IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 13, 2021, accepted May 9, 2021, date of publication May 24, 2021, date of current version June 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3083238

How Do Students Behave in a Gamified Course?–A Ten-Year Study

AMIR HOSSEIN NABIZADEH, JOAQUIM JORGE, (Senior Member, IEEE), SANDRA GAMA, AND DANIEL GONÇALVES
INESC-ID, Instituto Superior Técnico, University of Lisbon, 1649-004 Lisbon, Portugal

Corresponding author: Amir Hossein Nabizadeh (amir.nabizadeh@tecnico.ulisboa.pt)

This work was supported in part by the National Funds through the Fundação para a Ciência e a Tecnologia (FCT) under Project UIDB/50021/2020, and in part by the Project GameCourse, Portugal, under Grant PTDC/CCI-CIF/30754/2017.

ABSTRACT Gamified learning aims to motivate students using game elements. Although gamification can enhance students' enjoyment and engagement, it is unclear how different students behave in and interact with gamified contexts. To this end, we analyze how different students interact with a gamified course. We devised such an experimental course on Multimedia Content Production (MCP), and ran it for ten years. At each year, we modified it after students' feedback from the previous year. We determined student groups applying clustering techniques to learner performance data, independently analyzed the resulting clusters in terms of behavior, engagement, performance, and also compared those pairwise. Our analysis identified four different student groups (profiles/clusters) according to their performance and interactions with the course across all years. We found out that the best performing students were those that had significantly more interactions with course materials and consistently ranked highest. In addition, we found that performance indicators for students of all groups became stable within the first month after course start, allowing final grades to be predicted with high accuracy by then. Furthermore, all were deadline driven and became mainly active at the end of the semesters (indicating a lack of self-regulation skills). Moreover, we did not find any specific relation between students' groups and gaming profiles (Brainhex categories). Finally, we propose practical implications and guidelines for designing compelling gamified learning experiences.

INDEX TERMS Brainhex categories, gamification, gameful learning, student behavior, clustering.

I. INTRODUCTION

Gamification applies game design elements to non-game contexts [28] to increase users' enjoyment, and subsequently boost their motivation and enhance their engagement within a gamified context [14], [44], [64]. Gamification has been used in various contexts with different objectives, ranging from raising health awareness [17], teaching how to drive [33], to improving engagement with a course [10], [48], and enhancing enterprise risk management Education is one context that is widely explored by researchers.

Despite innovative educational approaches, the traditional and current educational methods are often considered boring and ineffective by students [29]. Hence, the main challenges of these methods are about enhancing students' motivation and engagement within a course [44]. Alternatively, gamified educational approaches attempt to tackle these challenges. These have a noticeable motivational power since they use different mechanisms to inspire students by providing joyful and fun game-playing. Game-like mechanisms encourage students to engage more with courses and perform different learning activities to earn more rewards (e.g., points, levels, experience points). Doing different learning tasks boosts students' learning and enhances their critical, collaboration, communication, and problem-solving skills [29].

Although gamification has shown promising results and studies demonstrate that it can enhance student performance, [32] and engagement [27], it is still not well explored how different students adjust to a gamified course and interact with it. Therefore, we intend to answer the following questions in our study:

RQ1 How many groups of students engage in a gamified course?
RQ2 How do students of different groups collect experience points (XP), and badges?
RQ3 How do these groups differ, considering average ranks and final grades?

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

RQ4 Are students deadline-driven regardless of group, or are they steadily engaged with the course throughout the semester?

RQ5 When does the performance of various groups become stable?

RQ6 Is there a specific relation among different student groups and their gaming profiles (Brainhex categories [55])?

RQ7 How does the ability to self-regulate vary between different student groups?

Furthermore, while other researchers evaluate the effectiveness of their gamified courses in a single iteration, they fail to assess the consistency of obtaining similar results over the years.

In this paper, we present a longitudinal study on a gamified course over a period of ten years: Multimedia Content Production (MCP). This is a MSc-level course, offered to students of the Computer Science Engineering MSc programme at Instituto Superior Técnico, the Engineering school of the University of Lisbon. It focuses on the digital representation and manipulation of different types of media (images, audio, video, etc.), but also on their creation and editing. It was gamified for the first time in the 2010-2011 school year, with the inclusion of game elements such as a leaderboard, experience points and levels, and achievements. The course was thereafter revised annually, based on an analysis of the pedagogical outcomes from previous instances and student feedback. This led, over the years, to the inclusion and removal of other game elements (quests, a skill tree, etc.) and, more importantly, it gave us thorough and longitudinal insights on how different students react to a gamified learning experience.

In order to analyze how different students performed and interacted with the course, we applied cluster analysis at the end of each academic year. We also conducted a cluster-wise assessment to evaluate student behavior, engagement, and performance within each cluster independently. We later compared each cluster to other clusters of the same and other years. According to the students' performance and behavior, we could mainly identify four student groups across all years. Our main contributions are as follows:

- Distinguishing and describing different students groups considering their performance and behavior.
- Analyzing the achievements, performance, engagement, interactions and characteristics of each group and compare it to others.
- Identifying how early we can predict each student' group and performance within a reliable accuracy.
- Analyzing whether student gaming profiles (Brainhex categories) have any specific relation with their groups.
- Suggesting substantial instructions for designing a gamified course.

The remainder of this paper is structured as follows: Section 2 introduces related work. Section 3 explains the research methodology of this study. Section 4 presents the elements of our gamified course as well as its evolution through the years. Our method to identify the students' groups and the data used for that are detailed in Section 5. In Section 6, the groups' behavior was analyzed and compared. Section 7 highlights some indications to design a more effective gamified course, while Section 8 presents our conclusion.

II. RELATED WORK

Different educational methods are introduced to support students and improve their learning activities toward their success. In that regard, several approaches aimed at enhancing students' performance and engagement with courses using the gameful learning. In [41], authors intended to boost the students' motivation and engagement by gamifying a Moodle-based course about Unified Modeling Language (UML). This course was organized in ten levels considering its original curriculum. Levels were classified into syntax and semantic levels. Students were rewarded 100 and 200 points for completing tasks within syntax and semantics levels, respectively. The accumulated points of students were used to estimate their levels, and a level could be completed by obtaining the minimum required points for that level. Whenever a student could complete a level, he/she could get a new badge. The effectiveness of this course was evaluated in a single semester with 22 students in autumn 2017. The results showed that the gameful learning was successful and students were satisfied with the course and found it useful.

In another paper, authors studied the experience of undergraduate students' with a gamified course about social networking technologies [23]. This course was already developed and in-used in a public university in the Northwest United States, and was designed for freshman students with no background on the subject. It included Experience Points (XP), a leaderboard, badges, and levels. In this course, all students' activities were assigned a certain XP, and students were graded using their accumulated XP throughout the course. They leveled up by accomplishing a certain amount of XP and got minor rewards by collecting badges. The top 10 students (i.e. got the highest XP) were listed on the leaderboard. For the evaluation, a survey was carried out with 139 students. It was conducted at three times during the course (beginning-middle-end). The results confirmed that the students had a positive influence about gameful learning on their achievements, learning, and engagement with the course.

Aleksic-Maslac *et al.* used a tool called Kahoot [26] to create fun quizzes for an ICT course at the Zagreb School of Economics and Management (ZSEM) [2]. During a semester, six Kahoot quizzes were conducted, and each student got a total point by answering the questions of each quiz. At the end of the semester, both professors and students were asked to state their level of satisfaction. The results showed that using Kahoot was an efficient method to motivate students for higher engagement.

Another gamified course experiment was presented in [16]. This course was a short one (lasted one month) about promoting entrepreneurship among B.Sc., M.Sc., and Ph.D. students in the field of Electronic Engineering at the University of

Genoa. It was relied on using three Serious Games (SG) [1] for business, which were: Hos Shot Business (HSB), Enterprise Game (EG), and SimVenture (SV). A range of topics was covered in this course having different level of difficulties in games, game-play, and home assignments. Like other studies, the students' ranks were determined considering their obtained cumulative scores. To evaluate this course, 34 students were divided into 16 teams. The results highlighted that their overall engagement with the course was high. Also, teachers had positive opinions about course acceptance and knowledge acquisition of the students.

The same authors conducted another gamified course using similar methodology two years later. This course was to stimulate entrepreneurial and innovative mindsets of B.Sc, M.Sc, and PH.D. students of non-business faculties, and to provide them with several significant operational and theoretical skills [3]. Here, the gameful learning was used to sustain students' motivation, and consequently, enhance their effectiveness in learning activities. During this course, students were engaged in a variety of learning activities and played with several SG. They received various topics with different level of difficulties in games, game-play, and home assignments. This course was carried out in Italy, Spain, and Netherlands, and it was concluded that a single game cannot be used to achieve all the gameful learning objectives (e.g. enhancing engagement, motivation, and performance). For that, more than one single game needs to be adopted in each course.

The effect of gamification on students' engagement and performance was studied in [21]. This study included the results of participants in a one-term ICT course enrolled at a school of education. Participants were interviewed to examine the relationships among gamification, engagement, and achievement. The results presented that gamified elements had a positive motivational impact on engagement and indirectly affected the academic achievement. Another study examined participants' willingness to join gamified activities where rewards were not directly tied to a course's grades [4]. For that, over two semesters, an optional gaming activity was added in five sections of a course (experimental group), and four sections acted as a control group. The findings (collected by pre-, mid-, and post-surveys) presented an important difference between experimental and control groups regarding hours spent on gamified activities. In [22], authors analyzed the effect of gamification on students of a management classroom. In this study, 44 elementary students attended a gamified management classroom (experimental group) and 42 attended a traditional classroom (control group). Then, the differences between the experimental and control groups were examined regarding their divergent thinking and creative tendency. The results showed that the verbal divergent thinking and creativity of the experimental group were enhanced in comparison with the control group. In [25], authors assessed the students' perception of the impact of badges and leaderboards in their motivation towards an introductory Software Engineering course. Hence, a survey with 18 participants was conducted for a quantitative evaluation, and a series of interviews with six participants was performed for a qualitative assessment. The results indicated that students were positive about badges, but had mixed feedback about the use of leaderboards.

An immersive gamified course was explained in [40]. It was presented in the academic year 2013-14, and was designed for twelve sessions. In each session, students (space school trainees) should activate one of the spaceship systems. Depending on the complexity of a system, it could include different number of missions. More complex systems had more short missions to complete. Completing these missions resulted in getting badges, and sessions could be passed by concluding missions and getting a certain amount of points. These points could be also obtained through assignments. The assignments were optional and students could select them as they like (harder assignments had more points). In this study, all missions, assignments, and achievements were transferred to Youtopia platform, which was designed for motivating and engaging students [66].

Pan *et al.* developed gamified forensic modules with intuitive designs and interactive dialogues [57]. They were designed for students with no related background to learn fundamental digital forensic content and explore the forensics procedures and technologies using interactive games. Here, the gameful learning was applied to keep students interested and engaged with the modules. Modules were implemented in a real computing environment with access to real forensics tools and evidence for enabling students to practice with forensics technologies. By the time of this paper, these modules were still un-evaluated. The authors intended to measure the effectiveness of their modules in the summer and fall semesters of 2015 at their institution and two other partners' colleges.

Another gamified e-learning course was introduced in [35]. It was designed based on a structural gameful learning [42], [53] for the bachelor's degree at the University of Plovdiv. The structural gameful learning was obtained by using various game elements. This course was arranged for ten weeks, and each week was corresponding to a level. Levels included quests and assignments (seven individual and one group assignments). When students completed the assignments, they were awarded by points, and getting enough points would allow them to go to the next level. In the last level, students should answer to a Constructivist On-Line Learning Environment Survey (COLLES) [6]. To analyze the effectiveness of the gamified course, it was presented together with a standard e-learning course, and the students were allowed to select their preferred one. Out of 113 students, 41 of them selected the gamified course and only 27 could complete it. Results showed that the students, who used the gamified course, were less confused and had higher grades and engagement with the course than the ones using the standard e-learning course. In addition, the understandability of students that used the gamified course was more than their expectation.

Dicheva *et al.* introduced a gamified course to enhance students' engagement with the course and encouraging them to do self-study [31]. It was designed for a data structure subject, and was using a OneUp learning platform [30], [50], [60]. This platform supports the application of game design principles and elements for a course. In this gamified course, the students were awarded using badges and a virtual currency. Badges were used to reward students for their performance while the virtual currency was used to reward their engagement with the course (e.g. attending classes). A quasi-experiment [20] was conducted for the evaluation. For that, 16 students (from fall 2017) formed the control group while 11 students (from spring 2018) shaped the experimental group. Both groups used the OneUp platform, but the gameful learning features were only activated for the experimental group. The log data of this platform together with the students' final grades were used to analyze their interactions and performance. In addition, a survey was conducted among the students in the experimental group to assess the usefulness of the gamified course. The results showed that the students found the gamified course useful while it increased their engagement and reduced their failing rates.

Barata *et al.* also proposed a gamified M.Sc. course to improve students' engagement and motivation [7]. It included six main game elements, which were, experience points (XP), a leaderboard, levels, Badges, challenges, and a skill tree. The skill tree was a precedence tree, where each node was a learning task that resulted in XP upon completion. Initially, six nodes were unlocked. Posterior nodes could be accessed if the anterior ones were completed. Students' behavior and interactions with the course was analyzed, and authors could distinguish four group of students. The evaluation results showed that the course enhanced students' performance and participation.

Although many studies shown that the gameful learning had positive impact on the students' motivation and engagement with the course, a few studies presented that the gameful learning had no or even negative influence on the students. For instance, in [38], students were analyzed to assess how gameful learning influenced their course engagement and behavior. To this end, their performance, motivation, and satisfaction were measured four times during a 16-week semester. For the final evaluation, the students across two courses were tested. One course was a gamified one using badges and a leaderboard while the other one was a non-gamified one (having similar curriculum). The results showed that the students using the gamified course were less satisfied, motivated and engaged with the course. In addition, these students had lower final grades in comparison with the ones using the non-gamified course. In [37], gamification was examined in relation to achievement goal orientation (students' preferences to various goals, outcomes, and rewards). For this purpose, achievement badges were added to a Data Structures and Algorithms course (N = 278), and students' feedback with various achievement goal orientation profiles was assessed. Moreover, the authors analyzed how students (most motivated by badges) differ from others in terms of behavior and achievement goal orientation. The results indicated no important differences in that regard. In [24] and [32], authors presented mixed results. These studies used social networking and gameful learning in an undergraduate course, and analyzed their influence on the students' behavior, engagement, and performance. The evaluation results stated that the students using this undergraduate course performed better in completing their assignments than the ones using a traditional e-learning course, but the students using the traditional course could learn more. In addition, students' engagement and grades remained low using the gamified course.

