

# 软件工程作业情况调查

清华大学计算机系软件工程课程教学团队

2019 年 12 月

## 1 简介

大作业是软件工程实践教学必不可少的环节 [4, 10]。随着敏捷开发在工业界的流行,不少软件工程课程在大作业中采用了敏捷开发的某种形式 [1, 3, 6, 7]。

我们试图探究学生在完成大作业工程中遇到的困难。我们通过调查问卷搜集数据,分析不同问题之间的关联,对开放性问题的回答文本进行主题分析。余下部分按如下方式组织。第2章介绍使用的问卷调查方法;第3章给出调查的直接结果;第4章尝试分析选项之间的潜在关联;第5章讨论一些发现,以及本文可能存在的问题;最后,第6章总结本文的调查结果以及后续需要解决的问题。

## 2 研究方法

本文采用匿名问卷调查的方式收集数据。问卷通过问卷星<sup>1</sup>设计和收集,问卷限制每个微信用户只能填写一次问卷,由此获得的微信名没有用于本文的分析。问卷在微信群中发放,该微信群由 2018 年选修课程的学生创建,大部分成员为 2019 年选修课程的学生,群中有助教和老师的参与;在课程进行过程中,相继有学生通过同学或助教加入微信群,截至停止回收问卷时,群中共有 241 名成员。在发放问卷的同时,发放了总金额 128 元的微信随机红包共 100 个,其中 3 个被助教抢到了。问卷题目主要包括如下方面,全部内容见附录 A。

- 背景信息。如院系、年级、项目、使用的语言;
- 作业相关情况。相关题目尽量采用李克特选项,避免设计的选项对答题者产生影响;
- 主观评估。如每周花费时间、贡献代码行数、大作业难度;
- 开放性问题。包括感到不知所措的情境、以及访问的网站或平台,分析开放性问题回答的方法与分析其他问题回答的方法完全不同。

## 3 调查结果

### 3.1 背景信息

课程早期分组名单中共有 228 人。发放问卷 48 小时内,共回收 96 份答卷,略低于发放的红包数

<sup>1</sup><https://www.wjx.cn>

量 97。通过邮件展开的对开发人员的调查回复率在 6% 到 36% [9]。我们的答卷大概覆盖 42% 的学生,高于 [9] 中报告的回复率。填写问卷用时如图 1,用时最短为 53 秒,没有回答选填的开放性问题;最长有几十个小时,可能是因为通过浏览器访问问卷链接但迟迟没有完成。回收的 96 份答卷中,年级分布如图 2,包括 90 名计算机系学生。96 人中,37 人在上大学之前学习过编程,占比 38.5%。

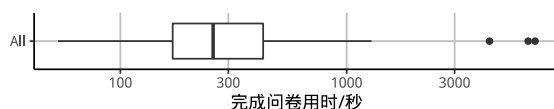


图 1: 完成问卷的时间分布

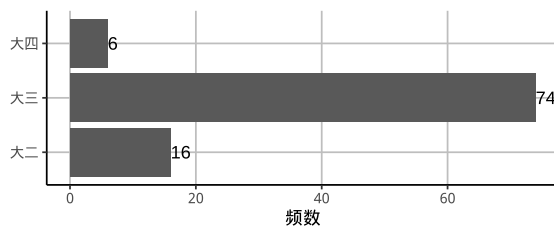


图 2: 年级分布

图 3 呈现了同学们使用的编程语言/框架情况。“其他”选项中的内容被展开,扩充语言/框架列表。由于没有出现在正式选项中,新加入的语言/框架可能与其他选项存在重复,或者没有被其他答卷纳入考虑。新加入的语言/框架包括: C#、Node.js、Docker、js/css/html、Java、kbengine、Python、websocket。

答卷的项目分布如图 4,计原软工联合实验与其他项目的差异很大,本文之后的分析除去了计原软工联合实验的 2 份答卷。

### 3.2 作业相关情况

七成学生参与编写单元测试,三分之二的学生参与配置 CI/CD,如图 5 所示。图 6 将“不同意”与“同意”在基准线两侧分别绘制。五点李克特选项存在趋中倾向,“不确定”被划入“不同意”的范畴,对于大部分问题,这意味着结论倾向于认为教学环节存在问题。大部分学生希望获得更多来自课程的帮助、大作业不集中在例会前完成、清楚每周的任务。半数学生不在本地执行单元测试、代码风格检

