The omitted variable: could DuoTest enable a new way to assess the link between individual and team performance in team-based learning?

Riccardo Bonazzi
University of Applied Sciences and Arts Western Switzerland (HES-SO)
riccardo.bonazzi@hevs.ch

Yviane Rouiller,
Haute Ecole Pédagogique du Valais
yviane.rouiller@hepvs.ch

Abstract

Imagine a class of students being allowed to do their final exam twice in a row: the first time, participants do their exam individually and with closed books (Exa01); the second time, they solve the same exam in groups and with open books (Exa02). If you think that all students will get a better grade in the second exam, you would be surprised by the results. This article is part of an ongoing project to develop a method for team-based learning named Testudo. We present an assessment technique called DuoTest, which uses a mixed model to (a) analyze data from individual and group exams and (b) determine the positive (or negative) effect of each team over the individual performances. Empirical results collected from 70 students show that individual exams are a weak predictor of the group scores, whereas the fixed effects of each team are a better predictor of Exa02.

1. Introduction

> *"The men of the first held their shields over their heads, and closed up, so that, owing to the density of the bucklers, it became like a tiled roof[...] in the shape of a tortoise (testudo)".* Polybius [25:11]

The Roman Testudo is a well-known example of a military formation, where soldiers put together their shields to achieve a common goal, such as to protect themselves against a threat or to let other soldiers walk upon it whenever they come to a narrow ravine. Nonetheless, such powerful feature came at a price, since Roman Testudo were said to be advancing slowly in combat, since soldiers had to coordinate themselves. Accordingly, the Roman Testudo and its trade-off could be used as a metaphor for a situation, where students are expected to work together and solve a problem as a team.

There are still mixed evidences on whether working in teams is an appropriate method to prepare students for the challenges of a constantly changing business environment: on the one hand, some teachers prefer to give instruction via teacher-centered methods (lectures with little text reading and student discourse), under the belief that the best way to ensure content learning is for the instructor to present all necessary information to students [16]. On the other hand, some scholars claim that traditional teaching methods do not enable all students to appropriately engage with the types of academic literacy constitutive to higher education [7, 11:6]. Hence, this article starts with a simple intuition to bridge the two viewpoints: if we assume that the team itself is an important outcome of a team project, could we assess, at the end of the course, if the students would have been more/less effective without it? Indeed, there is a consensus on the difficulty of correctly assessing the performance of each student in a team project [2], and most educators lack a simple tool to do it. Nonetheless, most of the previous works have considered the team as noise to be cancelled to assess the individual, whereas we consider it as the most important artefact of a course, which asks students to work in teams to solve real-world projects and reflect on what they learned by doing so.

According to Kolb [10], learning is the process whereby knowledge is created through the transformation of experience. *Group-based learning* is seen as a form of experiential learning and it has been termed differently through the years: (a) *small group learning* [26] include activities where the teacher lectures for 15–20 minutes and then asks students to pair with the student beside them to discuss a question, (b) *collaborative learning* involves carefully planned and structured group activities that are infused into a course of learning, whereas (c) *team-based learning* (TBL) makes intense use of small groups in that it changes the structure of the course, in order to develop and then take advantage of the special capabilities of high-performance learning teams [17]. According to its authors, TBL is an important opportunity for teamwork skill development, experiential learning, and learning from peers. However, TBL presents many challenges and is most appropriate in courses that meet two conditions: (1) students are required during the course to understand a significant body of information and (2) a primary goal of the course is to apply this content by

URI: https://hdl.handle.net/10125/70621
978-0-9981331-4-0

solving problems, answering complex questions and resolving issues [28]. Accordingly, our research question is: **how can we design a summative assessment of individual and team performance in a team-based learning scenario?**

The rest of the paper proceeds as it follows. Section 2 briefly reviews the existing body of knowledge to answer our research question. Section 3, 4 and 5 describe design science as our chosen methodology, highlight the relevant elements of the course which applies the Testudo method and then describe how to create and test the DuoTest prototype. Section 6 presents our preliminary findings, whereas section 7 concludes the paper by discussing the contribution and shortcomings of our work.

2. Literature review

In this section, we briefly assess the existing body of knowledge and define three constructs to avoid the *jingle fallacy* (constructs with the same name referring to different phenomena): (a) *team health*, which can be used to assess how well individuals work together in a team, (b) *transactivity*, to assess how each individual in a team can build on previous works from team members and (c) *immediate feedback assessment technique*, a tool used for summative evaluation in team-based learning that could be used to assess transactivity.

By using the keywords *"experiential learning" "team assessment" "individual assessment" review empirical*, we retrieved 28 results (link). The inclusion criteria of our practical screen were three: (a) no patents, (b) only conference and journal articles, (c) the chosen language was English. In the end, we obtained 10 articles.