Regardless of having negative or positive impact on the students' motivation, engagement, and performance, only a few studies investigated how different students behaved in a gamified learning environment. For example, in [19] authors examined the influence of different personalities traits and learning styles on students' engagement, perception, and performance on the course. Another instance is [46] where Lavoué *et al.* assessed whether the adaption of gaming features based on a player model enhance students' participation and motivation. This experiment was conducted once using 266 participants. One year later, Mbabu studied how students having different personality and learning styles interacted with an adaptive gameful learning tool that was designed for an e-learning platform [47]. This study was also evaluated using 158 students within a few weeks.

Besides the minority of studies that analyzed how students behaved in a gamified environment, they also often evaluated their courses in a single semester with a few students while obtaining consistent results over the years with a large number of students is ignored.

III. RESEARCH METHODOLOGY

To accomplish the main goal of this study, which is understanding how students of different groups behave in a gamified course, we have applied the Educational Design Research (EDR) approach [59] along with a formative study. EDR is an iterative development of solutions for practical educational issues. These solutions can be educational products, policies, or processes. EDR not only tries to solve important issues facing educational practitioners, but it simultaneously discovers new knowledge that can inform the work of others facing similar issues. Besides EDR, formative studies [36] were applied to gather data that could be helpful to understand students' behavior, diagnose problems, and improve the course design.

Using the mentioned research approaches, we annually collected students' interaction logs (e.g. what learning materials were seen by a student, how many posts he/she made, etc.) with a Learning Management System (LMS) used in the course (Moodle[1]). We also gathered students' performance data (e.g. obtained XP, badges, etc.) from our gamified

[1] www.moodle.org

TABLE 1. Game elements and changes over the years.

		2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020
Game Elements	**Leaderboard**	√	√	√	√	√	√	√	√	√	√
	Achievements	√	√	√	√	√	√	√	√	√	√
	Collab. Badges		√	√	√						
	Bragging					√	√	√	√	√	√
	XP	√	√	√	√	√	√	√	√	√	√
	Levels	√	√	√	√	√	√	√	√	√	√
	Skill Tree		√	√	√	√	√	√	√	√	√
	Quests			√	√	√	√	√			
	AvatarWorld				√	√					
	Skill Points		√								
Other	**Exam/Quizzes**	E	E	E	Q	Q	Q	Q	Q	Q	Q
	Total Students	35	56	58	84	68	95	95	135	121	99
	Num. Dropouts	0	4	4	8	7	10	5	18	12	15
	Num. Students	35	52	54	76	61	85	90	117	109	84
	Male/Female	M:32 , F:3	M:42 , F:10	M:41 , F:13	M:67 , F:9	M:53 , F:8	M:71 , F:14	M:71 , F:19	M:94 , F:23	M:77 , F:32	M:60 , F:24

course. Every year, the mentioned sets of data were collected from the enrolled students in the course (statistics are mentioned in Table 1). These datasets were analyzed using the R programming language at the end of each semester.

To analyze the collected data, we initially used an elbow technique together with a k-means clustering algorithm to identify the number of groups among students and distinguish their performance using their average accumulated XP over a semester. Then, we described the groups and used cluster analysis to compare them in terms of obtained badges, XP, final grades, and ranks. Their engagement and interaction with the course were then analyzed using their XP and logs data. At first, these sets of data were converted to binary sets (i.e. if a group member did an action or collected an XP in a day, it was considered as one, otherwise zero). These binary sets were then presented in form of density and scatter graphs to show groups' engagement and interaction with the course. Next, the stability of groups' performance was assessed using their ranks (calculated using average accumulated XP) over a semester. To evaluate whether groups had any relation with gaming profiles, we initially determined gaming profiles of groups' members using brainhex categories (questionnaires) and then examined differences between each pair of groups using a statistical hypothesis test (p-values). Finally, the self-regulation skills of students were assessed using the sum of XP and badges that they got within a month of a semester and compare them with other months from the same semester.

Besides the aforementioned data and analysis, we also collected students' feedback at the end of each semester using questionnaires. This feedback could be divided in two parts. The first part included students' opinions about the course itself, such as if it was engaging, motivating, creative, etc. The second part referred to the quality of the course elements, such as if they were interesting, or contributed for a greater workload.

After assessing how much we were successful in engaging, motivating, and providing enjoyment for students,

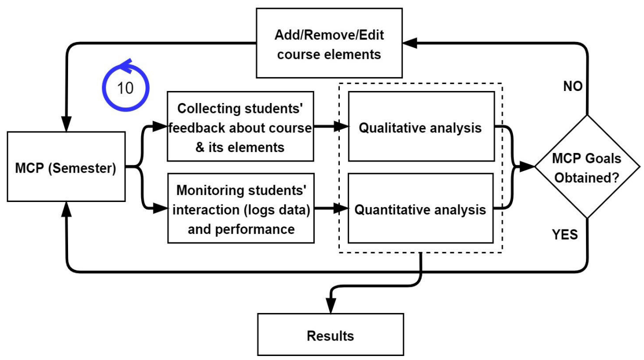


FIGURE 1. Research methodology.

we modified the course and added/removed/edited the course elements for the next iteration. After ten iterations of the course, we reached a set of design implications that can be helpful in the setting up of new gamified courses. They are mentioned in Section VII. A general view of our research methodology is presented in Figure 1.

IV. GAMIFIED COURSE

In this study, we gamified a course named Multimedia Content Production (MCP), which is designed for MSc students in the field of Information Systems and Computer Engineering. The main goal by gamifying the MCP was to provide enjoyment for the students to keep them motivated and engaged with the course. In MCP, students attend both theoretical lectures and practical labs. In theoretical lectures, students learn about different formats (audio, video, image, etc.) from an engineering standpoint (compression, formats, etc.) but also with an emphasis on the creation of aesthetic, impactful and high-quality media (in the Lab classes) [10]. Besides lectures and labs, students also join the discussions and complete online assignments via Moodle.

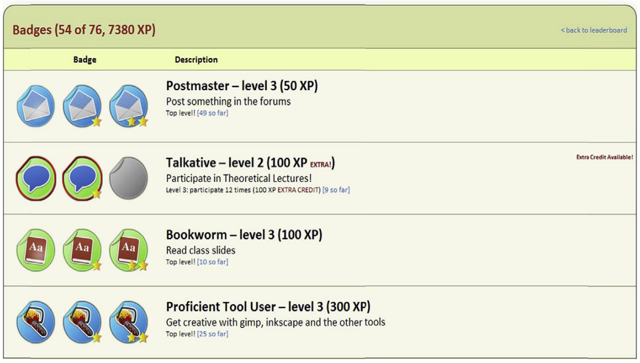


FIGURE 2. Badges and their levels.

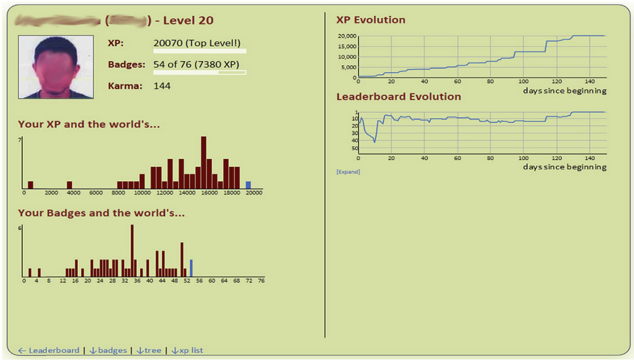


FIGURE 3. Dashboard.

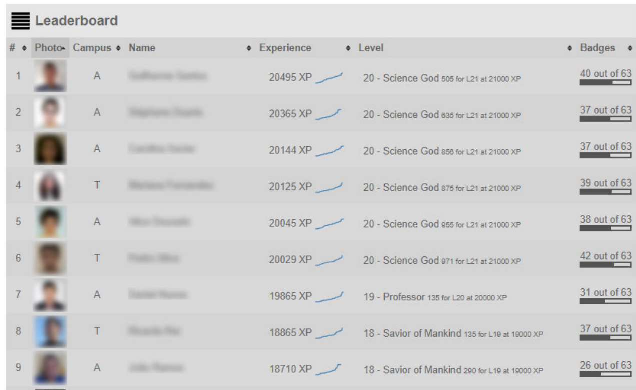


FIGURE 4. MCP leaderboard.

MCP is presented synchronized and identical across two university campuses. It is offered in English and is only available in the second semester of each academic year. This course has three sessions per week (each session lasts 1.5 hour), two theoretical lectures and one lab, and consists of various activities, such as quizzes, a multimedia presentation, lab assignments, and several other activities. Instead of receiving traditional grades, students earn XP for completing different course activities, including the traditional evaluation elements existing in the pre-gamified course (a multimedia presentation, lab assignments, final exam or quizzes) and the game elements. Performing specific course activities, such as attending lectures, finding bugs in class slides, or completing challenges, can result in obtaining badges. A badge can be achieved by completing an activity that might require a single iteration or up to three, with each iteration being worth a specific amount of XP and a badge (Badges are shown in Figure 2). Some of the badges award extra credit, as they reflect desirable behaviors but that cannot be mandatory due to the school's bylaws, and some award 0 XP and are there for bragging rights only. Getting a certain amount of XP results in achieving a new experience level. MCP includes 20 levels while 10 is the minimum level to pass the course. At the end of each semester, levels are converted to a 20-point grading system, which is the norm in our university. As means to communicate progress and provide feedback to students, a leaderboard (Figure 4) and a dashboard were used (Figure 3), which include students' XP, levels, and badges. For assessing students, MCP uses a quantum evaluation mechanism where students are awarded (by XP, badges, and levels) for performing any kind of course's activities (grade is granularized). Quantum grading allows students to transparently trace their progress whenever they participate in the course, down to the each XP. This is a marked departure from conventional so-called continuous evaluation schemes, which break exams and project assignments into smaller units in a rather discrete fashion.

This course has been modified annually considering the results that we got from analyzing the students' performance and interactions with the course and their feedback (collected by questionnaires) about the quality of the course and its elements. Depending on how much we were successful in achieving the MCP goals (enhancing students' enjoyment, motivation and engagement), the course was modified and its elements were added/removed/edited.

Various game elements are used during the years. The four main game elements are experience points (XP), levels, achievements and a leaderboard. In the first year, **achievements** were the only game element to obtain XP. For that, students needed to perform a set of tasks that we intended to encourage them to do, such as completing challenges, finding bugs in course slides, and attending lecturers. Besides getting XP, completing these achievements could result in getting badges. Some achievements awarded extra XP, which allowed students to obtain high grades without performing all mandatory tasks. There were $\approx$ 3000 extra XP but students could only earn up to 1000. This extra XP enabled us to reward desirable behaviors of the students that were not mandatory by the university's laws.

Another main element of the MCP is a **leaderboard** (an online webpage) that allows students to monitor their progress and compare it with others (Figure 4). This leaderboard also enables the students to clearly see what has been completed so far and what needs to be done. As presented in Table 1, it was used in all years without significant changes.

Skill Tree is a game element that was added to the course in the second year (2011-12). This element is designed to make the students more autonomous (select activities as they like). It is a precedence tree where each node refers to a learning activity that results in XP upon completion (Figure 5).

FIGURE 5. Skill tree.

Initially, five nodes are unlocked and subsequent nodes could be accessed if the anterior ones are completed [11]. In 2012-13, since students stated (in the satisfaction questionnaire) that the course needed more work than the other ones, we reduced 5% of total course XP from the final multimedia presentation and added it to the Skill Tree.

Skill Points were another game element that were only used in the second year. They were introduced to enhance the flexibility of students for selecting their preferred paths through the Skill Tree. Each type of course media (i.e. image, text, audio) had a different Skill Point. Skills of the Tree were unblocked by obtaining a certain number of points on related media. Anyhow, we were not successful in introducing these points due to lack of time management by students. As mentioned, these points were only used in 2011-12 and were dropped after that year. It needs to be noted that these points are different from XP (experience points), which can be obtained from various course activities. Also, these points are considered as XP in our analysis for 2011-12.

Quest was used for five years (2012-13 to 2016-17) to promote the collaboration toward a common goal among students. It was an online riddle where students should start by manipulating a multimedia content to find a URL for the next clue of the riddle. To enhance students' participation, they should contribute at least once to obtain the XP, and their contributions were posted in the forums and graded by the professors. Here, the amount of obtained XP was proportional to the accomplished Quest level and number of active participants. The Quest was dropped after five years since the students' feedback (ratings from 1 to 5) showed that, contrary to the first years in which it was used, they did not find it engaging ($\overline{X} = 3.07$, $\sigma = 0.47$) neither interesting ($\overline{X} = 3.22$, $\sigma = 0.41$) nor fun ($\overline{X} = 3.3$, $\sigma = 0.38$).

Finally, to boost the students' creativity and autonomy, and also to customize their learning experience using what they learned in the MCP classes, an **Avatar World** was added to the course for two years (2013-15). It was a 3D virtual world that evolved and grew by obtaining XP (emerging new buildings and characters). Students were represented by avatars, which could be customized or made by the techniques and tools introduced in the classes. Up to three percent of total XP could be obtained from it. Anyhow, Avatar World was not a successful attempt since it could not be well-integrated with the course and students had little to do there. In addition, we found relatively low levels of students' interest ($\overline{X} = 2.2$, $\sigma = 0.08$) and engagement ($\overline{X} = 2.12$, $\sigma = 0.03$) with it. To this end, it was removed from the course since 2015-16.