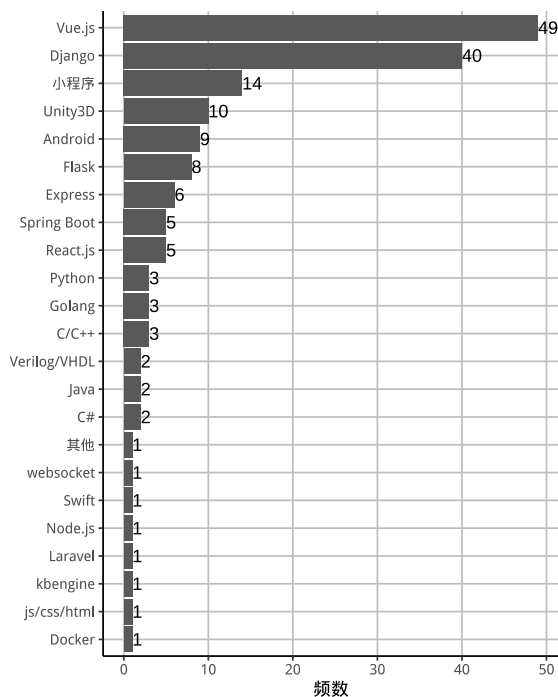


图 3: 编程语言/框架流行度

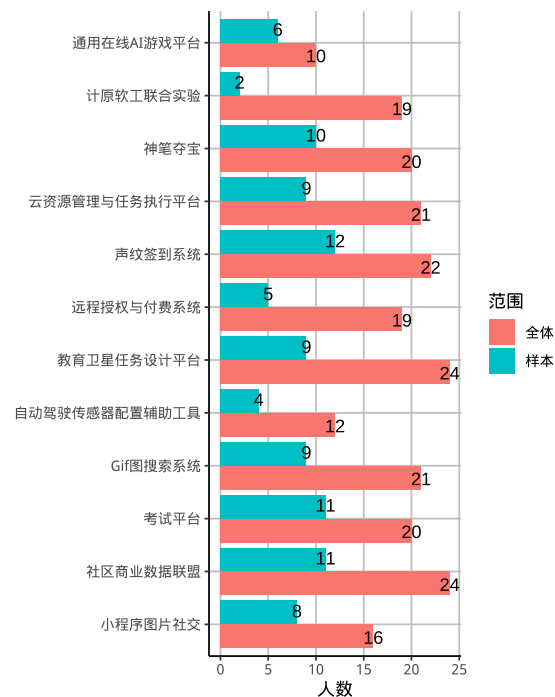


图 4: 项目分布

查，不参考 SonarQube 报告，认同单元测试作用的学生刚刚超过一半。

问题“在完成大作业的过程中，你会向谁寻求帮助？”以排序题的形式给出，访问问卷时，选项顺序随机。该题共 6 个选项，其中，“其他”选项可进一步填写内容，唯一说明的“其他”内容可归入“同学”。对于一份答卷，排在第  $i$  位的选项计  $7-i$  分，没有出现的选项计 0 分，每个选项的平均得分如图 7。主要的解决问题途径是“自行解决”和“同学”，排在第三位的是“老师/助教”。

### 3.3 主观评估

对于大作业的组内成员差异，评分倾向如图 8。20% 的受访学生倾向于同组同分，大部分学生倾向于提升贡献多的成员的分数。单选题型可能限制了受访者选择选项之间的组合，例如降低贡献少的成员分数同时提升贡献多的成员的分数。在“其他”选项中填写的内容如下

“组长给分，大致相同”

“贡献过低建议 0 分”

“增加小组成员互评环节”

“感觉缺乏一个能够衡量所有人工作量的标准”

受访者估计的每周用于完成大作业时间分布如图 9，贡献的代码行数如图 10。图 11 呈现了学生对大作业难度的主观评价，三分之二的学生觉得大作业有难度。

### 3.4 开放性问题

并非所有答卷都回答了开放性问题；在一份答卷中没有出现的内容，也并不代表答题者对相关内容没有看法。因此，本节不再统计某一类回答出现的数量，而只关心有哪些回答。

共有 63 份答卷给出了解决问题时访问的网站/平台，汇总如表 1。54 份答卷给出了完成作业过程中感到不知所措的情境，29 份答卷在问卷最后留言。对上述答卷文本采用主题分析方法 [2] 进行标注、归类，情境分类如图 12 所示。一个情境没有被包含在图示分类中，受访者按照问卷题目要求估计代码贡献行数时感到困难。