Team health from individual contributions (Team ← Individual - or T←I in short) Recent work from O'Neill et al. [23] presents a set of 18 questions to rapidly and reliably assess the *team health* by asking team members to describe their perception of team communication, adaptability, relationships and education. Other scholars have suggested that assessment in TBL should take into account the cognitive, affective and behavioral dimensions [2]. Indeed, students have multiple goals and motivations, which influence the team performance: mastery goals ("I want to learn new things") and social responsibility goals ("I want help my peers") prevail in effective teams, whereas belongingness goals (e.g., "I want my peers to like me") were more important than mastery goals in ineffective teams [9].

Team effect on the Individual performance (T→I:). To some degree, the group product will be codified in an artifact (e.g., group report, dialogue, diagram, etc.), but the individual experience of that collaborative learning event will be transposed to future collaborative learning events. [27]. Accordingly, the team effect can be associated to *transactivity*, that is the extent to which students refer and build on each other's' contributions and it can be measured by reflected in collaborative dialogue or individual products, or the extent to which students transform a shared artifact (e.g., a group report) [30].

Gap in the literature: IF-AT to assess transactivity. The *immediate feedback assessment technique* (IF-AT) form has (a) a series of boxes covered by an opaque, waxy coating similar to that found on scratch-off lottery tickets corresponding to the alternatives, with only one correct alternative having with a small star in it [14]. The authors found that students who did the final exam with the Immediate Feedback Assessment Technique (IF-AT) scored 10% more on average when they got partial credit for iterative responding (they could scratch more than one box). Although, this approach is already used in team-based learning scenarios [15], there is not a simple way to use it and assess how team transactivity influence individual performance.

3. Chosen methodology to develop and test the artefact

In this section, we position our study in the field of design science research [8] and we describe how we developed an artefact in the shape of a *prototype* [12], by following the guidelines of Peffers et al. [24].

(i) Identify problem and motivate. In section 1, we have described the opportunities associated with team-based learning and the current challenges associated with the assessment of individual and team performance.

(ii) Define objectives of the solution. As mentioned in section 2, we wanted to improve the immediate feedback assessment technique (IF-AT) by developing an online solution, which could allow students to do the final exam by themselves and then to get partial credits if they managed to correct their mistakes, by discussing with their team members. This way, we could measure the degree of transactivity in each team.

(iii) Design and development of the artefact: the DuoTest prototype. The underlying idea of *DuoTest* is simple: to allow students to do their final exams twice in a row. The first time, participants do their exam individually (Exa01); the second time, they solve the same exam in groups (Exa02). By comparing individual and team performances, the system induces the positive (or negative) effect of each group over the individual performances.

(iv) Demonstration. Section 5 illustrates in detail how the DuoTest can be made by using an open-source learning management system (Moodle) and how the data can be analyzed with R Studio to assess team health and transactivity.

(iv) Evaluation. We tested our prototype with three classes of undergraduate students undertaking the same course, for a total of 71 students attending the final exam in Sierre (Switzerland) the 20th of January 2020. We claim that the exam was (a) valid, since chosen questions provide useful information about the concepts seen in class, (b) reliable, thanks to the rule-driven correction of each question, and (c) recognizable, since it fully replicated the way students work during the semester.

(vi) Communication. Since our test took place in January, in the spring semester we shared our preliminary insights with colleagues, whose courses have been disrupted by the Covid-19 situation.

4. The artefact

Table 1 illustrates how DuoTest fits in the overall approach called *Testudo,* which splits the course in three sets of team-based learning activities and evaluations. At the beginning of the semester, students play a multi-round business simulation game [13]. In this phase, students are assigned to a new random group every week, to learn how to rapidly work together and take decision under uncertainty [5]. Starting from week 5, students form a group of max five team members. In this phase, students are assigned to a real project done with an external firm for eight weeks. All projects respect the five criteria for a project-based learning activity [29]: (a) projects are central to the curriculum, since the score given to the students reports will count as their midterm exam, (b) they are focused on problems that 'drive' students to encounter/struggle with the central concepts of a discipline, (c) they involve students in a constructive investigation, since students have to help the firm make sense of its data to find the solution, (d) they are student-driven to a significant degree, and (e) they are realistic and not school-like. Every week, students are asked to fill in a new section of the report and to submit it on a Moodle Workshop activity [20], where it will be assessed by their peers. During each class, the teacher briefly clarifies the required activities and facilitates discussions among team members. Slides are seldomly presented in class, since they are available to students in advance, together with check-up questions, as Moodle Lessons [21]. Accordingly, we state three hypotheses, which we would like to test:

- ***H1****: the individual performance of Exa01 has a positive and statistically significant effect over the individual performance of Exa02*. This statement is supported by all the reviewed literature on team-based learning

Table 1: Overall view of how the focus of this article (DuoTest) fits in the Testudo approach

Activities in Team-Based Learning [3, 10, 17]	**Participation score (DeTotus)** When: Weeks 1-12 How much: **20% of final score**	**Midterm Evaluation** When: Weeks 13-14 How much: **30% of final score**	**Final Exam with DuoTest** When: Week 15 How much: **50% of final score**
(i) Pre-reading	Interactive slides	---	---
(ii) Test: Individual Readiness Assurance	Peer-review (part B)	Individual score by role	Individual exam (Exa01)
(iii) Test: Team Readiness Assurance	Team project review	Team score by integration	Group Exam (Exa02)
(iv) Clarifying activity	15' debrief with professor	Written feedback by peers	--
(v) Knowledge application	60' project work (team/role)	Improve report	--
(vi) Report	Templates to fill-in (part A)	Self-Assessment before submit	--