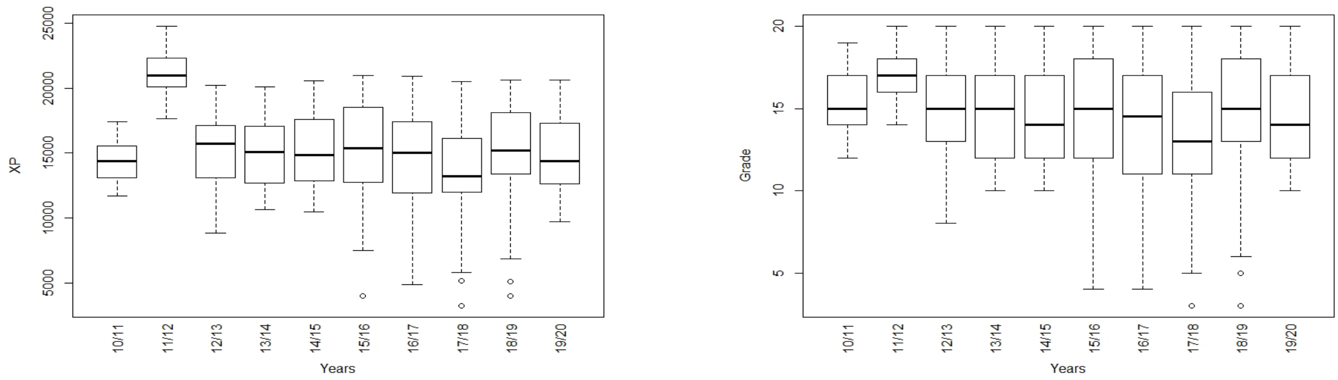
Table1 shows all of the game elements as well as the changes that are made into the course over the years. In this table, we also present the number of students who enrolled in the course and number of students who dropped out (withdrew from the course). Here, we see that a final exam is replaced with quizzes since 2013-14. This change was made since we noticed that several students had a low participation during a semester and became active only at the end of the semester, as the deadline looms. These quizzes allow students to assess themselves on a (almost) weekly basis and balance their participation through a semester.

Course modification and elements can directly influence the students' performance (e.g. XP and grades). In Figure 6, we show whether the students' XP and grades were changed over the year. We can see some variations through the years, especially since 2015-16. Up to this year, MCP had a rigid major\minor structure. So, to take MCP students needed to enroll in the Multimedia minor or major, and consequently, would be enrolled in other courses in the area. It implied that the students taking MCP were highly interested in it. Since 2015-16 and by restructuring the MCP, students could enroll in any course they like, and their reasons for that might vary, from deep interest to a perception of easiness. Figure 6 presents that the median of XP and grades were slightly decreased after 2015-16. Furthermore, the 25th percentile and minimum value were reduced more drastically, which could be due to the course restructuring. In 2011-12, XP and grades are noticeably higher than the other years. It happened since the initial version of the Skill Tree was added to the course without having any limitation on its XP. So, if the students worked more, they could get more XP from the Tree (and subsequently could get higher grades). To avoid this situation, after 2011-12, the amount of XP that students could get from the Tree was limited to a specific amount.

However, although the evolution of the course over the years caused meaningful performance changes among the students, we still do not know clearly how different students engaged with the course and interacted with it. In the next section, we explain a method that we used to identify various groups of students considering their performance on the course. We also detail the data that we applied for this identification.

V. IDENTIFY STUDENTS' GROUPS

During the semester, we noticed that students were interacting with the course differently. Some were highly active during a semester and achieved higher grades, while some performed well at the beginning of the course but they lost their interest after a while and focused mainly on the major evaluation



(a) XP Comparison.

(b) Grade Comparison.

FIGURE 6. XP and Grade comparison over years.

elements. There were also students that just performed enough to pass the course. Therefore, we found it of interest to identify different groups of students and assess how they interacted with the course. Considering groups rather than individuals also allows us for future extension of our course to a personalized one that can handle a large amount of data (a scalable course) for recommending gamified activities. In addition, it is of importance to find whether groups' performance is an adequate indicator to identify their groups. Hence, in this study, our main motivation is differentiating the students' groups considering their performance. For that, accumulated XP over a semester was estimated for each student, which allowed us to equally represent students with similar performance and clearly distinguish their ranks. In other words, if two students got similar amount of XP in the same day, with this data preparation method, we are able to distinguish their ranks since their previous XP was also taken into account. This approach was also used by Barata [11]. The amount of accumulated XP per day was used as attributes for the cluster analysis to group students by similarity of gaining XP. We used the K-means algorithm [39] for cluster analysis since it is easy to interpret and implement while has a linear complexity [51], [52], [54], [63].

A. NUMBER OF CLUSTERS

To identify the number of clusters (students' groups) we used the elbow technique. It runs K-means algorithm K times on the data (K is the number of clusters), and for each value of K estimates the Sum of Squared Errors (SSE). The results will be presented in form of a line graph (like an arm), where the elbow on the arm indicates the optimal number of clusters for that data. As shown in Figure 7, four is one of the most promising number of clusters for all years except 2010-11. For that year, three is a good candidate for the number of clusters. It can be because of not having enough learning activities and evaluation elements in the first year to clearly distinguish students' groups from each others. Anyhow, any of these numbers supports our initial assumption that there were different groups of students interacting with the course. Interestingly, our results are compatible with the results that are presented in a PhD thesis [61]. In that thesis, author also identified four groups of students considering their performance.

B. CLUSTER PERFORMANCE

In Section V-A, we could identify the optimal number of clusters for all years using the elbow technique and the accumulated XP of the students over time. In this section, we aim to present the average performance of each cluster using the same data. Here, we should state that each semester was different from another one in various terms, such as having different duration, learning activities, and evaluation elements. Therefore, the clusters from different years were generated using different collections of data.

As presented in Figure 8, all students' performance looked the same at the beginning of each semester since students were not fully enrolled in the MCP course or they still did not do any significant learning activity. During a semester and by doing enough activities by the students, their groups became more distinct. The performance of all clusters suddenly enhanced at the end of the semesters (i.e. end of May or beginning of June), which was due to obtaining a noticeable amount of XP from a multimedia presentation (it is a main course activity). After this presentation, since a little time (only two active weeks) and learning activities left, the clusters' performance often did not change significantly. In addition, the performance enhancement (at the end of the semesters) was larger in the first three years because the final exam was the main evaluation element that was located at the end of the course. After the first three years and by having quizzes, the course workload became more balanced and that enhancement got smoother. In the next section, we describe each of these clusters and analyze them in detail.

VI. GROUP ANALYSIS

In this section, we intend to describe each group, detail their characteristics and achievements, and analyze their

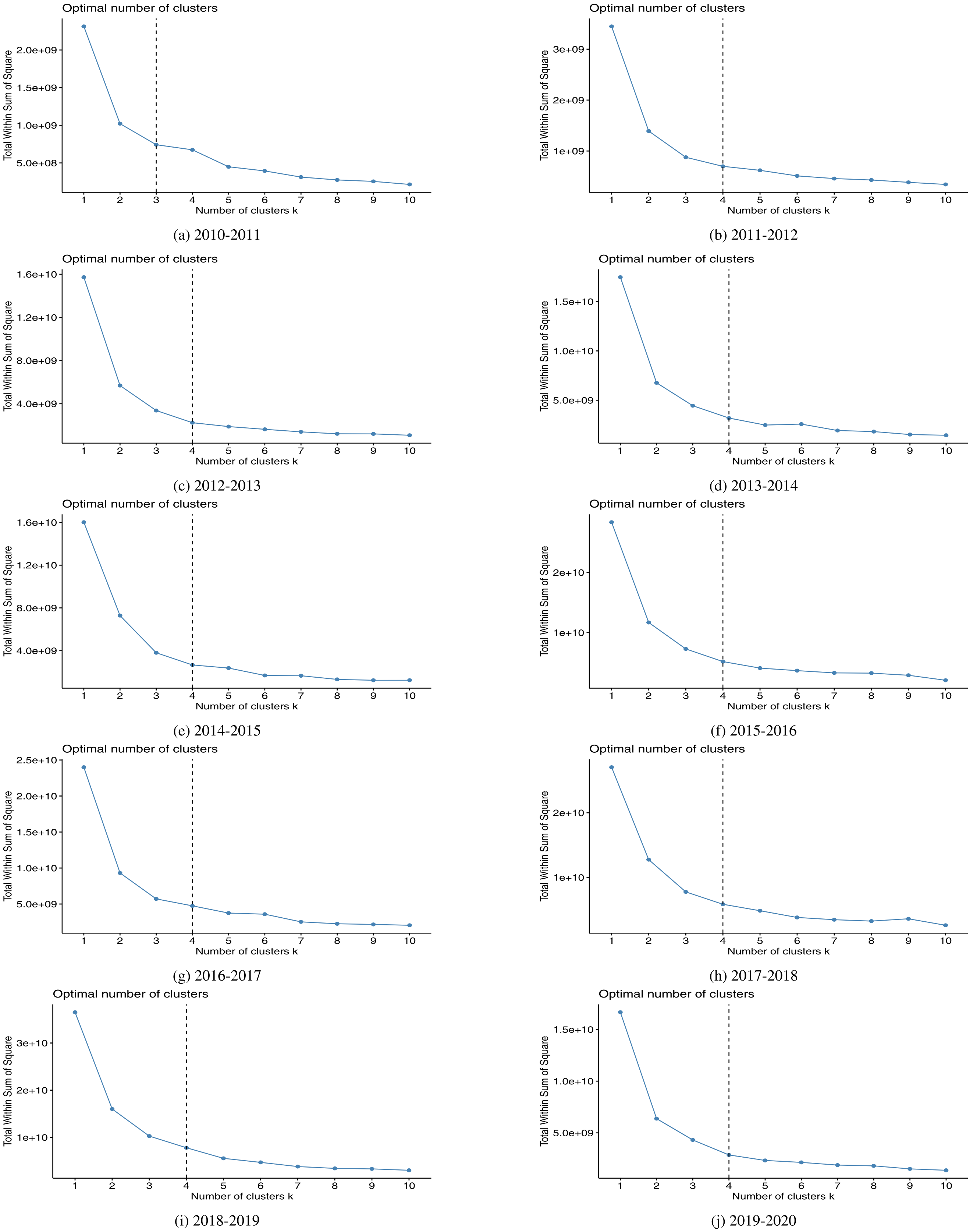


FIGURE 7. **Finding the optimal number of clusters via elbow technique and accumulated XP.**

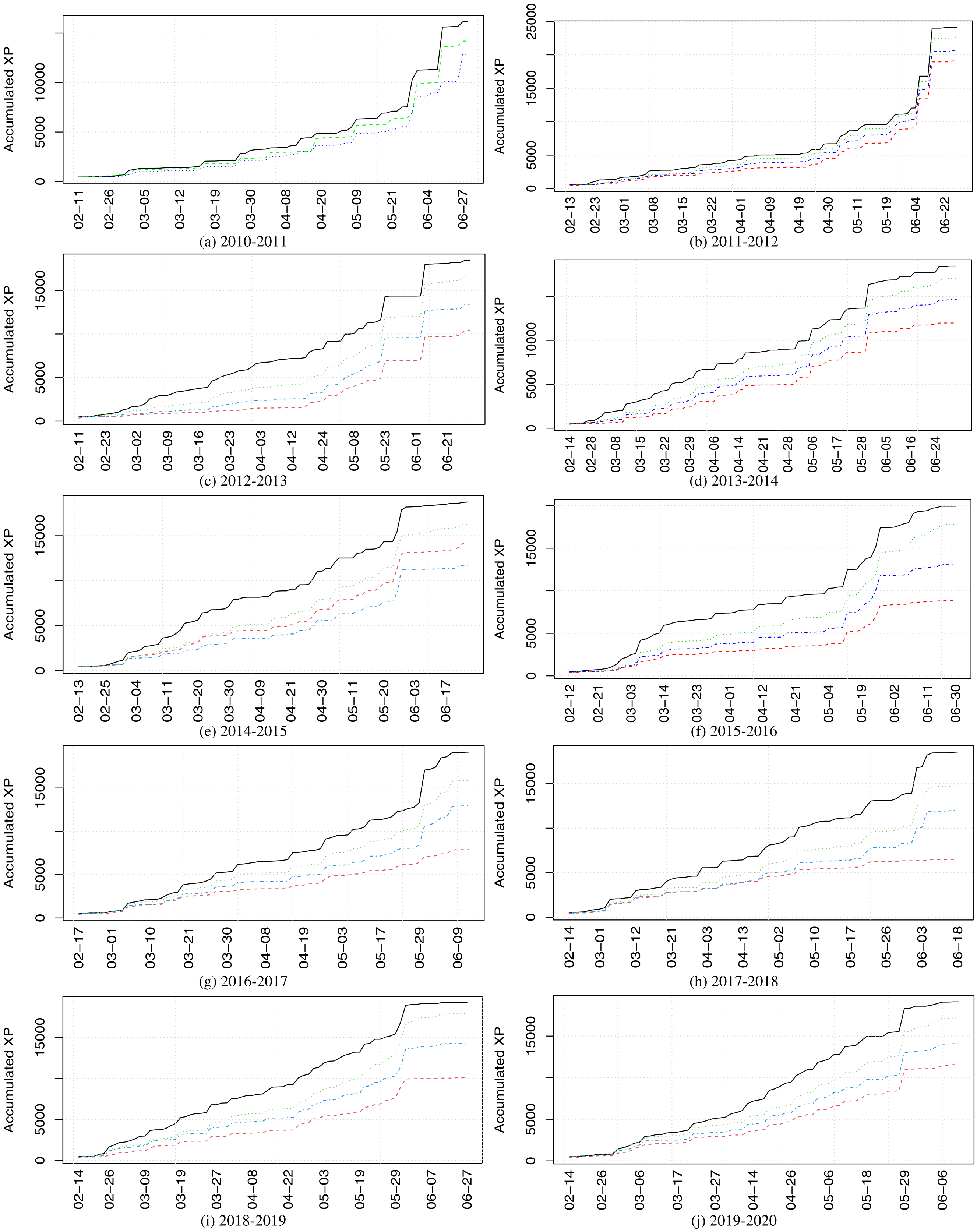


FIGURE 8. Average clusters performance using the accumulated XP. Here, black curves show the Achiever groups, green curves indicate the Regular, and blue curves present the Disheartened groups. Finally, red curves show the Underachievers.

interaction, engagement, and self regulation skill. Besides these, we also evaluate how early we can predict the students' groups and their performance with a high accuracy. In addition, we assess whether the gaming profile of the students had any relation with each student's group. This analysis allows us to better understand what students of a given profile like (Section VI-F). Finally, we study the students' feedback to evaluate how much we were successful in achieving our goals.

A. GROUP DESCRIPTION

As presented above, we distinguish four main student groups according to performance (**RQ1**). The first group is the Achievers, who concentrated on the achievements and obtained all the available XP. Due to the reason that they caught every opportunity to gain XP, they are called Achievers. These students had a higher XP accumulation curve (as shown in Figure 8), which presents that they were mainly ahead of other groups (positioned at the top of the Leaderboard). Regular students are the second group that their performance was good but lower than the Achievers, and balanced the achievements with the traditional assessment components. Their final grades were also close to the Achievers. Due to their behavior, they are named Regular students [8], [9]. In some studies, such as [10], they are called Late Awakeners since they seem to be like a fast rank loss in the beginning of the course followed by a progressive recovery (Figure 8-c, -f, -h).