表 1: 常用网站/平台

类别	举例
官方文档	Django, Vue.js, Vuetify.js, MDN, Unity, element-UI, 微信文档, Python, 从 pypi 跳转到的项目主页/文档
论坛	vuejs 论坛, unity 社区, 小程序社区, 微信开发者社区
博客	CSDN, 简书, 博客园, Medium
在线问答教程	StackOverflow, SegmentFault, 知乎菜鸟教程, YouTube
开源社区	GitHub
搜索引擎	Google, 百度, Bing
其它	gitlab, 软工平台帮助文档, gitlab 样例仓库

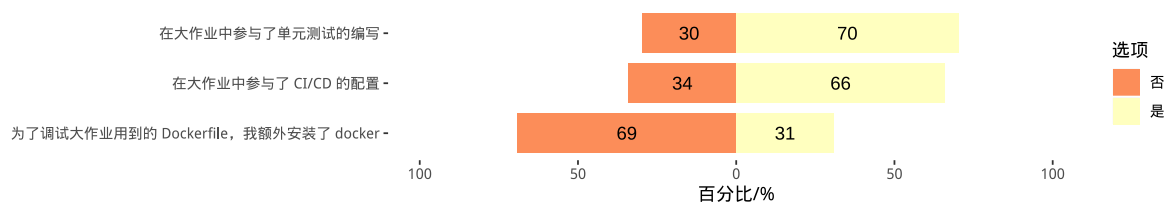


图 5: 是否参与单元测试与配置 CI/CD

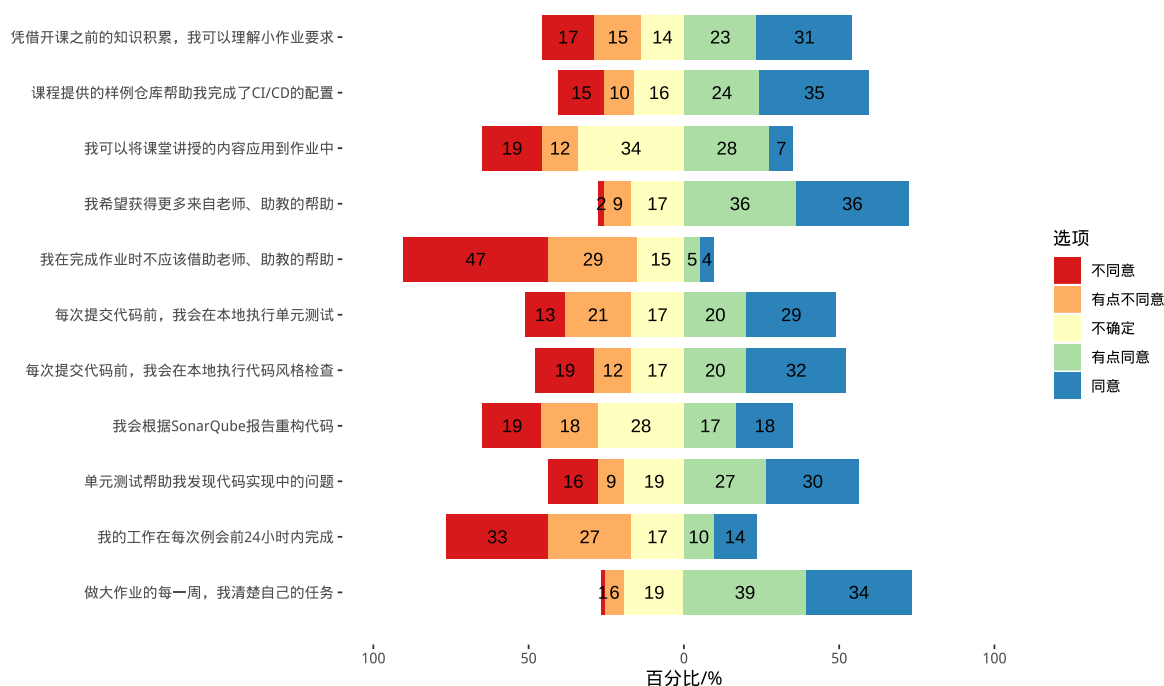


图 6: 作业相关情况

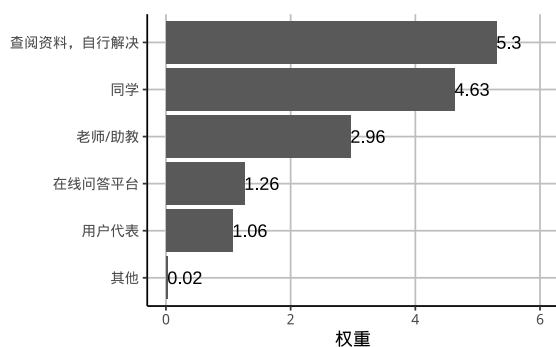


图 7: 在完成大作业的过程中, 你会向谁寻求帮助?