- **H2**: *the team performance (transactivity) has a statistically significant effect over the individual performance of Exa02*. If this hypothesis is correct, we should be able to see different improvement in different teams, depending on their degree of transactivity.
- ***H3:*** *the team performance has positive and statistically significant effect over the individual performance of Exa02*. H3 extends H2. Based on previous results of Maurer and Kropp [14] concerning IF-AT with partial credit, we could assume that a student having the possibility to correct his mistakes by discussing with his team will improve his final score.

5. Demonstration of how DuoTest works

Before the exam, we create a Moodle Quiz activity [18] with ten questions: five theoretical questions and five questions about a case study. The type of the ten questions is Short Answer [22]: this information will be relevant when we explain how to analyze the data after the exam. In the parameters of the Moodle Quiz activity, hereinafter referred to as *Exa01*, we set the duration at 35 minutes. Then, we copy the Quiz activity a second time, hereinafter referred to as Exa02. This way, the questions of Exa02 are the same of Exa01. In the parameters of Exa02, we set the beginning of the activity 5 minutes after the end of Exa01, to allow students the logistical time to setup their teams in the class. The duration of Exa02 is set at 20 minutes, which brings the total to 60 minutes. Finally, in the Moodle Gradebook [19], we set the score of the final exam as the average between Exa01 and Exa02.

During the exam, students are expected to do Exa01 without additional material and by themselves. When Exa01 is over after 35 minutes, each student assembles with the team members, with whom he has been working between week 5 and 12. Students can talk among them during Exa02 and they have access of any type of material. Indeed, Exa02 recreates the conditions that the team has lived during the semester and allows educators to assess in detail the dynamics of each team.

After the test, each answer is corrected by using a special feature of Short-answer questions: the educator defines a set of rules in the parameters of each question, and the answers of all students are corrected automatically by Moodle. This assures a coherent assessment all along and it increases the rigor of the overall process. In the end, we can export from Moodle a table, like the one shown in Table 2.

Here, we assume that we have only one question and two students in each group, who answered individually in the first exam Exa01 and then tried to find the right answer together for Exa02. As we can see, Student 01 answered correctly (Exa01 = 1), but after the discussion with Student 2 she changed her answer and made a mistake (Exa02 = 0). Meanwhile, student 3 successfully managed to help Student 4 improve his answer in Exa02. Finally, Student 5 and 6 managed to work together and find the right answer in Exa02, even if they made a mistake during Exa01. By exporting the two datasets from Moodle (Exa01 and Exa02) we can combine them to assess the changes in score and compute the group effect on the individual performance of each student. The group A had a low transactivity, since the change in scores was negative on average. Group B had a positive transactivity, but it concerned only one student out of two, whereas Group C had a high transactivity, since both students improved their scores after working together.

In the next page, Figure 3 shows the real data from the two exams of 70 students in 16 groups, and it assigns a colored dot to each student in a team, to visualizes the change in performance across the two exams.

To get familiar with the Figure 3, we suggest starting with the dot from G07 that has the lowest score for Exa01 and that does not change between Figure 3a and 6.2b: Group 07 had a student, who attended the exam to help his team, but who did not gave answers to most of the questions since he was about to drop out the university (row 18 in the table of Appendix A). The next section will analyze more in depth how to convert intuitions seen in Figure 3 into quantitative measures.

Table 2: Fictive example of student scores from Exa01 and Exa02, to assess group effect (transactivity)

STUDENT	GROUP	EXA01
Student1	A	1
Student2	A	0
Student3	B	1
Student4	B	0
Student5	C	0
Student6	C	0

STUDENT	GROUP	EXA02
Student1	A	0
Student2	A	0
Student3	B	1
Student4	B	1
Student5	C	1
Student6	C	1

