Interrelated analysis of interaction, sequential patterns and academic achievement in online learning

Denizer Yildirim
Faculty of Open and Distance Education, Ankara University

Yasemin Usluel
Department of Computer Education and Instructional Technology, Hacettepe University

This study aimed to examine the behaviour of learners across a whole system and in various courses to reveal the interrelation between learners' system interaction, age, programme features and course design. We obtained data from the system logs of 1,634 learners enrolled in distance learning programmes. We performed hierarchical clustering analysis to describe system interactions; then, we carried out a sequential pattern analysis to examine navigational behaviours by clusters. The results showed that the system interactions (e.g., content, live lesson, assignment, exam, discussion) across the whole system differ by age and programme. The behaviour profiles of the learners changed when different course designs were presented. Learners who interacted more with any component (e.g., live lesson or content) according to their needs were more successful than those with limited interaction and assessment-oriented (those with limited interactions outside of the assignment). In an information and communication technology course, learners whose system interactions were sufficient to receive rewards were more likely to succeed. The sequential pattern analysis showed that the assessment-oriented cluster interacted with the assignment in the midterm weeks; the award-oriented cluster interacted with the content or completed their assignment and received an award. Consequently, it is difficult to determine or generalise the intervention unless the system, programme and course design features are standard.

Implications for practice or policy:

- Course designers can use the assessment activities or motivation factors such as awards to increase students' system interactions.
- Course designers should not determine or generalise interventions unless the system, programme and course design are standard.
- Researchers should not only focus on data but also consider the contextual characteristics of data.

Keywords: online learning, learner profile, age and programme features, course design, cluster analysis, sequential patterns

Introduction

Based on actual usage data in online learning environments, meaningful and realistic results can be obtained regarding analysing learner behaviour (Cantabella et al., 2018; Soffer et al., 2017). However, the effectiveness of distance education is a multidimensional process. Therefore, more holistic approaches are needed, similar to recent studies dealing with different aspects of distance learning (Choudhury & Pattnaik, 2020; Zhang et al., 2019). Studies have analysed learners' behaviours (participation, interaction and navigation patterns) in online learning environments. Interaction and participation patterns differ according to the discipline of study (Soffer et al., 2017), course structure (Kahan et al., 2017), target audience, platform, culture (Cohen et al., 2019; Zhong et al., 2017), educational level and course content (Cerezo et al., 2016; Soffer et al., 2019), navigational patterns (Lin et al., 2017; Shih, 2018). Therefore, the insights obtained from these studies are not generalisable for increasing the effectiveness of distance education. Because each distance education system contains specific features, it is not easy to apply recommendations in different contexts.

When analysing learner behaviours in online learning environments, it is necessary to determine behavioural patterns by considering the relationships between several dimensions (e.g., age, programme features, course designs). First, age and the programme or discipline may be another factor in determining online behaviour. Since learners in distance education are between the ages of 18 and 70 (Bravo-Agapito et al., 2021), the behavioural patterns of learners in different age groups are also different. Bravo-Agapito et al. described the profile of students based on learners' interactions with the Moodle learning management system (LMS). Profiles in Bravo-Agapito et al.'s research show learners' groups who behave differently. When the characteristics of groups (n = 5) were examined, some groups were characterised as having a lower age (e.g., groups 1 and 2). For example, the average age of group 1 was found to be lower than that of group 3 (group 1: n = 214, X_{age} = 35.03; group 3: n = 87, X_{age} = 40.05). Also, group 1 was composed mainly of Computer Science, Journalism and Psychology students, and group 3 consisted of Civil Engineering students. Therefore, it can be argued that programme features and age may be determining factors for learners' profiles.

Moreover, we can begin to understand learners' behaviours and performance indicators when we examine how students interact with the system or course designs (Nguyen et al., 2017; Nguyen et al., 2018; Rienties & Jones, 2019). Different course designs make it difficult to improve student learning experiences in online learning environments (Botelho et al., 2019; Gkontzis et al., 2019; Sandoval et al., 2019). For example, Sandoval et al. showed that predictive power changed according to the courses chosen. Botelho et al. tried to predict students' attrition in an environment where they used clues to solve mathematical problems. Gkontzis et al. attempted to predict learners' attrition in a discussion forum, content (e.g., videos) and quiz components. Therefore, the variables related to course designs may be another reason for changing learners' behaviour.