The performance of the third group (Disheartened group) was lower than the Regular group, and it looks that they ignored some of the course activities. They normally started the course at a pace similar to the Achievers, but soon they lost their interest and fell behind in terms of XP acquisition. Their average Leaderboard positions, which was close to the Achievers at the beginning of a semester, dropped significantly as the semester evolves. This is the reason that we called them Disheartened. Finally, the last group is named Underachievers since the students in this group were showing little interest and engagement with the course. They also had the lowest performance and just did enough activities to pass the course, and were mainly positioned at the bottom of the Leaderboard. Although the Achievers and Regular students had a high performance and engagement with the course, the other two groups were disengaged and their performance was relatively low [8]–[10].

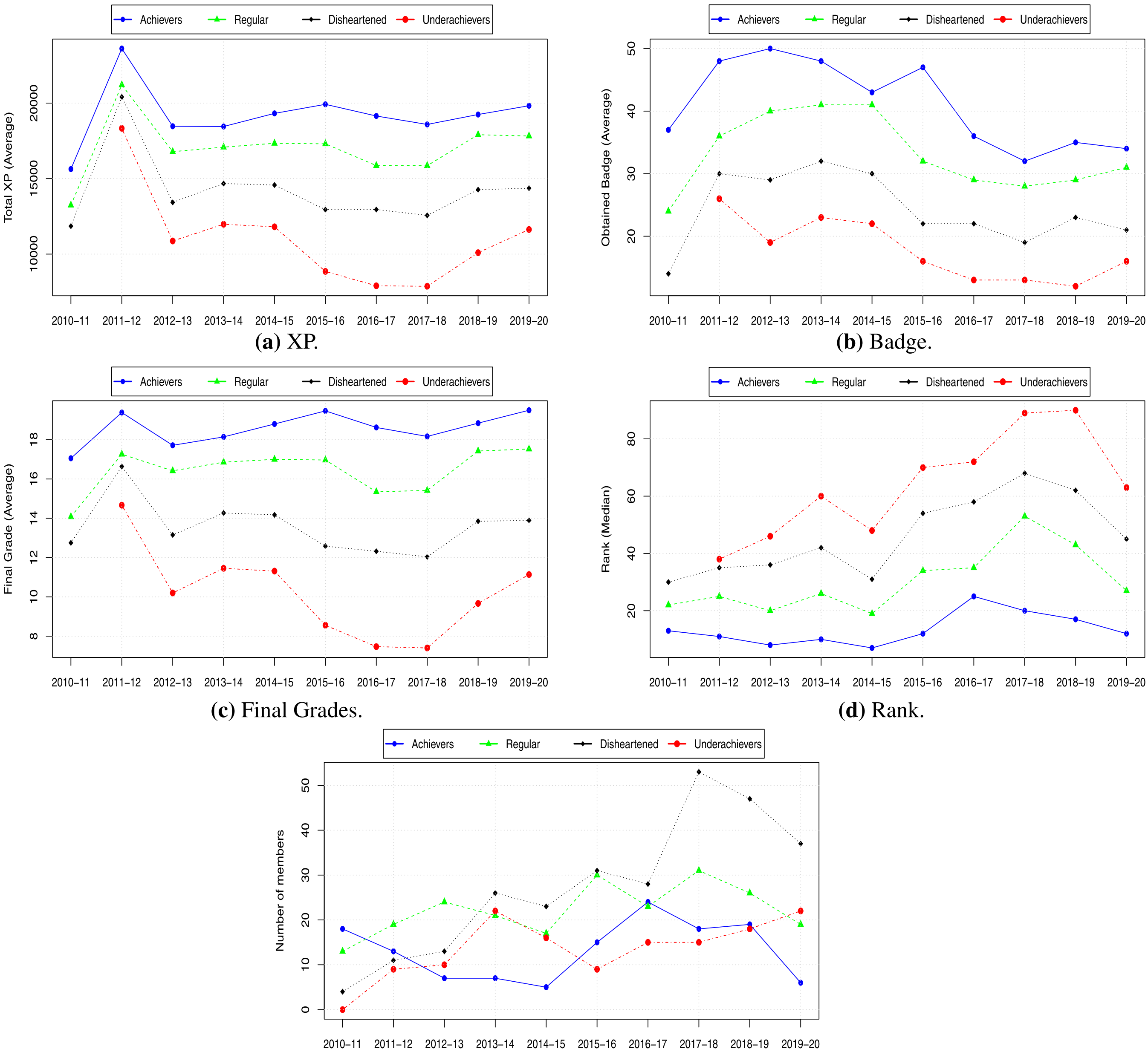
As mentioned in Section V-A, authors of [61] also found four groups of students. It is of interest that our groups are compatible with the ones that are presented in [61]. The performance of Achievers matched the performance of the students in the Advanced group of that study. Both groups almost completed all the learning tasks and got most of the points. Regular group that had the second best performance can be compared with the Good Participants group. The final grades of these groups were slightly lower than the Achievers and Advanced groups. The behavior of Disheartened was compatible with the Keen group. Both groups ignored some of the course activities. Finally, Underachievers had similar behavior to the Novice group and they did just enough to pass thee course.

B. GROUP ACHIEVEMENTS AND CHARACTERISTICS

This section presents some of the groups' statistics as well as showing their obtained XP, badges and final grades. In Figure 9, we compared the identified groups in terms of obtaining XP and badges. In the same figure, we also compared their final grades, ranks, and their size through the years. Here, we do not have the Underachievers group for the year 2010-11 since we only found three rigid groups among the students (see Figures 7 and 8). As shown in Figure 9 (sub-figures a to c), Achievers outperformed others in getting XP and badges, and had higher final grades in all years. They also obtained better ranks constantly (sub-figure d) (**RQ2 and RQ3**). To compute the ranks, we used median on the accumulated XP of each group's members. Median is opted since it is more robust than mean to outliers. In this computation, we ignored the students' ranks during the first four weeks of each semester (before March 15) due to having much variation (more details in Section VI-E). The groups' ranks increased through years, which is due to having more students.

Furthermore, we compared the size of all groups. As presented in Figure 9, Achievers and Underachievers (had the highest and lowest performance) were often in minority while the other two groups (had almost average performance) were the major ones. It matches most of the courses where only a few students out- or under-perform the rest of the students while the performance of the majority is around average. In terms of XP, the groups' XP (Figure 9-a) was noticeably higher in 2011-12, which was due to not having any restriction on the XP gained from the Skill Tree. This problem was fixed for the next years by restricting the achievable amount of XP from the Tree. Regarding the badges, the groups' badges reduced since 2015-16, which could be due to course restructuring in that year. Before the course restructuring, the enrolled students often were highly interested in course while after the restructuring their reasons for course enrolment varied from high interest to perception of easiness (more detail in Section IV). So, they all were not that motivated to earn many badges.

We also analyzed the performance of groups in earning badges from different levels of the Skill Tree. In every year, for each level, we summed the number of obtained badges by each group and divided it by the size of that group. In Figure 10, each bar refers to a level of the Tree while every color indicates a semester (eight colors because the Tree is in-use since 2012-13). As presented in this figure, all groups had better performance in level one since the tasks were less complicated. Groups' success got lower by raising the levels since the tasks became more challenging. As expected, Achievers had the best performance in comparison with others in obtaining badges from all levels.



(a) XP.

(b) Badge.

(c) Final Grades.

(d) Rank.

(e) Size.

FIGURE 9. Comparing the groups XP, badges, rank, final grades, and size through the years.

C. COURSE ENGAGEMENT

One of the current educational challenges is to keep students effectively engaged with a course [43]. Hence, in this section, we intend to show how much we were successful with that. For this purpose, we considered the density of activities and the collected XP by each group in every semester. For that, we initially built a binary matrix for every year in a way that whenever a group member did an action (e.g. made a post, got a badge, performed a task) or collected an XP in a day, we regarded it as one otherwise it considered as zero. Then, these matrices were used to generate the density graphs.

As shown in Figure 11, the densities of all groups in all years almost follow the same pattern. At the beginning of each semester, due to awarding an initial XP to all students as well as having simple tasks that students could get their XP, the density of all groups is one. Then, all densities dived drastically till end of April. After April, by getting close to the end of the semesters, students increased their activities and obtained more XP to get better grades or even pass the course. Within this time (after April) and by approaching the end the course, we can also observe some shifts in groups' performance (change in average XP) that was never happened

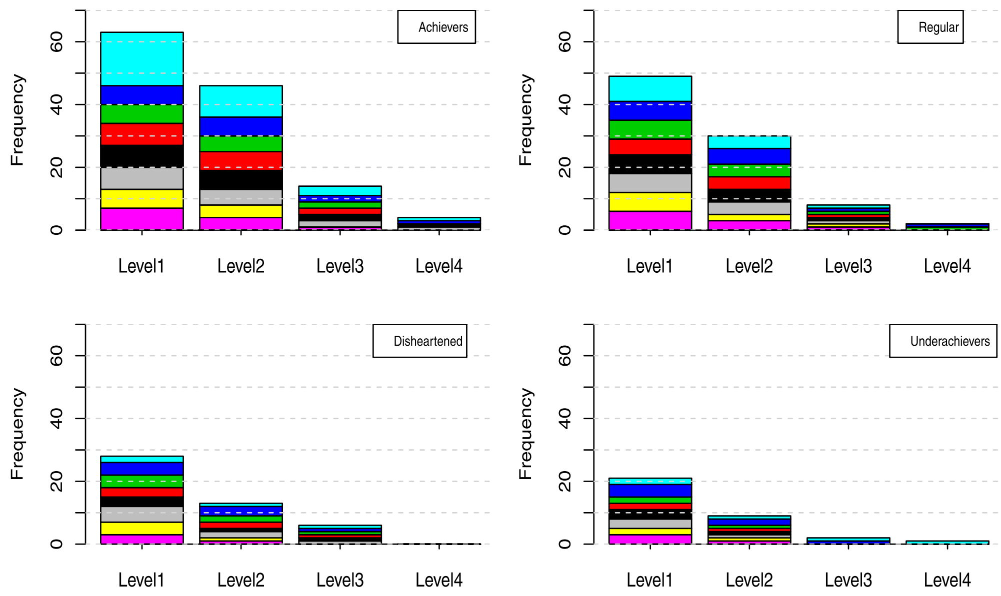


FIGURE 10. Earned badges from each level of the skill tree by each group.

before April. It confirms that the students are often deadline driven (**RQ4**).

Similar pattern could be observed in Figure 12, where the density activity of all groups dropped until the end of April and raised after this month. In this graph, we ignored the initial XP that was awarded to all students. These XP were to motivate the students to continue with MCP and to solve the grade rounding issue at the end of the semester. Here, it is noticeable that the density of several groups became one (or close to one) at the end of the semesters, which shows that the groups were mainly active in that time, but sometimes without obtaining XP (compare the results in Figure 11 with 12).

In both figures, we noticed that the students' XP and activities were in the minimum in April. It is due to the reason that the semester break is in this month and students go for holidays for almost one week. Therefore, their activities, and subsequently their XP dropped significantly in this month. Also, these graphs show that the students are mostly deadline-driven and become active at the end of the semesters. In order to have them well-engaged with a course, we need to define activities that are carefully distributed during a semester, which keeps the students interacting with the course continuously.

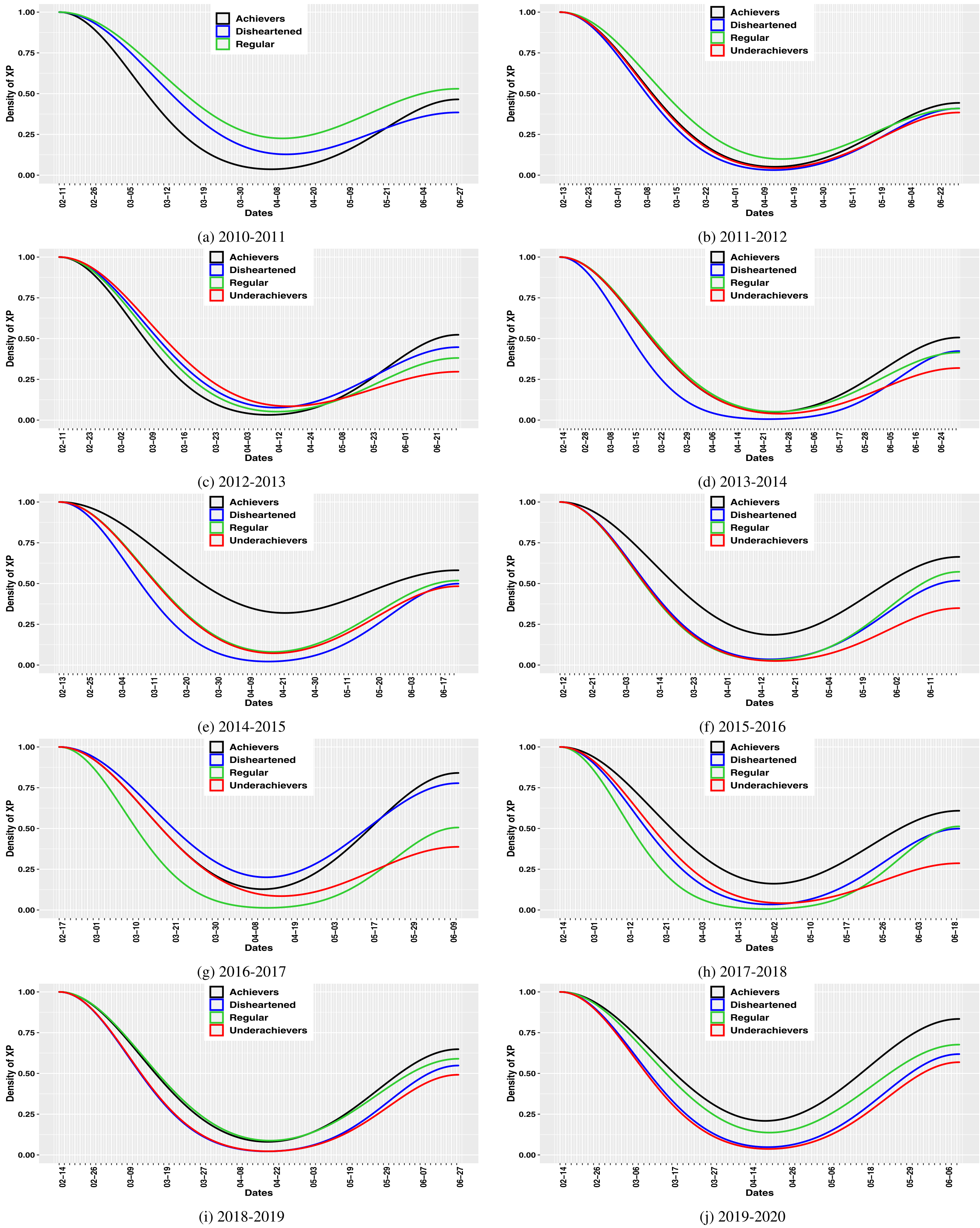
D. STUDENT INTERACTION WITH THE COURSE

Students' interaction with the course is an informative type of data that we used to analyze whether there is a relation between the students' interaction and their performance. For that, we employed the students' log data the we collected from the Moodle. We then generated scatter plots using this data. In Figure 13, each dot indicates that a student interacted with the course in a day. As shown in this figure, Achievers interacted frequently while the Underachievers interaction was quite sparse. It also presents that the Regular groups were interacting with the course more than the Disheartened ones but less than the Achievers. Therefore, we can conclude that there is a direct relation between the students' interaction and their performance in the course. So, whenever a student is more active and interacts with the course frequently, he/she mainly gets a better grade than the ones with a lower level of interaction. This statement is also confirmed by the results shown in Figure 14, where the students got higher final grades whenever they interacted more with the course.