GROUP	DELTA	EFFECT
A	-1	Low
A	0	Low
B	0	Medium
B	+1	Medium
C	+1	High
C	+1	High

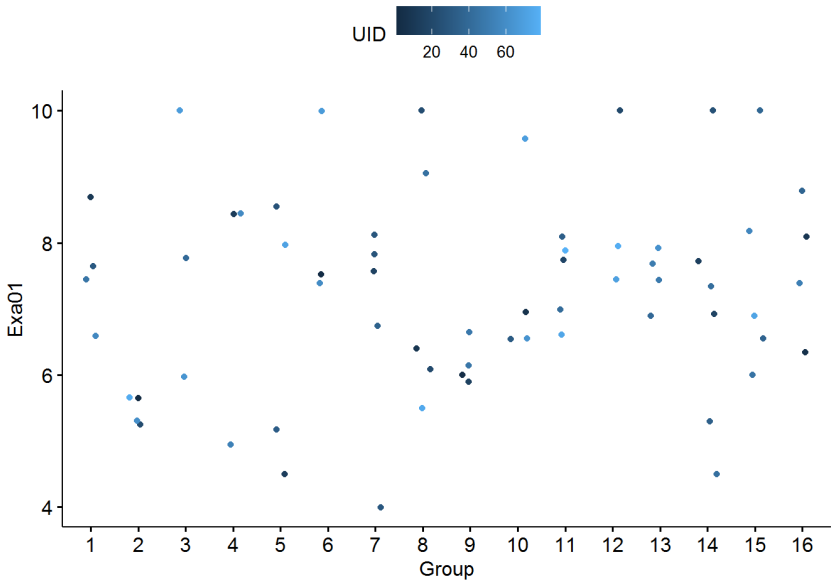


Figure 3a: Individual Exam Scores

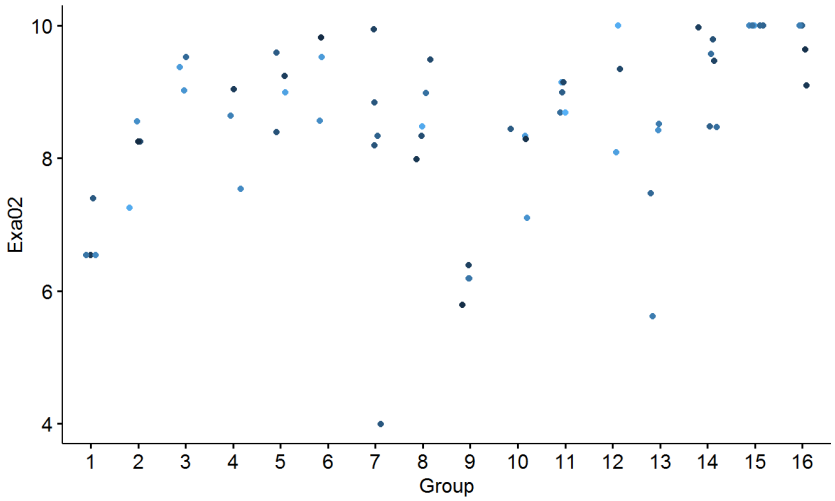


Figure 3b: Group Exam Scores

6. Preliminary findings

This section analyses the results of the individual and the group exams, which are illustrated in Figure 3. The complete dataset is available in Appendix A. As shown in the Table 2 for Group A, some teams performed worse in the second exam, the groups G01, G10 and G13 being examples of students, who decided to change some correct answers into wrong answers after discussing with the rest of the team. One can also find examples of the Groupe B in Table 2, such as G02 (which had a strong concentration of scores below 6/10 and shifted up above 8/10) or groups G14 and G15, where one student managed to lift the scores of all the team members Finally, G16 is an example of students with average scores for Exa01 working together to get high scores in Exa02, as shown by the fictive Group C in Table 2.

To assign some quantitative data to our assessment, we start by scaling the raw data and properly compare the coefficients of each variables.

Table 4 illustrates that the performance of the first exam (Exa01) positively influences the score of the second exam (Exa02), with a coefficient of 0.22 (hence Exa02 = 0.22*Exa01). The value of p = 0.08 is below 0.10. Therefore, **we cannot reject the hypothesis H1**, and we affirm that there is a weak causal effect between the Ex01 (done individually) and Exa02 (done in group). Nonetheless, the Adjusted R2 = 0.03 suggests that the explanatory power of this model is fairly low.

Assessing team performativity with DuoTest. We assign a binary variable for each student groups.

Table 4: Exa02 as a function of individual exam (model 01) and team transactivity (model 02)

Variable	M1: Individual	M2: Group
Intercept	0.00 (1.00)	-2.05 (0.00) ***
Exa01	0.22 (0.08)	0.06 (0.418)
Group 01		0.46 (0.243)
Group 02		1.69 (0.000) ***
Group 03		**2.63** (0.000) ***
Group 04		1.89 (0.000) ***
Group 05		**2.48** (0.000) ***
Group 06		**2.61** (0.000) ***
Group 07		**2.24** (0.000) ***
Group 08		**2.09** (0.000) ***
Group 10		1.57 (0.000) ***
Group 11		**2.33** (0.000) ***
Group 12		**2.47** (0.000) ***
Group 13		1.11 (0.006) ***
Group 14		**2.66** (0.000) ***
Group 15		**3.25** (0.000) ***
Group 16		**2.97** (0.000) ***
Adjusted R2	*0.03*	*0.71*

For example, the group 01 will have 1 in a column called G01. Since the group G09 seems to have the worst performance in Figure 3, it will have 0 for each group variable and it will be treated as baseline. Such baseline allows us to add only 15 variables for the 16 groups. To assess the performance of each team, we look at the coefficients of each group, which mitigates the negative effect of the intercept. Table 4 shows that the group effect is always above 0.46 (almost eight times the standardized effect of Exa01 = 0.06).

The coefficient of each team allows to distinguish positive/negative effects that were common among all teams (such the learning effect due to the fact of doing again the same exam, the advantage of switching from close book to open book, and the challenge of working in teams Vs working alone) and what happened in each team.