### 3.4.1 上手困难

持续集成 (Continuous Integration, CI) 与持续部署 (Continuous Deployment, CD) 是规范开发过程的重要实践, 在课程中, CI 涵盖单元测试、代码风格检查与 SonarQube 分析。课程提供了 Django、Flask、Golang、React.js、Maven 等框架/语言的 CI/CD 样例仓库, 依然没能覆盖学生在大作业中使

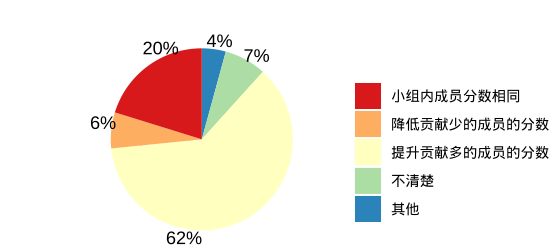


图 8: 对于大作业的组内成员差异, 你更倾向怎样给分?

用的全部工具, 例如 Ant、Laravel、Unity、CI 中使用 MongoDB。学生在参考样例仓库后, 配置 CI 以及部署在教学平台时依然存在困难。对于 CI 中的测试, “由于单元测试的环境和实际部署的环境并不完全相同而出现了单元测试时不会发生的问题”。框架的部署模式本身也和开发模式不同, 例如 “部署之后找不到静态文件”。

课程要求的工具的学习成本集中体现在 “git 冲

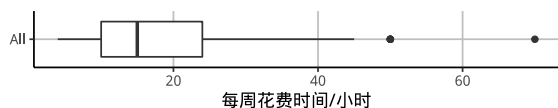


图 9: 每周用于完成大作业的时间估计/小时

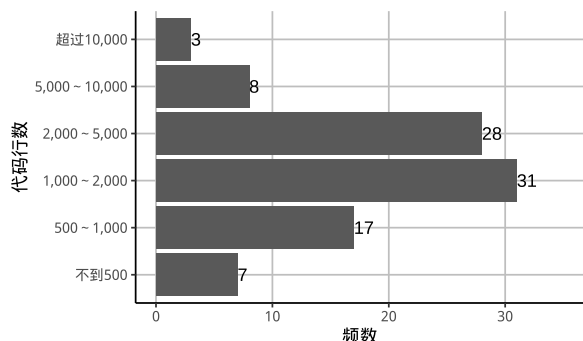


图 10: 最后一次例会的大作业版本中, 贡献代码行数估计

突不会解决”、单元测试、学习新框架、代码重构。有学生希望“助教讲讲单元测试和 *mock* 吧”; 为大作业提供过渡的第一次小作业本身也让学生觉得不知所措, 仅 54% 的受访者表示, 凭借开课之前的知识积累可以理解小作业的要求。

此外, 一些困难是因为工具本身的不成熟。“神笔夺宝”项目使用用户代表提供的 SDK 进行开发, 但“用户代表给的 *sdk* 不知道怎么用”。小程序在课程使用的框架中流行度位列第三, 而“微信小程序自己的(有)*bug*”; 对于“远程授权与付费系统”项目, 在小程序中调用 NFC 和蓝牙的技术方案最终发现不可行。

### 3.4.2 缺少技术支持

大作业实现中, 或多或少会遇到课堂上不会涉及的技术, 例如异步编程、游戏的帧同步、UI 设计。学生在自行查找解决方案时可能会遇到不知道要查什么或查不到的困境。一些回答摘录如下。

“不知道哪里出了 *bug*, 明明看起来是对的。。”

“某个功能在测试时没问题, 但是展示时出现问题”

“当实现新功能时, 查找网上所有的示例代码都不能正常运行时, 会不知所措。”

“有一些前端问题, 网上有很少有有价值的参考资料, 所以问题解决耗时过多”