Like Figures 11 and 12, in Figure 13 we also observe that in April all groups interacted less with the course, which is due to the semester break. In addition, the groups' interaction became sparse at the end of the semesters ($\approx$ after the first week of June). It is because that around this time students delivered their multimedia presentation (one of the main course activities at the end of the semesters) and there was not much time (there are only two active weeks in June) and course activities left to complete. Hence, the groups' interaction dropped significantly. Moreover, in the first five years, the initial interaction of the groups with the course was sparse while it became denser since 2015-16. It might be due to the course restructuring in this year and better distributing of the game elements throughout the course. Similarly to the results presented in Figure 9, in Figure 13 we also see that the Regular and Disheartened groups are often the larger ones while the other two are smaller.

E. PERFORMANCE STABILITY

One of the main challenges in educational environments is to identify the students' groups (Profiles). It assists professors to provide more suitable learning materials for the students that match their competency levels. This identification needs to be



(a) 2010-2011 (b) 2011-2012

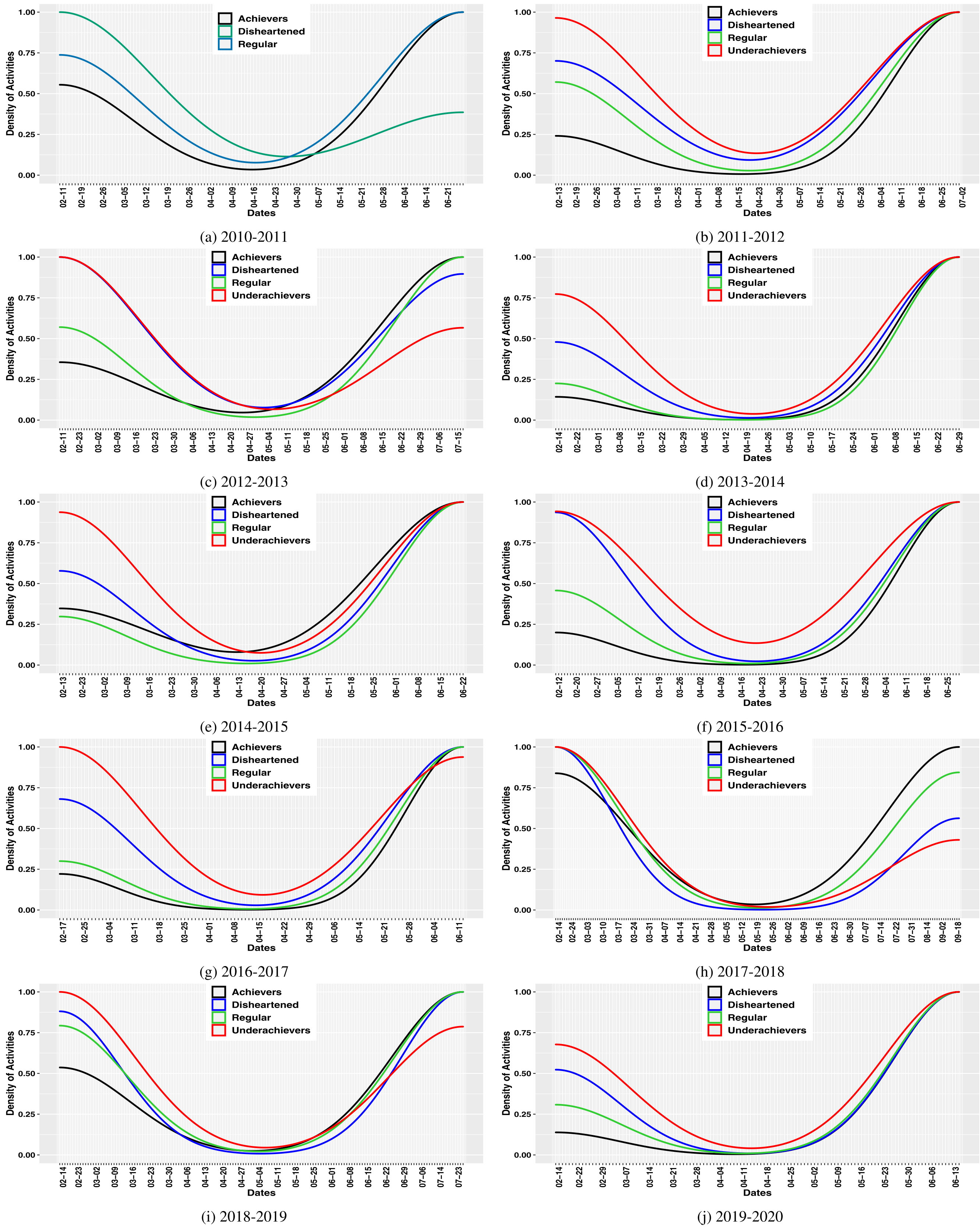
(c) 2012-2013 (d) 2013-2014

(e) 2014-2015 (f) 2015-2016

(g) 2016-2017 (h) 2017-2018

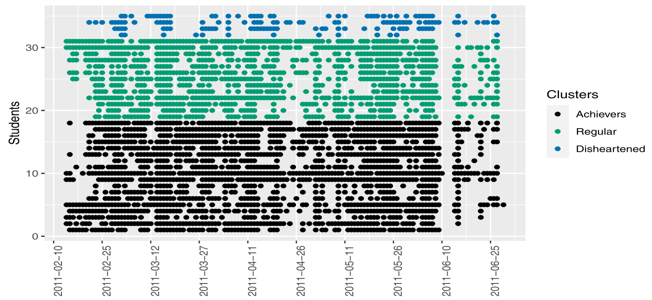
(i) 2018-2019 (j) 2019-2020

FIGURE 11. **Density of obtained XP by groups.**

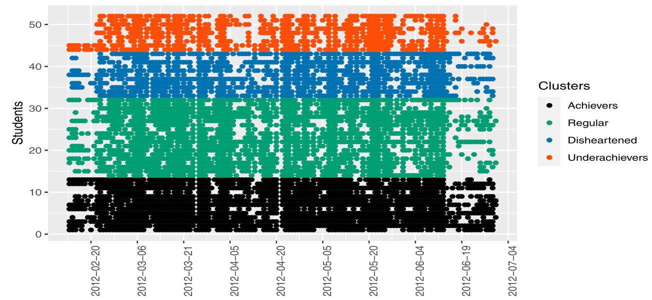


(a) 2010-2011 (b) 2011-2012 (c) 2012-2013 (d) 2013-2014 (e) 2014-2015 (f) 2015-2016 (g) 2016-2017 (h) 2017-2018 (i) 2018-2019 (j) 2019-2020

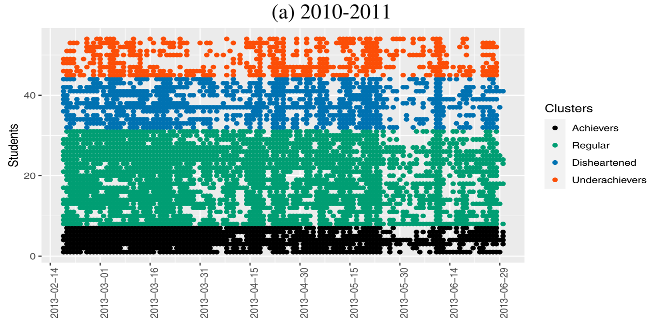
FIGURE 12. Density of activities by groups.



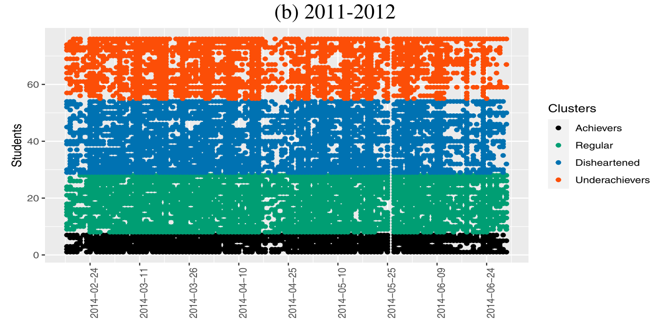
(a) 2010-2011



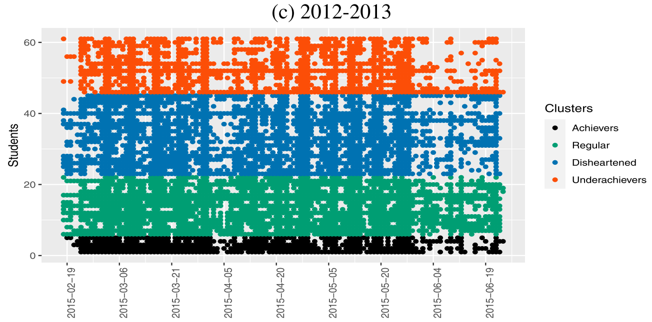
(b) 2011-2012



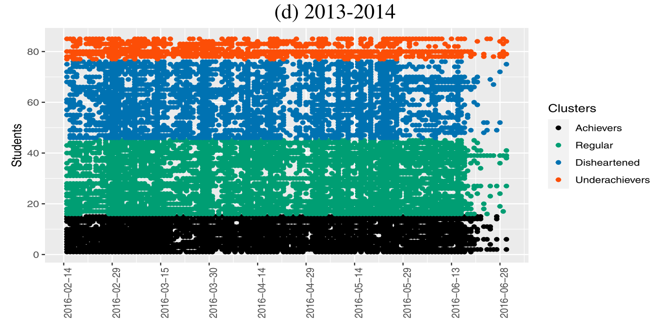
(c) 2012-2013



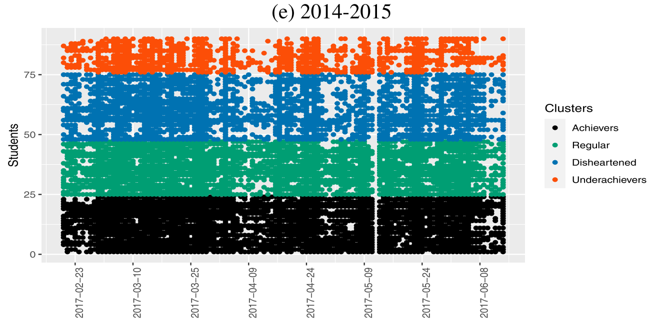
(d) 2013-2014



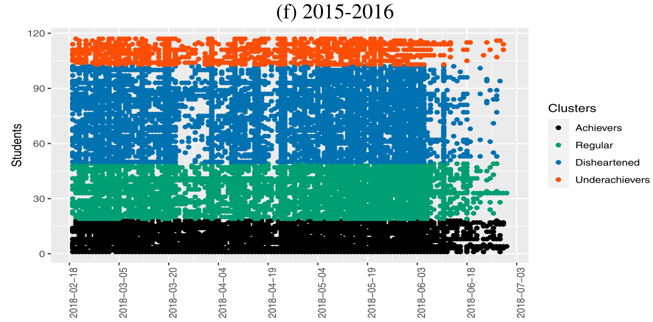
(e) 2014-2015



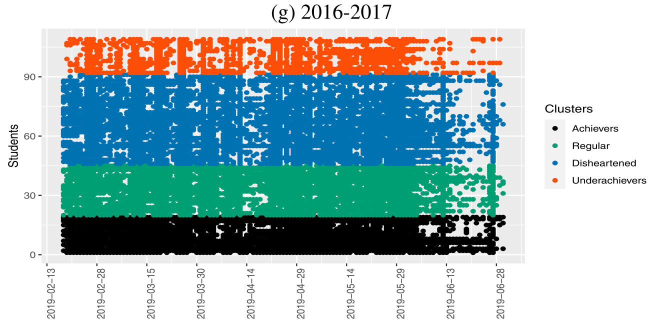
(f) 2015-2016



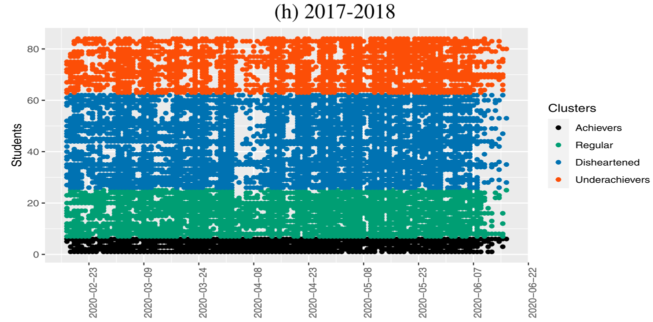
(g) 2016-2017



(h) 2017-2018



(i) 2018-2019



(j) 2019-2020

FIGURE 13. Students Interaction with MCP.

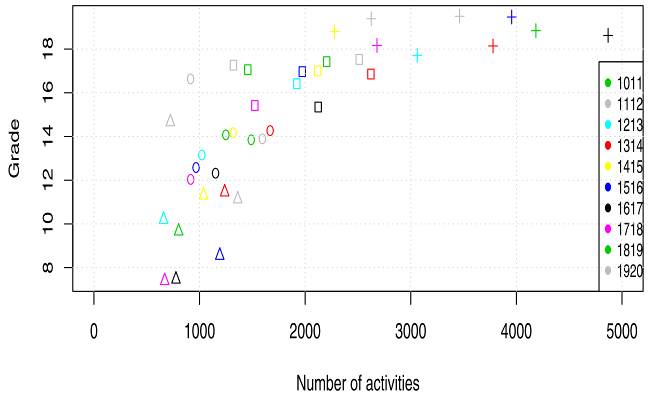


FIGURE 14. Relation between students' interaction and their final grades. Here, a year can be shown by three or four dots depending on the number of students' groups for that year. Achievers +, Regular □, Disheartened ○, Underachievers △.

made as soon as starting a course with a high accuracy since it gives more time to the professors to better guide different groups of students. Hence, in this section, we study how soon the students' performance gets stable for having a more accurate identification. For that, we initially estimated the accumulated XP of each student through a semester. This data enabled us to differentiate students that got similar XP in the same day. Then, the daily ranks of students were computed using the mentioned data.

