系（第2级),

以及学院（第3级）的数据结
构网络）

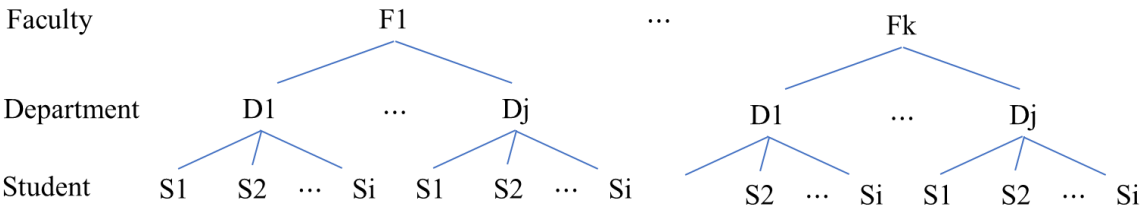
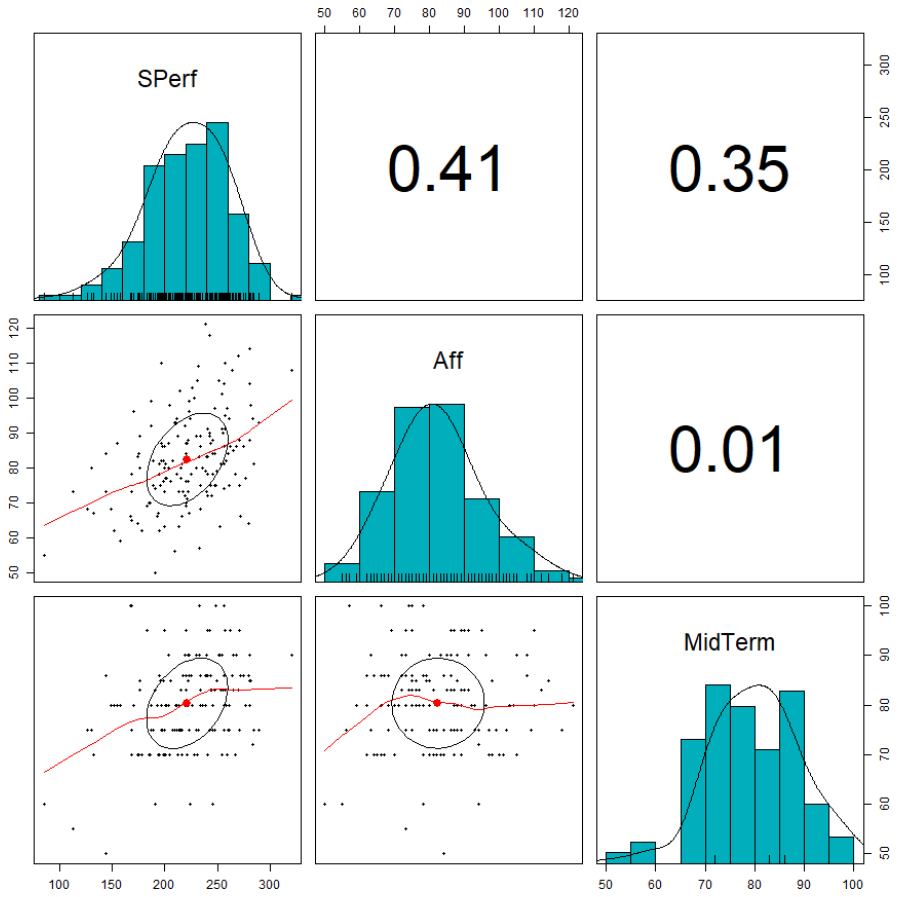


图 3. SPerf、Aff 与期中考试之间的散点图和皮尔逊相关



注：SPerf 与 Aff 和期中考试之间的关联分别为（r = 0.41 和 r = 0.35）。预测变量（Aff 和期中考试）之间的独立性也以可视化方式呈现（r = 0.01）。