Indeed, we can see that the coefficient of some groups compensates for the value of the intercept (-2.05). For example, G08 has a coefficient equal to 2.09; hence, on average, the team members of Group 08 had slightly higher scores in the second exam. Groups with positive effects in Table 4 have coefficients in bold.

The resulting model 02 (M2) has a very good Adjusted R2 (0.71). In M2, the coefficient of Exa01 (0.06) is not statistically significant anymore (p = 0.418), leading us to **confirm hypothesis H2** and to affirm that *the team effect (transactivity) increases the explanatory power of our model*. Indeed, one could assume that the increase in the value of the R2 would be the consequence of using more variables; but the Adjusted R2 automatically adjusts the R2 of the model to take this effect into account. Moreover, the regression diagnostics in Appendix B does not indicate any further issues. Another way to read Table 4 is to read the p-values associated with the coefficients of the groups. With the exception of G01, the probability that the coefficient of each team would be the same as G09 (the baseline) is almost none, leading us to reject the null hypothesis in our case, which states that there is no group effect and that it is all random.

Nonetheless, the analysis of the coefficients shows that **we cannot confirm nor reject hypothesis H3**, which state that *the team has a positive effect on the individual performance*. The quantitative analysis rejoins the insights already described by Table 2 and shown in Figure 3: there are four possible scenarios. Teams with low transactivity have a grey coefficient in Table 4, and describes situations where (a) students with wrong answers did not find right answers (G09), or (b) students with wrong answers convinced students with right answers to switch in the wrong direction (G04, G10, G13). Teams with good transactivity are associated with (c) students having the right answers and convincing their colleagues to switch in the right direction (G14 and G15) or (d) students with wrong answers working together to find the right answer (G16). This leads us to underline the need shown in the literature review of a detailed analysis of the effect of each individual on the team (Team←Individual) illustrated in Figure 3 and the effect of the team on each individual (T→I) shown in Table 4.

A final remark should be done for G02, and its surprising negative coefficient in Table 4. Figure 3 shows that the score Exa02 on average increased from Exa01. Nonetheless, the quantitative analysis shows that students of group G02, who got the best Exa01 results, are those who got the worse Exa02 results afterwards. Indeed, this can be visually noticed as well, if one looks attentively at the colors of the dots for G02 the Figure 3a and 3b. Once this situation is acknowledged, it is possible to notice the same trend in G03, G06, G08 and G12: the students with the highest Exa01 scores that pushed up the scores of his/her colleagues were not the one with the best scores in the second exam.

7. Discussions and conclusions

This article started by using the metaphor of the Roman Testudo to describe how students learn to cooperate in order to deal with problems in their future careers. Our study suggests that what seems to be a single phenomenon (team performance) is in reality composed of assorted heterogeneous elements [4]: *team health*, which depends on each team member (T←I), and *transactivity*, which influences the future performance of each team member (T→I) and that we called "*the omitted variable*" in the title of the article.

Accordingly, we wanted to look for new ways to design a final exam to assess individual and team performance in a *team-based learning* (TBL) course. Such objective is relevant and persisting in the field of study of information systems, since TBL is increasingly used to teach university students how to work together and solve complex problems in a growing number of fields, and we were missing of a structured and simple way to perform summative assessment. Moreover, many experts agree that the current situation concerning Covid-19 will speed up existing trends in digital tools for education, and we believe that the assessment online of team-based learning will be one of them. Nonetheless, since our test took place at the beginning of 2020, the data collected describes a situation before Covid-19 and new data is required to update the model.

Clarifying our initial assumptions. Our initial assumption is that teams have a positive effect on the learning experience. Moreover, our approach might be biased towards TBL as a form of teaching. Nonetheless, our intent is to bridge forms of experiential learning with classic testing techniques such as written exams. The screening process shown in section 2 illustrates how we have selected and reviewed previous works from the fields of team-based learning, project-based learning, and software solution to assess students. Even though, such works are complementary, a paper that combines these three views to develop an artefact is missing.

There is not a simple way to use immediate feedback assessment. The research gap identified in section 02 concerned the lack of a simple tool to assess individual to team (T←I) and team to individual (T→I) effects. Therefore, we have decided to create a theory of design and action [6], which explains how to do something and gives explicit prescriptions for teachers to construct a new type of final test for TBL classes, which we called DuoTest. Our preliminary findings show promising results that needs to be replicated in other classes and other topics. That will allow to take into account the changes in the pedagogical scenario due to Covid-19.

Next steps. So far, DuoTest extends existing solutions for immediate impact assessments [14], since it allows to obtain deeper insights at a fraction of their cost. Nonetheless, this assessment tool is only as good as the team-based learning environment where the course is situated. Some teachers might be concerned about the effort required to setup the overall system, but it might end requiring less effort than a standard class. That is why, in our future work, we will (a) present the longitudinal data collected in the 13 weeks before the exam, (b) describe how to predict the level of transactivity of each team and (c) reflect on reducing the number of summative tests required in a course.