In Figure 15, we presented the students' ranks through the semesters. Here, each group is highlighted in a different color, and the average final grade of each group is mentioned on the right side of the graphs. In these graphs, students with better performance have lower ranks (located at the bottom of the graphs), and vice versa. As presented in Figure 15, the students' ranks varied a lot in the first four weeks of each semester. Then, they got stable till the end of the semesters (**RQ5**). It could be due to the reason that the students were new to the course and they still did not do much activities to gain XP (i.e. their ranks are similar). Hence, obtaining a small XP could affect their ranks significantly. After almost the first month ($\approx$ March 15) and by earning a considerable amount of XP, the students' ranks did not change noticeably anymore. This trend can be observed in all years and the course changes and various game elements did not influence this trend. It indicates that theoretically we should be able to identify the students' groups and predict their performance with high accuracy around this time. According to our results for predicting the students' groups and performance, which are not mentioned in this paper, we could achieve the accuracy of $\approx$ 85% (via Random Forest algorithm [56], [62]) around mid March. Therefore, soon after starting the course we are able to identify the students' groups and performance accurately.

F. GAMING PROFILES

In this section, we intend to analyze whether there is a relation between the students' groups (indicating their performance) and their gamer profiles. The importance of this analysis is, the better we know what a profile likes, the better we can provide gaming experiences for it. There are several models to determine the gaming profiles of the students, such as the one presented by Richard Bartle [12], [13] that divides students into four groups, or the one called Demographic Game Design 1 (DGD1) [15]. DGD1 was introduced by Chris Bateman, and is based on the Myers-Briggs personality model [49]. Among all the models, we selected the Brainhex [55] since it is one of the most complete models that works based on the previous ones, such as DGD1. In addition, its questionnaire is available and online,[2] which makes it easy to administrate and access. This model is based on neurobiological responses inherent to playing games [11]. It includes seven player archetypes, and classifies players into primary (main) and secondary (sub) classes. The seven archetypes are: Achiever, Conqueror, Daredevil, Mastermind, Seeker, Socializer, and Survivor.

TABLE 2. Student groups and the Brainhex classes.

	Main Class			Sub Class		
	Achiever	Conqueror	Mastermind	Achiever	Conqueror	Mastermind
Achievers	21.7%	24.3%	25%	21%	28.2%	19.7%
Regular	16.6%	24.3%	23%	20.5%	21.7%	18.5%
Disheartened	10.7%	34.1%	20.9%	21.9%	20.4%	20.9%
Underachievers	15.6%	33%	19.1%	16.5%	20.8%	22.6%

To determine the Brainhex classes of the students (i.e. main-class and sub-class), we initially used the aforementioned questionnaire at the beginning of a semester. We then related the collected information to the students' clusters [11]. The results for all years are shown in Figure 16. Although it was rough to determine the major classes in some years, Mastermind, Conqueror, and Achiever were the three major primary and secondary classes for all clusters (students' groups).

Mastermind refers to the students that enjoy solving puzzles and strategic games while concentrating on the most efficient decisions. Achiever addresses the goal-oriented students who like to collect special achievements and points. Finally, the Conqueror players like challenges, enjoy defeating tough rivals and beating others. We believe that the MCP was not attractive for this group since there was never an actual rival to defeat. Anyhow, it might come from the fact that the course Leaderboard encouraged the students' competitiveness.

As shown in Table 2, although the percentage of the three mentioned classes (main and sub) varied for each student group, in total around 60 to 70 percent of each group was formed by the mentioned classes and the remaining percentage was composed of the other four classes. To show that there was not any significant difference among the student groups considering the Brainhex classes, we estimated the p-values using all main and sub classes for each pair of student groups. As presented in Table 3, there was not

[2]http://www.survey.ihobo.com/BrainHex/

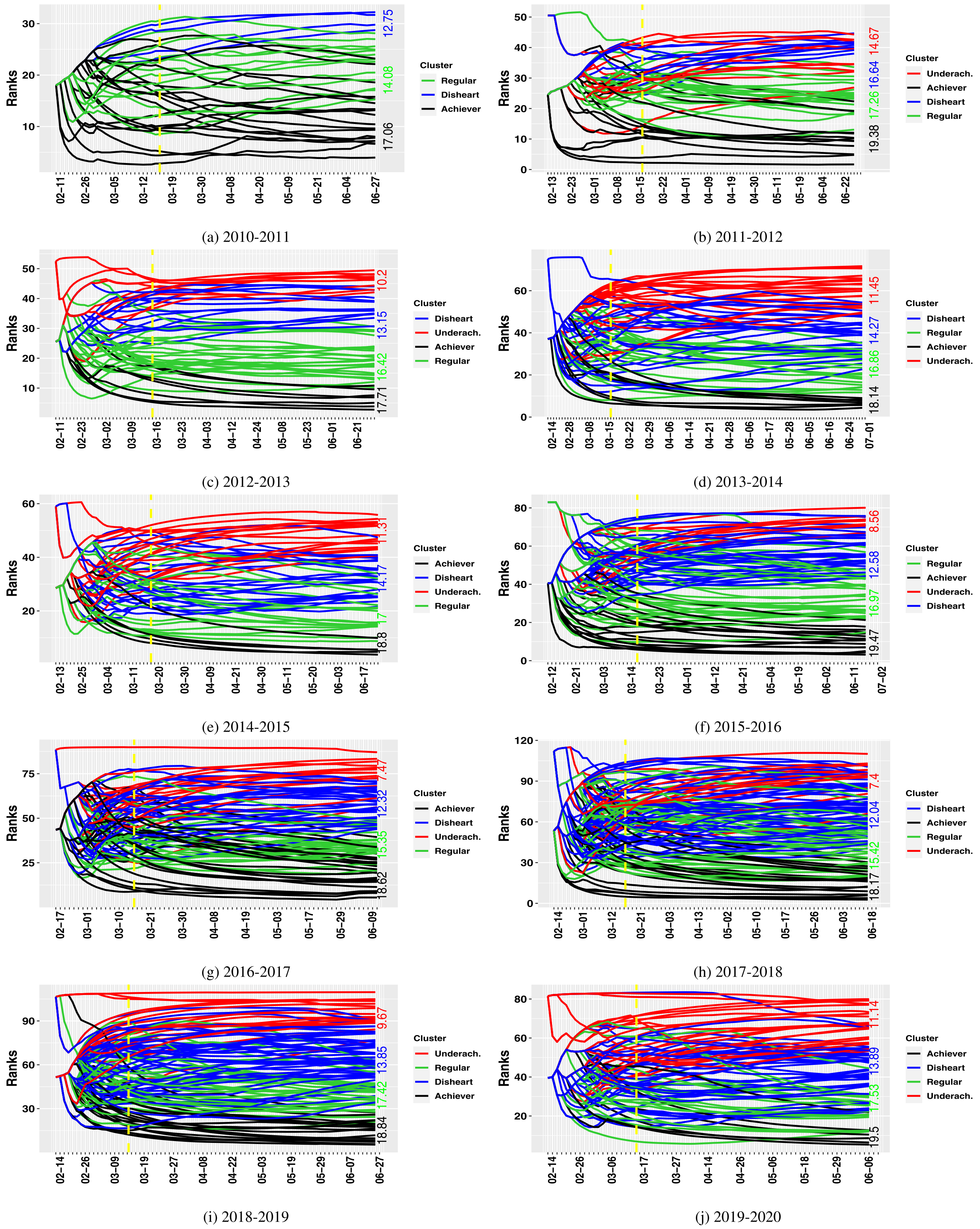
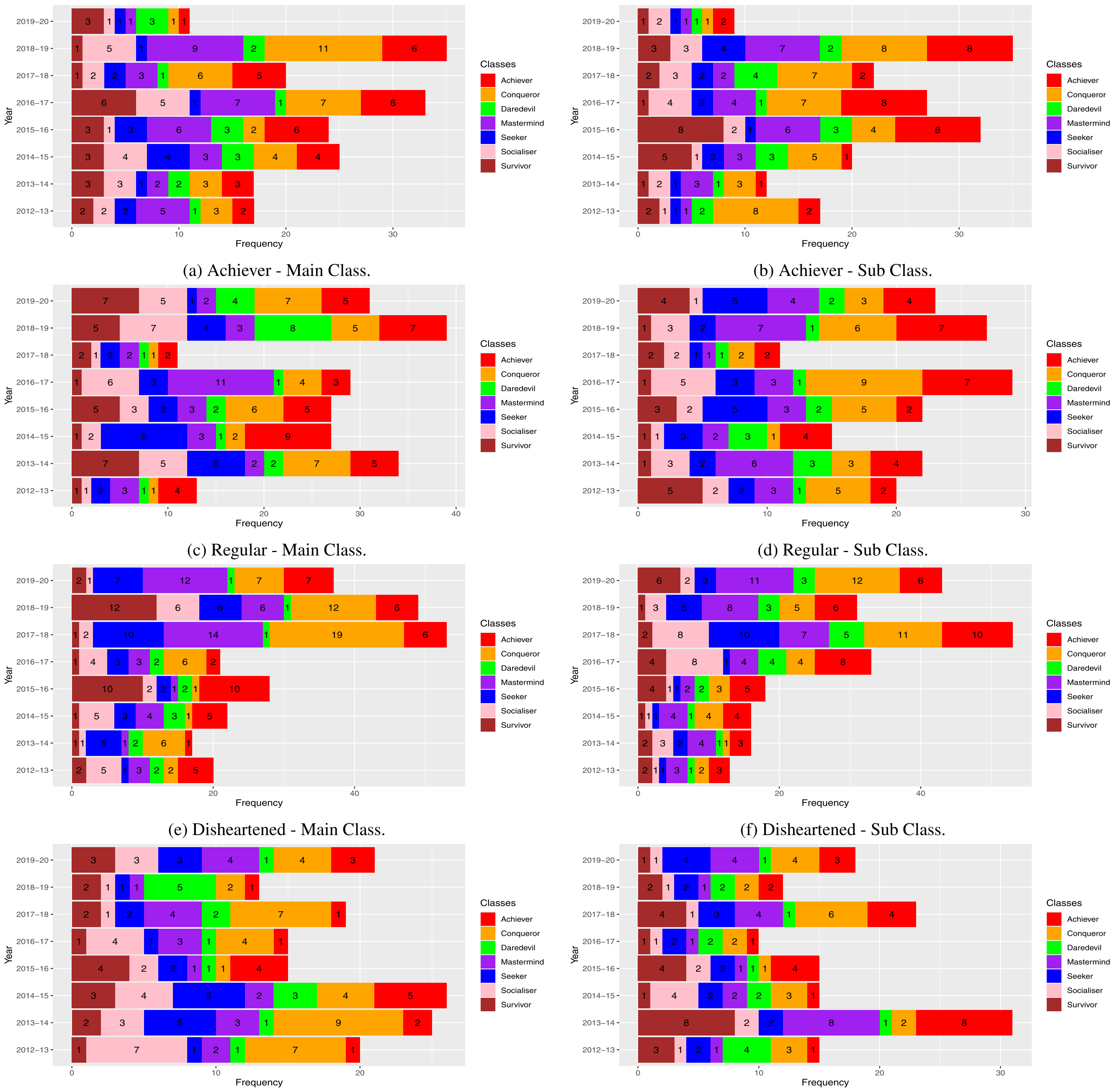


FIGURE 15. **Performance stability of students. Black: Achievers, Green: Regular, Blue: Disheartened, Red: Underachievers.**



(a) Achiever - Main Class. (b) Achiever - Sub Class.
(c) Regular - Main Class. (d) Regular - Sub Class.
(e) Disheartened - Main Class. (f) Disheartened - Sub Class.
(g) Underachiever - Main Class. (h) Underachiever - Sub Class.

FIGURE 16. Brainhex categories for each student's cluster.

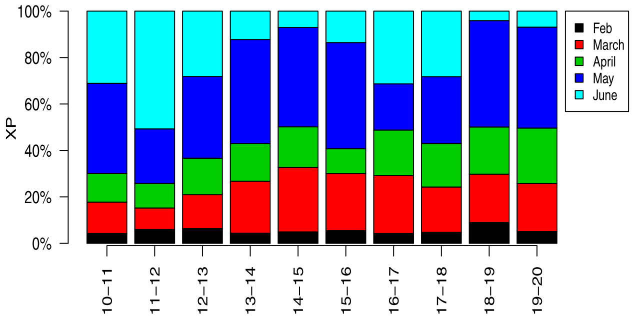
TABLE 3. P-values across student groups. Significance = 0.05.

	Main Class	Sub Class
Achiever-Regular	0.94	0.93
Achiever-Disheartened	0.46	0.31
Achiever-Underachievers	0.45	0.37
Regular-Disheartened	0.49	0.29
Regular-Underachievers	0.39	0.23
Disheartened-Underachievers	0.21	0.05

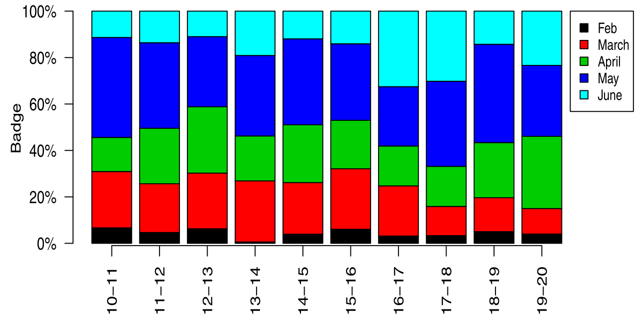
TABLE 4. P-values across all years. Significance Level = 0.05.

	1213-1314	1314-1415	1415-1516	1516-1617	1617-1718	1718-1819	1819-1920
Main Class	0.45	0.66	0.49	0.69	0.65	0.96	0.54
Sub Class	0.36	0.49	0.27	0.68	0.59	0.94	0.41

an important difference among the student groups. We also estimated the p-values among consecutive years considering all Brainhex classes to assess whether the course changes influenced the proportion of the classes through years. According to our results presented in Table 4, the proportion of the classes did not change significantly. From all these results, we can conclude that there was not any specific



(a) XP.