### 3.4.3 交流

有的小组在开发过程中“没有明确的迭代计划”。接口的约定可能出现问题, 例如, “两个前端分别同样内容给出不同的接口要求, 而且还和我之前和他们说的并写在文档里的我自己设的接口不一样”; “前后端的对接”也可能存在困难。

队友的情况也会影响学习体验。队友不回复消息、事情多、调不出 *bug*、没有进度、退课都会让人感到焦虑, 类似情况如果发生在关键路径上更是让人无奈。“凌晨一点半接到需求电话”的队友也是倍感压力。

### 3.4.4 得分

不明确的计分方式会让学生迷茫, 有学生提到, “分工的时候后端单元测试分给了组长, 但最后评分的时候似乎不参与单元测试会扣分?” 一些建议包括为课堂参与提供总评加分、引入“小组成员互评环节”、将评价标准替换为“完善了什么提升体验的细节”而非“实现了多少功能”。答卷文本中也包含了一些反思。

“虽然说个人觉得需要根据组内贡献体现分数差异, 但有时候也不大容易评判贡献。一是因为代码行数其实不是绝对的评判标准, 二是其他组员研究的模块, 组里面其他人也不太清楚难度。”

### 3.4.5 教学计划

一份答卷表示喜欢现有的大作业时间安排, “其实我挺赞成这种把大作业分布在一个较长的周期然后每周检查, 这样可以强行避免自己最后赶 *ddl*, 写出来的东西质量也相对较高”。助教在工作中存在问题, 一份答卷提醒助教“注意自己言行”。其他建议包括,

- “随机分组”或可分摊遇到合作困难的队友的情况;
- 可以借鉴“国外真正的世界一流大学的计算机系的软工课是怎么开设的”;
- “可以优化作业的类型配比, 感觉那种主要需求是搭个有多种权限网站的项目比例还是太高了”;
- 将软工放到大二, “可以早点学会 *git*”, “可能同学们找实习也更方便”, 现有工作量“适合大四不在意绩点且时间多的同学”。

学生对教学计划也存在疑惑,

“本课程最奇怪的地方就是上课内容和大作业内容完全独立, 上课第 8 周之后讲的内容对大作业其实是有帮助的, 但那个时候大作业基本上都截止了。”

“对于大部分同学而言, 之前没有这样的实际开发经验, 可能不太能体会到其重要性, 也不知道该如何将其贯彻大作业中。”

基于现有的大三上学期整体教学计划, 软件工程课程大作业主体安排在前半个学期, 以尽量避免与计算机组成原理大作业的交集。教学计划考量、以及课程意义与应用场景, 需要在课程的开始明确。