One solution suggested by one of the reviewers of this paper concerned the possibility to conduct team-based learning tasks by formative assessments throughout a course; that option is currently being tested. Another remark made by another reviewer concerned the possibility to quantitatively assess the team health with our approach: it is already possible to visually assess the concentration of scores for each team, but we intend to assess if there is a correlation between the variance of Exa02 scores in a team and the 18 dimensions of O'Neill et al. [23].

Acknowledgements: *this work was supported by the fund of the University of Applied Sciences and Arts Western Switzerland (HES-SO) for pedagogical research.*

The authors would like to thank Hervé Barras, Pierpaolo Maglietta Pollari and the two anonymous HICSS reviewers for their feedbacks and useful insights.

A previous version of this paper has been discussed at the e-Bled conference 2020 [1]. *Accordingly, the data analysis of the coefficients has been improved, the overall framework of testudo has been described and the description of the different levels of transactivity has been included as a possible explanation for hypothesis 3.*

8. References

[1] Bonazzi, R., and Y. Rouiller, “The Omitted Variable: Could DuoTest Enable a New Way to Assess Team Performance in Team-Based Learning?”, *Bled 2020 Proceedings*, 1–17.

[2] Brazhkin, V., and H. Zimmerman, “Students’ Perceptions of Learning in an Online Multiround Business Simulation Game: What Can We Learn from Them?”, *Decision Sciences Journal of Innovative Education 17*(4), 2019, pp. 363–386.

[3] Christensen, J., J.L. Harrison, J. Hollindale, and K. Wood, “Implementing team-based learning (TBL) in accounting courses”, *Accounting Education 28*(2), 2019, pp. 195–219.

[4] Davis, M.S., “That’s interesting! Towards a phenomenology of sociology and a sociology of phenomenology”, *Philosophy of the social sciences 1*(2), 1971, pp. 309–344.

[5] Edmondson, A.C., *Teaming: How organizations learn, innovate, and compete in the knowledge economy*, John Wiley & Sons, 2012.

[6] Gregor, S., “The nature of theory in information systems”, *MIS quarterly*, 2006, pp. 611–642.

[7] Hake, R.R., “Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses”, *American journal of Physics 66*(1), 1998, pp. 64–74.

[8] Hevner, A., S. March, J. Park, and S. Ram, “Design Science in Information Systems Research”, *Management Information Systems Quarterly 28*(1), 2004, pp. 75–105.

[9] Hijzen, D., M. Boekaerts, and P. Vedder, “Exploring the links between students’ engagement in cooperative learning, their goal preferences and appraisals of instructional conditions in the classroom”, *Learning and Instruction 17*(6), 2007, pp. 673–687.

[10] Kolb, D.A., *Experiential learning: Experience as the source of learning and development*, Pearson Education, Inc., Upper Saddle River, New Jersey, 2015.

[11] Lea, M.R., and B.V. Street, “The" academic literacies" model: Theory and applications”, *Theory into practice 45*(4), 2006, pp. 368–377.

[12] March, S.T., and G.F. Smith, “Design and natural science research on information technology”, *Decision support systems 15*(4), 1995, pp. 251–266.

[13] Martin-Rios, C., and N. Erhardt, *Organizational Design Simulation: Evolving Structures*, Harvard Business School Publishing, Boston MA, 2019.

[14] Maurer, T.W., and J.J. Kropp, “The Impact of the Immediate Feedback Assessment Technique on Course Evaluations”, *Teaching & Learning Inquiry: The ISSOTL Journal 3*(1), 2015, pp. 31–46.

[15] Mazur, E., *Peer instruction: A user’s manual*, American Association of Physics Teachers, 1999.

[16] McKeachie, W.J., and M. Svinicki, “How to make lectures more effective”, *Teaching tips: Strategies, research, and theory for college and university teachers. New York: Houghton Mifflin*, 2006, pp. 57–73.

[17] Michaelsen, L.K., A.B. Knight, and L.D. Fink, “Team-based learning: A transformative use of small groups in college teaching”, 2004.

[18] Moodle, “Quiz activity - MoodleDocs”, 2019. https://docs.moodle.org/38/en/Quiz_activity

[19] Moodle, “Grader report - MoodleDocs”, 2019. https://docs.moodle.org/38/en/Grader_report

[20] Moodle, “Workshop activity - MoodleDocs”, 2019. https://docs.moodle.org/38/en/Workshop_activity

[21] Moodle, “Lesson activity - MoodleDocs”, 2019. https://docs.moodle.org/38/en/Lesson_activity

[22] Moodle, “Short-Answer question type - MoodleDocs”, 2020. https://docs.moodle.org/38/en/Short-Answer_question_type

[23] O’Neill, T.A., L. Pezer, L. Solis, et al., “Team dynamics feedback for post-secondary student learning teams: introducing the ‘Bare CARE’ assessment and report”, *Assessment & Evaluation in Higher Education*, 2020, pp. 1–15.

[24] Peffers, K., T. Tuunanen, M.A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research”, *Journal of management information systems 24*(3), 2007, pp. 45–77.