(b) Badge.

FIGURE 17. Proportion of obtained XP and Badges per month.

relation between students' groups and their gaming profiles (**RQ6**).

G. SELF-REGULATION

In e-learning courses, whether gamified or not, one of the main issues is the lack of self-regulation skill of the students regardless of their groups (profiles). It results in missing a considerable amount of time during a semester while being pressured by the learning tasks and activities before the due dates. Thus, this skill can be considered as a significant one for assisting students to better distribute their workload through a semester. To this end, we assessed the self-regulation skill of the students in all years (Figure 17). This assessment was based on the number of XP and badges earned by the students from beginning to the end of a semester.

In Figure 17, we observe how students' activities were biased towards the end of the semesters. Taking into account that there were only two active weeks in February and June, we immediately notice that in June students were more active than February, even if we deduct the effect of the final exam in the first two years. Also, the multimedia presentation that was in May/June might explain a bump in XP, but it could not justify an increase in badges since it only awarded XP. Furthermore, we can see that in May students were more active than April (the semester break is in April) and March. As it is shown, in all years, almost half of the rewards were obtained in May and June. Moreover, in 2011-12, the amount of XP obtained in June was significantly higher than the rest of the years. It is because that the initial version of the Skill Tree was introduced and it was not well integrated with the course and its tasks were not well distributed over the semester. So it caused a bump in XP.

We can conclude that the most of the students were lacking the self-regulation skill, and were mostly deadline driven (**RQ7**). It was also mentioned in their feedback (collected since 2012-13) that their effort was unbalanced and they did not study regularly through a semester (Figure 18). In 2015-16, the average score was slightly higher. It could be due to the reason that the course was restructured in that year,

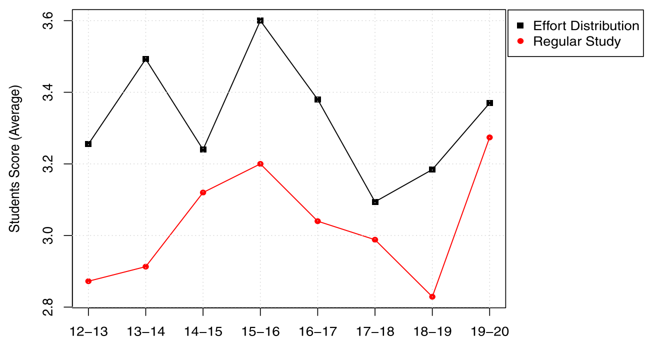


FIGURE 18. Students' score (1 to 5) to their effort distribution and regular study over a semester.

and it became more flexible. Nonetheless, this score was not significantly high.

H. STUDENTS' FEEDBACK

The students' feedback on the gamified MCP was collected since 2012-13. For this purpose, we provided a short questionnaire to collect their opinions about various aspects of the course. Initially, we focused on their general feedback on the course, like if it was competitive, likable, creative, interesting, extendable to other subjects, or even if the students could learn more with it in comparison with the other courses. As presented in Figure 19, most of the students had a strong agreement on the creativity of the course and confirmed that it was engaging, likable, extendable, interesting, motivating, and well-received. In spite of having their agreement on the mentioned aspects, they disagreed that they could learn more with the MCP than the other courses.

It is worth mentioning that in 2019-20, where almost all universities and courses were negatively influenced by the Covid-19 pandemic, the students were positive about MCP and found it motivating and engaging (Figure 19). This, shows that MCP was robust to this issue, and the pandemic had almost no impact on its effectiveness. Also, the results presented in previous sections confirmed that the students' engagement and interaction did not change in 2019-20.

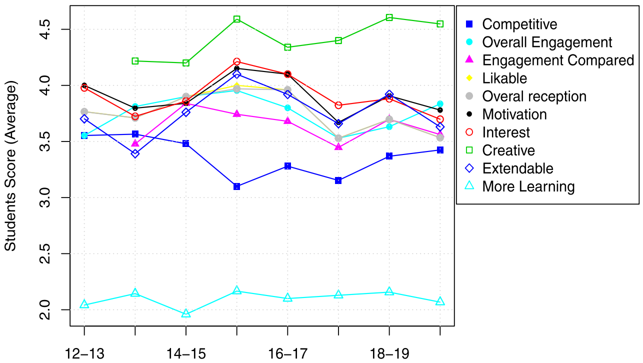


FIGURE 19. **Student feedback on the general aspects of MCP.**

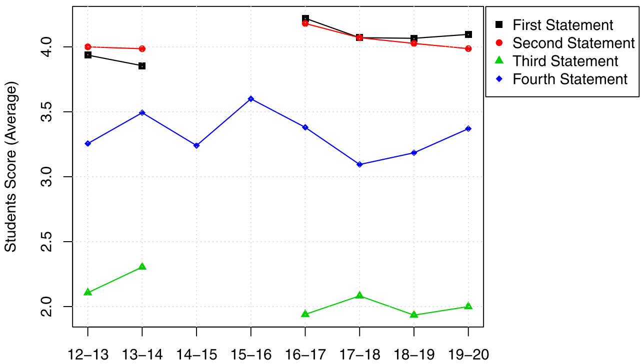


FIGURE 20. **Students' feedback on the quality of game elements. In 2014-15 and 2015-16, we could not collect the students' opinions about the first, second, and third statements.**

The students' feedback on the quality of the game elements was collected using four statements. Like Figure 19, we asked the students to rate the statements using a five-point Likert scale. These statements were:

1) Game elements made the course more interesting.
2) Game elements made the course more engaging.
3) MCP would be better off without game elements.
4) Game elements contribute for a greater workload when compared with traditional courses.

Results are presented in the form of a line graph in Figure 20. A brief look at this figure shows that the students highly agreed that the course was more interesting with the elements and they made it more engaging. Furthermore, they strongly disagreed that the course could be more effective without the elements. Finally, they were not sure whether the elements made the course more demanding or not (Fourth statement). According to these results, we conclude that our game elements were well-designed and worked in a promising way that the students did not feel exhausted using them.

In addition to the mentioned students' opinions, we collected their feedback on how successful was the MCP in boosting the students' autonomy, creativity, and whether they could learn some practical skills that could be useful for their future. For that, we asked the students to give us their feedback (rate from 1 to 5) on three statements that focused on the mentioned criteria. The three statements were:

1) The course allowed me to get rewards for things I like to do.
2) The course allowed me to be creative.
3) The course taught me useful skills for my future.

The results are summarized in Figure 21. As shown in this figure, students agreed that the course was autonomous and flexible (First statement), and it highly triggered their creativity (Second statement). Moreover, in MCP, students needed to use various tools and software to complete assignments and tasks, which they found it useful for their future (Third statement). In this figure, we can see a notable drop in 2017-18. It could be due to not having the Quest element in the course anymore. Also, the peak in 2015-16 (for the Second and Third statements) is because of the course restructuring in that year, which caused that the students found the course more creative and valuable for the future.

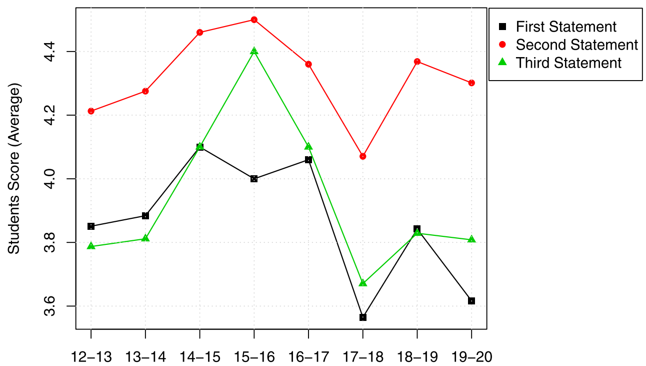


FIGURE 21. **Students' feedback on the course autonomy, creativity, and usefulness.**

VII. DESIGN IMPLICATIONS

After analyzing the students' data for all years, we are in the position to suggest a set of design guidelines, which can be helpful to design a more effective gamified learning environment. In this section, we discuss them in detail.

A. MULTI ACTIVITIES

In MCP, we aimed at enhancing student autonomy, interaction, motivation, creativity, and engagement with the course. Through years, we have realized that these goals cannot be achieved using only a single learning activity and a combination of activities is required. For example, the Leaderboard would motivate the students to be engaged more with the course for getting higher grades and ranks, and subsequently, it would influence their interaction, while the Skill Tree would enhance the students' autonomy since it enables them to perform the preferred activities for obtaining rewards. Similar results were presented in [3], where the authors stated that a single game cannot be used to achieve all the gameful learning goals, such as enhancing engagement, motivation, and performance.

B. QUANTUM ASSESSMENT MECHANISM

As explained in Section VI-C, students are deadline-oriented and often become active at the end of the semester. This lag leads to idle time during the semester and a frenzied flurry of activity just before deadlines. To avoid these issues and to better distribute students' workloads over a semester, it is essential to have an assessment mechanism that awards students as they participate in a course. As mentioned in Section IV, MCP uses a quantum assessment mechanism, and students are granted (via XP, levels, and badges) whenever they are performing course activities. This mechanism provides a transparent, traceable, and incremental continuous assessment of student performance down to the individual XP. Figure 8 shows that in the first three years (when MCP grade was not well granularized) there was a bump in XP at the end of the semester because the main assessment took place at the end of the course. Since 2013-14 and by having quizzes instead of the final exam, students became more active during the semester and the grade bump became smoother (i.e. performance curves became more linear).

C. FOUR STUDENT GROUPS

The main goal of our study was to analyze how different student groups interacted with a gamified course. For that, we distinguished each grouping by applying clustering techniques to student performance data. As already presented in Figure 7, we identified four distinct clusters. Considering that the course was modified in each year, we conclude that the underlying four-cluster model is both stable and resilient to external changes. Interestingly, the same results were obtained by Nabizadeh [61]. He also confirmed that four is the most promising cluster configuration among all student populations considered.

D. EARLY GROUP IDENTIFICATION

We analyzed the performance (ranks) for each cluster across all years (Figure 15). We noticed that performance varied a lot at the start of each semester ($\approx$ first four weeks). This could be due to lack of early engagement in collecting XP. Thus, even collecting a few XP would affect a student's rank remarkably. After the middle of March (almost one month after semester start) and collecting significant XP, student ranks became more stable. This implies that within a month after course start we can identify clusters and predict student performance with high accuracy. This enables us to tailor the gamified experience in the way that best suits different student groups early in the course.

E. INITIAL ENGAGEMENT

In our analysis, we noticed that the students were almost inactive in the first two weeks of each semester, and then slowly started interacting with the course. This time can be considered as a golden time since if we get the students to work with the course as soon as starting a semester, we would be able to collect more interaction data. This data is essential to have an accurate prediction of the students' groups and their performance. This accurate prediction in the early stages of the course gives us enough time to better guide the students by tailoring the gamified activities in the way the best match the students' preferences and groups. Therefore, finding a strategy to make the students interacting with the course right after starting it is of importance. One strategy could be placing simple learning tasks that have extra rewards (grades) at the beginning of the course to convince the students for interacting with it.

F. BRAINHEX NOT SIGNIFICANT

As presented in Section VI-F, Achiever, Conqueror, and Mastermind were the three most frequent classes of Brainhex (main and sub) among all groups of the students. Therefore, we can conclude that there was not any specific relation between a single class and a students' group. Our statement was also confirmed in [9], where the authors predicted the students' groups using different collection of data. They got the accuracy of 71.70% via the BayesNet algorithm [34] using the Brainhex classes, while after ignoring the BrainHex categories, they achieved the accuracy of 79.63%. They also used other algorithms, such as Logistic regression [45], for the prediction task. In some cases, the accuracy enhanced using Brainhex classes while in other cases the accuracy got worse. Hence, we can conclude that the Brainhex classes might not be an informative and significant type of data, especially for the prediction task.

G. GENDER NOT IMPORTANT

During our study, we analyzed whether the students' genders could influence the accuracy of predicting their performance, and consequently, their groups. For that, students' performance was predicted using their logs data and their past grades (on the same course) with and without using gender information. Random forest [18], Naive Bayes [65], and K-Nearest Neighbor (KNN) [58] were used for the prediction task (Table 5). In both cases, the accuracy results were the same and did not change. In Table 5, we present, as a sample, the results for years 2010-11, 2014-15, and 2018-19. As can be seen, using or omitting gender information had no impact in the prediction's accuracy. Values for the other years, while not show for brevity's sake, follow the same pattern. Therefore, we can deduce that gender has no impact on predicting students' performance and groups.

TABLE 5. Prediction accuracy of students' performance with and without using genders data.

	Random Forest			Naïve Bayes			KNN		
	10-11	14-15	18-19	10-11	14-15	18-19	10-11	14-15	18-19
With Gender	0.834	0.869	0.910	0.664	0.746	0.728	0.761	0.641	0.659
Without Gender	0.834	0.869	0.910	0.664	0.746	0.728	0.761	0.641	0.659

VIII. CONCLUSION

In this paper, we studied how different groups of students behaved in a gamified environment and interacted with it. For that, we developed a gamified course called Multimedia Content Production (MCP), and collected the students' interaction data with it for ten years. Students' groups were determined using clustering techniques, which were applied to the students' performance data. In general, we often distinguished four groups of students (clusters) in all years.

By analyzing and comparing these groups, we noticed that all groups were deadline driven and became active as the deadline loomed. Also, the performance and final grades of all groups were accurately predictable within the first month after starting the course since within this time the students' performance got stable and did not change noticeably anymore. Furthermore, we concluded that the students' groups did not have any relations with their gaming profiles. Finally, several practical implications for designing a gamified course were suggested, such as having a continuous evaluation mechanism (i.e. awarding students as they participate in a course) throughout a semester that helps to keep students engaged with a course and distribute the workload. There are still a number of gaps that follow from our implications, such as how to engage students in early stages of a course, which would benefit from further research.

In the future, we want to analyze what students (especially the regular, disheartened, and underachiever groups) wanted to learn, what they learned, what their difficulties were, and how their performance can improve to reach the Achievers (the group with the best performance). To this end, we want to work on personalizing gamified activities for the regular, disheartened, and underachiever groups to improve their learning performance.

REFERENCES

[1] A. Ahmed and M. J. D. Sutton, "Gamification, serious games, simulations, and immersive learning environments in knowledge management initiatives," *World J. Sci., Technol. Sustain. Develop.*, vol. 14, nos. 2–3, pp. 78–83, Apr. 2017.