图 11: 大作业难度评价

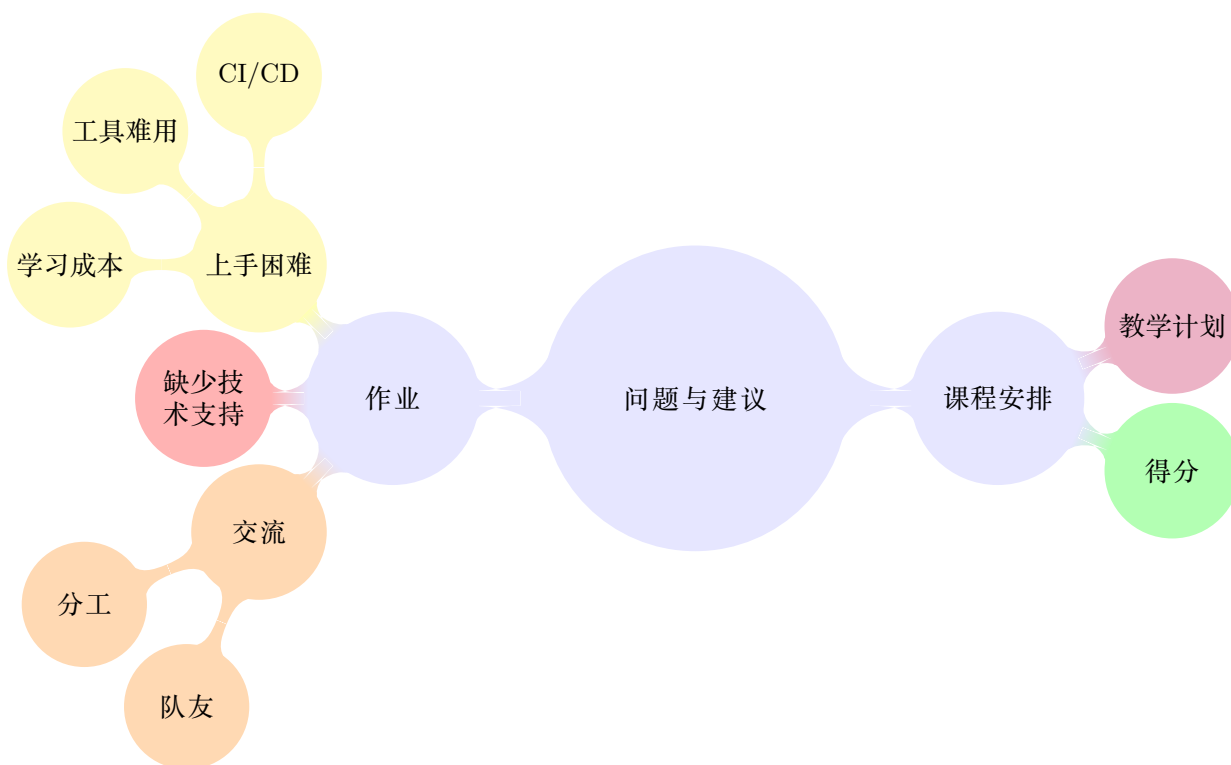


图 12: 答卷文本模式图

## 4 相关性分析

对于作业相关情况的 11 个问题，将五点李克特选项的“同意”、“有点同意”归为“同意”，“不同意”、“有点不同意”、“不确定”归为“不同意”，以简化分析过程。本章对选项之间的关联进行探讨。

### 4.1 样例仓库的作用

对“额外安装 docker”与“参与 CI/CD 的配置”、“额外安装 docker”与“样例仓库帮助完成 CI/CD 的配置”进行卡方检验得到的  $p$  分别为 0.5178、0.7238，没有表现出相关性。“参与 CI/CD 的配置”与“样例仓库帮助完成 CI/CD 的配置”的列联表如表 2，进行卡方检验得到  $p = 0.000794 < 0.01$ ，一定程度上表明样例仓库可以帮助降低 CI/CD 配置的学习成本。

### 4.2 本地单元测试

对“参与单元测试的编写”与“在本地执行单元测试”进行卡方检验， $p = 0.3204$ ，没有表现出二

表 2: “参与 CI/CD 的配置”与“样例仓库帮助完成 CI/CD 的配置”列联表。进行卡方检验， $p = 0.000794 < 0.01$

	没有配置 CI/CD	配置 CI/CD
样例没用	21	17
样例有用	11	45

者的相关性。“单元测试可以发现代码实现中的问题”与“在本地执行单元测试”的列联表如表 3，进行卡方检验得到  $p = 0.001649 < 0.01$ ，这意味着了解单元测试的意义与遵循测试要求之间存在联系。

表 3: “单元测试可以发现代码实现中的问题”与“在本地执行单元测试”列联表。进行卡方检验， $p = 0.001649 < 0.01$

	单元测试没用	单元测试有用
本地不执行测试	29	19
本地执行测试	12	34

### 4.3 寻求帮助

“凭借开课前知识积累可以理解小作业要求”与“希望获得更多来自课程的帮助”的列联表如表 4，二者没有表现出相关性，即便可以理解小作业要求的学生也希望获得更多来自课程的帮助。

表 4: “凭借开课前知识积累可以理解小作业要求”与“希望获得更多来自课程的帮助”列联表。进行卡方检验， $p = 0.5189$

	不能理解小作业	可以理解小作业
不需要帮助	10	16
需要帮助	33	35

抽取“在完成大作业的过程中，你会向谁寻求帮助”一题中，“老师/助教”的排名，与“希望获得更多来自课程的帮助”的联合分布如表 5。57 份答卷将“老师/助教”排在前三位；15 份答卷从未向老师、助教寻求帮助，但仍然希望获得更多来自课程的帮助。

表 5: “希望获得更多来自课程的帮助”与“学生以老师、助教为帮助来源的排名”列联表

排名	希望获得更多来自课程的帮助		
	不同意	不确定	同意
1	1	1	4
2	0	2	4
3	1	7	37
4	1	1	8
5	0	1	0
从不	7	4	15