[25] Polybius, “Histories, Book 28”, In *Greek texts and facing English translation*. Harvard University Press, 1922.

[26] Springer, L., M.E. Stanne, and S.S. Donovan, “Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis”, *Review of educational research 69*(1), 1999, pp. 21–51.

[27] Strijbos, J.-W., “Assessment of (computer-supported) collaborative learning”, *IEEE transactions on learning technologies 4*(1), 2010, pp. 59–73.

[28] Swanson, E., L.V. McCulley, D.J. Osman, N. Scammacca Lewis, and M. Solis, “The effect of team-based learning on content knowledge: A meta-analysis”, *Active learning in higher education 20*(1), 2019, pp. 39–50.

[29] Thomas, J.W., “A Review of Research on Project-based Learning”, 2000, pp. 49.

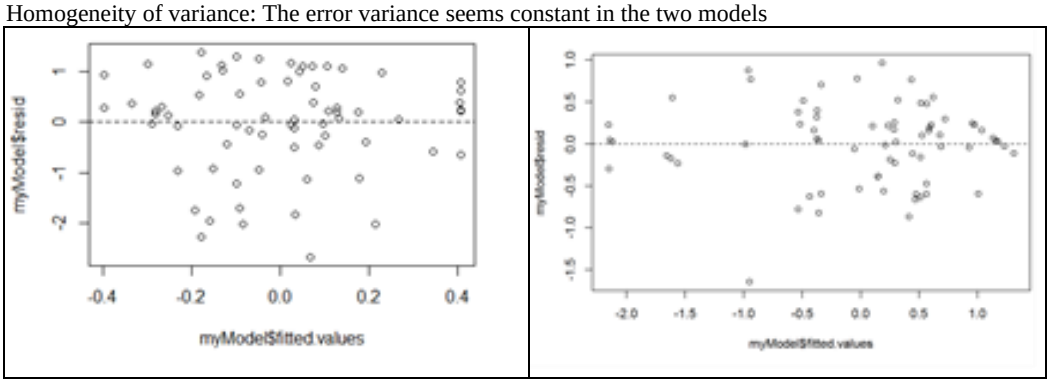
[30] Weinberger, A., K. Stegmann, and F. Fischer, “Knowledge convergence in collaborative learning: Concepts and assessment”, *Learning and instruction 17*(4), 2007, pp. 416–426.