[2] K. Aleksic-Maslac, B. Sinkovic, and P. Vranesic, "Influence of gamification on student engagement in education," *Int. J. Educ. Learn. Syst.* vol. 2, pp. 76–82, May 2017.

[3] A. Antonaci, F. M. Dagnino, M. Ott, F. Bellotti, R. Berta, A. De Gloria, E. Lavagnino, M. Romero, M. Usart, and I. Mayer, "A gamified collaborative course in entrepreneurship: Focus on objectives and tools," *Comput. Hum. Behav.*, vol. 51, pp. 1276–1283, Oct. 2015.

[4] D. D. Armier, Jr., C. E. Shepherd, and S. Skrabut, "Using game elements to increase student engagement in course assignments," *College Teach.*, vol. 64, no. 2, pp. 64–72, Apr. 2016.

[5] P. Bajdor and L. Dragolea, "The gamification as a tool to improve risk management in the enterprise," *Annales Univ. Apulensis, Ser. Oeconomica*, vol. 13, no. 2, p. 574, 2011.

[6] J. Baker, "Constructivist online learning environment survey," in *Handbook of Research on Electronic Surveys and Measurements*. Hershey, PA, USA: IGI Global, 2007, pp. 299–301.

[7] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "So fun it hurts—Gamifying an engineering course," in *Proc. Int. Conf. Augmented Cognition*. Berlin, Germany: Springer, 2013, pp. 639–648.

[8] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "Gamification for smarter learning: Tales from the trenches," *Smart Learn. Environ.*, vol. 2, no. 1, p. 10, Dec. 2015.

[9] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "Early prediction of student profiles based on performance and gaming preferences," *IEEE Trans. Learn. Technol.*, vol. 9, no. 3, pp. 272–284, Jul. 2016.

[10] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "Studying student differentiation in gamified education: A long-term study," *Comput. Hum. Behav.*, vol. 71, pp. 550–585, Jun. 2017.

[11] G. Barata, S. Gama, J. A. P. Jorge, and D. J. V. Gonçalves, "Relating gaming habits with student performance in a gamified learning experience," in *Proc. 1st ACM SIGCHI Annu. Symp. Comput.-Hum. Interact. Play*, Oct. 2014, pp. 17–25.

[12] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit MUDs," *J. MUD Res.*, vol. 1, no. 1, p. 19, 1996.

[13] R. A. Bartle, *Designing Virtual Worlds*. Indianapolis, IN, USA: New Riders, 2004.

[14] S. Barzilai and I. Blau, "Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences," *Comput. Educ.*, vol. 70, pp. 65–79, Jan. 2014.

[15] C. Bateman and R. Boon, *21st Century Game Design* (Game Development Series). Newton, MA, USA: Charles River Media, 2005.

[16] F. Bellotti, R. Berta, A. De Gloria, E. Lavagnino, A. Antonaci, F. M. Dagnino, and M. Ott, "A gamified short course for promoting entrepreneurship among ICT engineering students," in *Proc. IEEE 13th Int. Conf. Adv. Learn. Technol.*, Jul. 2013, pp. 31–32.

[17] P. Brauner, A. C. Valdez, U. Schroeder, and M. Ziefle, "Increase physical fitness and create health awareness through exergames and gamification," in *Proc. Int. Conf. Hum. Factors Comput. Informat.* Berlin, Germany: Springer, 2013, pp. 349–362.

[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[19] P. Buckley and E. Doyle, "Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market," *Comput. Educ.*, vol. 106, pp. 43–55, Mar. 2017.

[20] D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. New York, NY, USA: Ravenio Books, 2015.

[21] Ü. Çakıroğlu, B. Başıbüyük, M. Güler, M. Atabay, and B. Y. Memiş, "Gamifying an ICT course: Influences on engagement and academic performance," *Comput. Hum. Behav.*, vol. 69, pp. 98–107, Apr. 2017.

[22] P.-Z. Chen, T.-C. Chang, and C.-L. Wu, "Effects of gamified classroom management on the divergent thinking and creative tendency of elementary students," *Thinking Skills Creativity*, vol. 36, Jun. 2020, Art. no. 100664.

[23] K. Davis, H. Sridharan, L. Koepke, S. Singh, and R. Boiko, "Learning and engagement in a gamified course: Investigating the effects of student characteristics," *J. Comput. Assist. Learn.*, vol. 34, no. 5, pp. 492–503, Oct. 2018.

[24] L. de-Marcos, A. Domínguez, J. Saenz-de-Navarrete, and C. Pagés, "An empirical study comparing gamification and social networking on e-learning," *Comput. Educ.*, vol. 75, pp. 82–91, Jun. 2014.

[25] M. R. de Almeida Souza, K. F. Constantino, L. F. Veado, and E. M. L. Figueiredo, "Gamification in software engineering education: An empirical study," in *Proc. IEEE 30th Conf. Softw. Eng. Educ. Training (CSEE&T)*, Nov. 2017, pp. 276–284.

[26] R. Dellos, "Kahoot! A digital game resource for learning," *Int. J. Instruct. Technol. Distance Learn.*, vol. 12, no. 4, pp. 49–52, 2015.

[27] P. Denny, "The effect of virtual achievements on student engagement," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2013, pp. 763–772.

[28] S. Deterding, R. Khaled, L. E. Nacke, and D. Dixon, "Gamification: Toward a definition," in *Proc. CHI Gamification Workshop*, Vancouver, BC, Canada, 2011, pp. 1–4.

[29] D. Dicheva, C. Dichev, G. Agre, and G. Angelova, "Gamification in education: A systematic mapping study," *J. Educ. Technol. Soc.*, vol. 18, no. 3, pp. 75–88, 2015.

[30] D. Dicheva, K. Irwin, and C. Dichev, "OneUp: Supporting practical and experimental gamification of learning," *Int. J. Serious Games*, vol. 5, no. 3, pp. 5–21, Sep. 2018.

[31] D. Dicheva, K. Irwin, and C. Dichev, "OneUp: Engaging students in a gamified data structures course," in *Proc. 50th ACM Tech. Symp. Comput. Sci. Educ.*, Feb. 2019, pp. 386–392.

[32] A. Domínguez, J. Saenz-de-Navarrete, L. de-Marcos, L. Fernández-Sanz, C. Pagés, and J.-J. Martínez-Herráiz, "Gamifying learning experiences: Practical implications and outcomes," *Comput. Educ.*, vol. 63, pp. 380–392, Apr. 2013.

[33] Z. Fitz-Walter, P. Wyeth, D. Tjondronegoro, and B. Scott-Parker, "Driven to drive: Designing gamification for a learner logbook smartphone application," in *Proc. 1st Int. Conf. Gameful Design, Res., Appl.*, 2013, pp. 42–49.

[34] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, no. 2, pp. 131–163, Nov. 1997.

[35] M. Gachkova, M. Takev, and E. Somova, "Learning and assessment based on gamified e-course in moodle," *Math. Inform.*, vol. 61, no. 5, pp. 444–454, 2018.

[36] C. V. Gipps, *Beyond Testing: Towards a Theory of Educational Assessment*. London, U.K.: Psychology Press, 1994, pp. 1–193.

[37] L. Hakulinen and T. Auvinen, "The effect of gamification on students with different achievement goal orientations," in *Proc. Int. Conf. Teach. Learn. Comput. Eng.*, Apr. 2014, pp. 9–16.

[38] M. D. Hanus and J. Fox, "Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance," *Comput. Educ.*, vol. 80, pp. 152–161, Jan. 2015.

[39] J. Hartigan and M. Wong, *A K-Means Clustering Algorithm*. Milton Keynes, U.K.: New Haven, 1979.

[40] M. Jakubowski, "Gamification in business and education—Project of gamified course for university students," in *Proc. Annu. ABSEL Conf. Develop. Bus. Simulation Experiential Learn.*, 2014.

[41] M. Jurgelaitis, V. Drungilas, and L. Čeponienė, "Gamified moodle course for teaching UML," *Baltic J. Mod. Comput.*, vol. 6, no. 2, pp. 119–127, 2018.

[42] K. M. Kapp, *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. Hoboken, NJ, USA: Wiley, 2012.

[43] A. Khan, O. Egbue, B. Palkie, and J. Madden, "Active learning: Engaging students to maximize learning in an online course," *Electron. J. E-Learn.*, vol. 15, no. 2, pp. 107–115, 2017.

[44] S. Kiesler, R. E. Kraut, K. R. Koedinger, V. Aleven, and B. M. Mclaren, "Gamification in education: What, how, why bother," *Acad. Exchange Quart.*, vol. 15, no. 2, pp. 1–5, 2011.

[45] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic Regression*. Berlin, Germany: Springer, 2002.

[46] E. Lavoué, B. Monterrat, M. Desmarais, and S. George, "Adaptive gamification for learning environments," *IEEE Trans. Learn. Technol.*, vol. 12, no. 1, pp. 16–28, Jan./Mar. 2018.

[47] K. M. Mbabu, R. O. Oboko, and S. Kamunya, "An adaptive gamification tool for e-learning platform," *Open J. Inf. Technol.*, vol. 2, no. 2, pp. 41–52, Dec. 2019, doi: 10.32591/coas.ojit.0202.03041m.

[48] C. I. Muntean, "Raising engagement in e-learning through gamification," in *Proc. 6th Int. Conf. Virtual Learn. (ICVL)*, 2011, pp. 323–329.

[49] I. B. Myers and M. H. McCaulley, *Myers-Briggs Type Indicator: MBTI*. Palo Alto, CA, USA: Consulting Psychologists Press, 1988.

[50] A. H. Nabizadeh, A. M. Jorge, and J. P. Leal, "Long term goal oriented recommender systems," in *Proc. 11th Int. Conf. Web Inf. Syst. Technol. (WEBIST)*, Lisbon, Portugal, vol. 1, 2015, pp. 552–557.

[51] A. H. Nabizadeh, A. Mário Jorge, and J. Paulo Leal, "RUTICO: Recommending successful learning paths under time constraints," in *Proc. Adjunct Publication 25th Conf. User Modeling, Adaptation Pers.*, Jul. 2017, pp. 153–158.

[52] A. H. Nabizadeh, D. Gonçalves, S. Gama, J. Jorge, and H. N. Rafsanjani, "Adaptive learning path recommender approach using auxiliary learning objects," *Comput. Educ.*, vol. 147, Apr. 2020, Art. no. 103777.

[53] A. H. Nabizadeh, J. P. Leal, H. N. Rafsanjani, and R. R. Shah, "Learning path personalization and recommendation methods: A survey of the state-of-the-art," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113596.

[54] A. H. Nabizadeh, A. M. Jorge, and J. P. Leal, "Estimating time and score uncertainty in generating successful learning paths under time constraints," *Expert Syst.*, vol. 36, no. 2, Apr. 2019, Art. no. e12351.

[55] L. E. Nacke, C. Bateman, and R. L. Mandryk, "BrainHex: Preliminary results from a neurobiological gamer typology survey," in *Proc. Int. Conf. Entertainment Comput.* Berlin, Germany: Springer, 2011, pp. 288–293.

[56] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?" in *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit.* Berlin, Germany: Springer, 2012, pp. 154–168.

[57] Y. Pan, D. Schwartz, and S. Mishra, "Gamified digital forensics course modules for undergraduates," in *Proc. IEEE Integr. STEM Educ. Conf.*, Mar. 2015, pp. 100–105.

[58] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[59] T. Plomp, "Educational design research: An introduction," in *Educational Design Research*. Enschede, The Netherlands: Netherlands Inst. for Curriculum Develop., 2013, pp. 11–50.

[60] H. N. Rafsanjani, A. Ghahramani, and A. H. Nabizadeh, "ISEA: IoT-based smartphone energy assistant for prompting energy-aware behaviors in commercial buildings," *Appl. Energy*, vol. 266, May 2020, Art. no. 114892.

[61] A. H. N. Rafsanjani, "A long term goal recommender approach for learning environments," Faculdade de Ciências da Universidade do Porto, Porto, Portugal, Tech. Rep. 101504217, 2018.

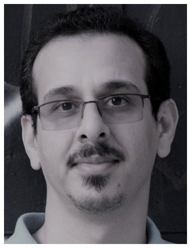
[62] A. H. N. Rafsanjani, N. Salim, A. R. Aghdam, and K. B. Fard, "Recommendation systems: A review," *Int. J. Comput. Eng. Res.*, vol. 3, no. 5, pp. 47–52, 2013.

[63] L. Rokach and O. Maimon, "Clustering methods," in *Data Mining and Knowledge Discovery Handbook*. Berlin, Germany: Springer, 2005, pp. 321–352.

[64] A. Todd, "Why gamification is malarkey," *Morning Watch, Educ. Social Anal.*, vol. 44, nos. 1–2, pp. 1–7, Mar. 2017.

[65] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inf. Sci.*, vol. 44, no. 1, pp. 48–59, Feb. 2018.

[66] Youtopia. (2009). *Youtopia Website*. Accessed: Jul. 3, 2020. [Online]. Available: http://www.youtopia.com/info



AMIR HOSSEIN NABIZADEH received the M.Sc. degree in information technology management from the University Technology Malaysia (UTM), in 2013, and the Ph.D. degree in computer science from the University of Porto, in 2018. He is currently a Postdoctoral Researcher with the INESC-ID, Instituto Superior Técnico, University of Lisbon. His main research interests include machine learning, data mining, recommender systems, e-learning, and MOOC.

JOAQUIM JORGE (Senior Member, IEEE) received the Ph.D. degree from the Rensselaer Polytechnic Institute, in 1995. He coordinates the Interaction and Graphics Research Group, INESC-ID. He is currently a Computer Graphics and Multimedia Professor with the School of Engineering, Instituto Superior Técnico, Universidade de Lisboa. His research interests include multimodal user interfaces, VR/AR/MR, 3D visualization, and advanced learning techniques. He is a Fellow of the Eurographics Association. He is a Distinguished Member and a Speaker of the ACM. He is the Editor-in-Chief of the *Computers and Graphics* and *Graphics and Visual Computing* journals.

SANDRA GAMA is currently a Researcher with the Visualization and Multimodal Interfaces Group, INESC-ID. She is a Professor of computer science with the Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal. Her research interests include information visualization, multimodal user interfaces, and human–computer interaction.

DANIEL GONÇALVES is currently a Researcher with the Visualization and Multimodal Interfaces Group, INESC-ID. He is a Professor of computer science with the Instituto Superior Técnico (IST/UL), Lisbon, Portugal. His research interests include information visualization, personal information management, and human–computer interaction and accessibility. He is a member of the ACM and the Portuguese Computer Graphics Group (the National Eurographics Chapter).