## 5 讨论

### 5.1 发现与认识

基于调查结果，我们对学生作业完成情况主要有如下发现与认识。

发现 1 在大作业与课程内容之间，需要有指导的练习。

学生独立完成任务的能力与完成作业需要的能力之间可能存在差距，此时，适当的帮助可以促进学生完成任务并提升能力上限 [8]。小作业中，学生可以练习课堂讲授的内容。为了应对不同基础的学生，小作业需要提供细致的说明。

问题 1 实现大作业需求时缺少技术支持。

代码调试是学生遇到的难点之一，发生在计算机系学生学习的各个阶段，目前还没有课程讲授如何调试代码。代码修复是软件工程研究方向之一，但研究成果大多止步于实验室。

另一个难点是由需求引入的。助教不可能熟悉学生使用的各种语言、框架，能提供的帮助有限。同时，并非所有用户代表都能对需求提出明确有效的解决方案，需求方提供的 SDK 甚至可能增加任务复杂度。

发现 2 在完成大作业的过程中，一半左右的学生没有贯彻测试、代码风格检查等最佳实践。

不遵循课程要求完成作业，将使教学质量打折扣。相关性分析表明，认为单元测试有用与在本地执行单元测试之间的相关性在统计意义上明显。这意味着，让学生了解到教学实践各部分的意义，有助于学生接受课程介绍的最佳实践。

问题 2 部分学生存在感不强，不能与队友进行有效合作。

小组作业中可能存在学生划水，但学生之间也确实存在差异。基于问卷调查的结果，本文无法对部分学生存在感不强的原因作出解释。

问题 3 大作业需要明确评价标准。

工作量是评价个体差异的简单手段，但工作量本身缺少衡量标准，不同项目之间的难度差异也会影响对工作量判断。软件生产力估计是软件工程的研究方向之一，越精确的估计也需要花费更高的成本 [5]。

发现 3 学生需要了解课程的意义，有的学生希望课程前移。

说明课程意义有助于帮助学生课程。目前大三上学期各课程工作量向学生施加了巨大压力。前移软件工程课程，也可以让学生尽早运用课程中介绍的工具、最佳实践和原理。

### 5.2 有效性检验

#### 5.2.1 构造有效性

构造有效性表征研究方法本身的有效性。微信群没有包含全部选课学生，不是所有选课学生都填写了问卷，调查结果与总体情况之间的差异不明确。问卷本身有其局限性，难以判断选择背后的决策因素。例如，对于在大作业中存在感不强的学生，问卷结果只能呈现其他学生的态度，而不能了解参与程度低的原因。

#### 5.2.2 内部解释性

内部解释性指研究对象是否被正确度量。问卷是一种方便的统计方法，问卷议题由研究者设定，可能影响答题者的判断。我们采用五点李克特选项设计单选题，以尽量避免研究人员引入的选项对答题者的额外影响。对工作时间和代码贡献行数的估计可能与事实存在偏差，相关数据没有进一步进行解释。开放性问题允许受访者更完整地表达自己的意见，但对文本进行主题分析时，不免受到研究人员既定认识的影响。

### 5.2.3 外部解释性

外部解释性指研究结果是否适用于其他群体。答案是否定的。调查结果在不同届学生之间可能存在差异。不同课程、不同院系、不同学校不可照搬调查结果。

## 6 总结

本文通过调查问卷搜集数据，对结果进行相关性分析，并对答卷文本进行主题分析。我们发现，

- 在大作业与课程内容之间，需要有指导的练习；
- 在完成大作业的过程中，一半左右的学生没有贯彻测试、代码风格检查等最佳实践；
- 学生需要了解课程的意义，有的学生希望课程前移。