Appendix A: Complete dataset with the raw data

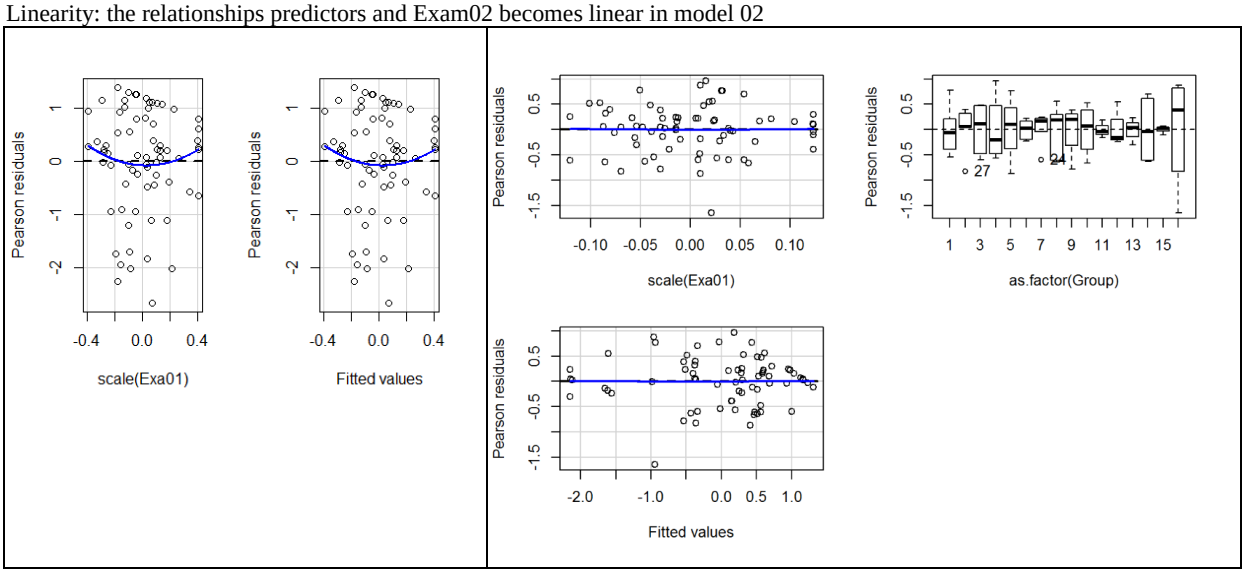
UID	Group	Class	Exa01	Exa02
1	1.00	2.00	5.50	8.48
2	10.00	3.00	8.54	8.39
3	8.00	1.00	7.34	9.57
4	12.00	3.00	7.64	7.39
5	16.00	3.00	6.89	7.47
6	9.00	3.00	6.54	8.44
7	3.00	1.00	7.39	8.57
8	16.00	3.00	7.69	5.62
9	7.00	2.00	6.90	10.00
10	2.00	1.00	5.25	8.25
11	11.00	3.00	7.77	9.52
12	8.00	1.00	10.00	9.79
13	7.00	2.00	8.79	10.00
14	14.00	2.00	4.94	8.64
15	7.00	2.00	7.39	10.00
16	5.00	2.00	7.44	8.09
17	6.00	3.00	6.60	9.14
~~18~~	~~4.00~~	~~2.00~~	~~4.00~~	~~4.00~~
19	16.00	3.00	7.92	8.42
20	15.00	3.00	6.89	10.00
21	2.00	1.00	5.40	8.65
22	13.00	1.00	6.64	6.19
23	15.00	3.00	8.18	10.00
24	14.00	2.00	8.43	9.04
25	7.00	2.00	8.09	9.09
26	9.00	3.00	6.55	7.10
27	9.00	3.00	9.58	8.34
28	2.00	1.00	5.65	7.25
29	1.00	2.00	6.08	9.48
30	15.00	3.00	6.55	10.00
31	12.00	3.00	7.45	6.54
32	4.00	2.00	8.12	8.84
33	8.00	1.00	5.29	8.47
34	14.00	2.00	6.19	7.39
35	8.00	1.00	7.72	9.97
36	12.00	3.00	8.69	6.54
37	7.00	2.00	6.34	9.64
38	6.00	3.00	7.74	9.14
39	10.00	3.00	5.18	9.59
40	3.00	1.00	6.32	9.77
41	4.00	2.00	7.57	9.94
42	12.00	3.00	6.59	6.54
43	1.00	2.00	6.40	7.98
44	5.00	2.00	7.95	10.00
45	14.00	2.00	8.44	7.54
46	6.00	3.00	7.88	8.69
47	15.00	3.00	10.00	10.00
48	3.00	1.00	7.43	8.72
49	15.00	3.00	6.00	10.00
50	8.00	1.00	4.50	8.47
51	1.00	2.00	9.05	8.98
52	4.00	2.00	7.82	8.19
53	2.00	1.00	5.65	8.25
54	5.00	2.00	10.00	9.34
55	4.00	2.00	6.74	8.34
56	9.00	3.00	6.95	8.29
57	10.00	3.00	7.97	8.99
58	11.00	3.00	5.97	9.02
59	13.00	1.00	5.90	6.39
60	1.00	2.00	10.00	8.33
61	16.00	3.00	7.44	8.52
62	10.00	3.00	4.50	9.24
63	6.00	3.00	8.09	8.99
64	13.00	1.00	6.00	5.79
65	3.00	1.00	7.52	9.82
66	3.00	1.00	9.99	9.52
67	2.00	1.00	5.30	8.55
68	11.00	3.00	10.00	9.37
69	8.00	1.00	6.92	9.47
70	6.00	3.00	6.99	8.69
71	13.00	1.00	6.14	6.19

Appendix B: Regression diagnostic for model 01 (left) and model 02 (right)

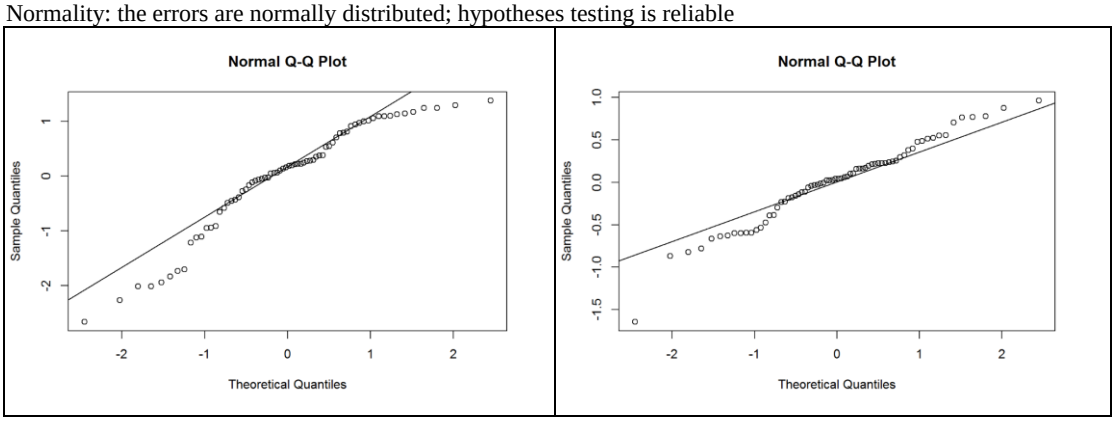
Homogeneity of variance: The error variance seems constant in the two models



Linearity: the relationships predictors and Exam02 becomes linear in model 02



Normality: the errors are normally distributed; hypotheses testing is reliable



Multicollinearity of the second model: when VIF > 10 a variable merits further investigation

	VIF	Df	GVIF^(1/(2*Df))
scale(Exa01)	1.32	1	1.15
as.factor(Group)	1.32	15	1.01