同时，课程也面临如下困难，

- 实现大作业需求时缺少技术支持；
- 部分学生存在感不强，不能与队友进行有效合作；
- 大作业需要明确评价标准。

在未来的课程中，需要考虑在评分中引入课堂表现和组内互评的部分、设计小作业，并参考其他学校计算机系软件工程课程的课程设计。

## 致谢

感谢同学们对课程的支持。感谢用户代表在课程教学实践中的付出。感谢腾讯云提供的云计算资源。

## 参考文献

- [1] Lukas Alperowitz, Dora Dzvonyar, and Bernd Bruegge. Metrics in agile project courses. In *Proceedings of the 38th International Conference on Software Engineering Companion - ICSE '16*. ACM Press, 2016. doi: 10.1145/2889160.2889183.
- [2] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. doi: 10.1191/1478088706qp063oa.
- [3] Bernd Bruegge, Stephan Krusche, and Lukas Alperowitz. Software engineering project courses with industrial clients. *ACM Transactions on Computing Education*, 15(4):1–31, dec 2015. doi: 10.1145/2732155.
- [4] Carlo Ghezzi and Dino Mandrioli. The challenges of software engineering education. In Paola Inverardi and Mehdi Jazayeri, editors, *Software Engineering Education in the Modern Age*, pages 115–127. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-68204-2.
- [5] Capers Jones. 软件工程通史：1930-2019. 清华大学出版社, 2017. ISBN 978-7-302-46133-3.
- [6] Phillip A. Laplante. An agile, graduate, software studio course. *IEEE Transactions on Education*, 49(4):417–419, nov 2006. doi: 10.1109/te.2006.879790.
- [7] M. Marques, S. F. Ochoa, M. C. Bastarica, and F. J. Gutierrez. Enhancing the student learning experience in software engineering project courses. *IEEE Transactions on Education*, 61(1):63–73, Feb 2018. ISSN 1557-9638. doi: 10.1109/TE.2017.2742989.
- [8] Saul A. McLeod. What is the zone of proximal development?, 2019. URL <https://www.simplypsychology.org/Zone-of-Proximal-Development.html>.
- [9] E. Smith, R. Loftin, E. Murphy-Hill, C. Bird, and T. Zimmermann. Improving developer participation rates in surveys. In *2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 89–92, May 2013. doi: 10.1109/CHASE.2013.6614738.
- [10] Hans van Vliet. Reflections on software engineering education. In Paola Inverardi and Mehdi Jazayeri, editors, *Software Engineering Education in the Modern Age*, pages 1–10. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-68204-2.

## A 调查问卷

编号	问题	题型
1	院系 计算机系 其他	单选
2	年级 大一 大二 大三 大四 其他	单选
3	你参与的项目是	单选
4	你在大作业中使用的编程语言/框架有 其他	多选 填空
5	以下描述是否符合你的情况	单选，是/否
5.1	我在上大学之前学习过编程	
5.2	我在大作业中参与了单元测试的编写	
5.3	我在大作业中参与了 CI/CD 的配置	
5.4	为了调试大作业用到的 Dockerfile，我额外安装了 docker	
6	在软件工程课内完成小作业、大作业的过程中，你在多大程度上认可以下论述	五点李克特选项
6.1	凭借开课之前的知识积累，我可以理解小作业要求	
6.2	课程提供的样例仓库帮助我完成了 CI/CD 的配置	
6.3	我可以将课堂讲授的内容应用到作业中	
6.4	我希望获得更多来自老师、助教的帮助	
6.5	我在完成作业时不应该借助老师、助教的帮助	
7	在软件工程课内完成大作业的过程中，你在多大程度上认可以下论述	五点李克特选项
7.1	每次提交代码前，我会在本地执行单元测试	
7.2	每次提交代码前，我会在本地执行代码风格检查	
7.3	我会根据 SonarQube 报告重构代码	
7.4	单元测试帮助我发现代码实现中的问题	
7.5	我的工作每次例会前 24 小时内完成	
7.6	做大作业的每一周，我清楚自己的任务	
8	在完成大作业的过程中，你会向谁寻求帮助？ 同学 老师/助教 在线问答平台 查阅资料，自行解决 用户代表 其他	排序，选项顺序随机
9	关于大作业的组内成员差异，你更倾向怎样给分？ 小组内成员分数相同 降低贡献少的成员的分数 提升贡献多的成员的分数 不清楚 其他	填空 单选
A	你大概每周花多少小时用于完成软件工程专业大作业？	填空
B	最后一次例会的大作业版本中，你大概贡献了多少行代码？ 不到 500 500 ~ 1,000 1,000 ~ 2,000 2,000 ~ 5,000 5,000 ~ 10,000 超过 10,000	单选
C	你觉得大作业难吗？	五点李克特选项



编号	问题	题型
D	在完成小作业、大作业的过程中，你会访问哪些网站或平台来帮助你解决遇到的问题？最多 5 个	
E	在完成小作业、大作业的过程中，哪些情境让你不知所措？最多 5 条	
	你希望获得调查报告吗？	单选
	不用了	
	邮箱	填空
	你愿意就相关话题与助教聊聊天吗？大概 15-60 分钟	单选
	不要	
	邮箱同上一题	
	其他联系方式	填空
	有什么想说的吗？	