This study aimed to examine learners' behaviour across the whole system and in various courses to reveal the interrelation between students' interaction, multiple variables (such as age, the registered programme) and course design. For this purpose, we applied holistic educational data mining (EDM) with clustering and sequential pattern analysis (SPA) on learners' behaviour. The research questions (RQs) are presented below:

1. How are learner interactions clustered across the whole system?
2. Do learner interactions across the whole system vary according to age and distance education programmes?
3. How are learner interactions clustered in different online lessons?
    3.1. Is there a significant difference in the academic achievement of learners clustered in different online lessons?
    3.2. How do the interactions of learners vary by week?
    3.3. What navigational behaviours do the learners demonstrate?

Educational data in the current study

While analysing learner behaviours using EDM, various educational data based on self-report or observation may be considered (Han & Hellis, 2020). Observational data include learning and teaching analytics and assessment analytics (Han & Hellis, 2020; Ifenthaler & Yau, 2020). These variables consist of analytics such as frequency of usage, participation mode, time-related metrics or formative or summative assessment scores (Agudo-Peregrina et al., 2014; Bravo-Agapito et al., 2021). In this context, the issue of which variables to consider is essential. Sandoval et al. (2018) stated that selecting low-cost variables is more practical and helpful in producing more generalisable results. For this purpose, this study referenced the frequency of usage and the participation mode. Time-related metrics (e.g., time spent on a task, individual work time, time spent on an activity) were not calculated because the calculation is labour-intensive (Bravo-Agapito et al., 2021; Sandoval et al., 2018).

The frequency of usage indicates the number of times a behaviour is repeated. Many researchers base their analysis on the frequency of usage when examining learners' interactions and participation (Cerezo et al., 2016; Cohen et al., 2019; Kahan et al., 2017; Soffer et al., 2019). The frequency of usage is calculated according to the frequency of action within a certain time frame (e.g., number of views, number of submissions). Participation mode handles behaviour actively and passively. Some studies have focused on participation mode in their analyses by classifying learner actions as passive or active based on learners' interactions in the system (e.g., Gavilanes-Sagnay et al., 2018). Agudo-Peregrina et al. (2014) found that active participation (e.g., sending a message, submitting homework or assignments) of the learners is lower than their passive participation (e.g., reading messages, reading homework or assignments).

Profiles of learner behaviours
Some studies have shown that behaviour patterns change according to the way participation and interaction are handled (e.g., access to learning resources, effort, or procrastination) in courses with similar activities (e.g., Bravo-Agapito et al., 2021; Cerezo et al., 2016; Cohen et al., 2019). Even when interaction and participation are evaluated within the same context (e.g., learners' activities and environment), differences in behavioural patterns can be observed (Kahan et al., 2017; Van den Beemt et al., 2018). Although these results provide information about learner behaviour, they do not contain enough details regarding the learner's interaction with course design, age and the registered programmes in distance online learning. In this context, this research examined online courses that provide different learning activities in course design to demonstrate the importance of elements related to course design on learners' behaviour using the EDM approach.

Navigational paths and SPA

SPA may be used to analyse learners' navigation paths across a system (Campagni et al., 2012; Dráždilová et al., 2010). Some studies have focused on the potential uses of SPA (Munk & Drlik, 2011; Zhou, 2010), whereas some have looked into the design of the system or were based on repetitive sequential patterns (Wang et al., 2012; Wong et al., 2019), and others have analysed sequential interactions by labelling learners as successful or unsuccessful (Lin et al., 2017; Shih, 2018). Most studies have demonstrated sequential patterns in learner behaviour in an activity (Lin et al., 2017; Shih, 2018; Wang et al., 2012). However, in Wong et al.'s study, the order of learners in the video, discussion and assessment was analysed according to whether the students' self-regulated learning prompt videos were watched or not. Therefore, by examining the order of learners' learning activities according to different categories, more course design information may be obtained. The current study examined sequential patterns based on all activities rather than a single activity and classified learner behaviour in a single activity as active or passive according to participation mode.

Methods

Participants

The participants consisted of three groups chosen from 1,634 learners enrolled in distance education programmes at Ankara University in Turkey. The participants were selected according to the RQ as we wanted to define the learner behaviours across the whole system and in various lessons. The first group (all students; n = 1634) was chosen to answer RQ1 and RQ2; the second group (students taking three different online lessons (MD: Medical Documentation, A: Anatomy, ICT: Information and Communication Technologies) together in the same programme; n = 95) was chosen to answer RQ3 and RQ3.1; and the third group (students taking ICT; n = 112) was chosen to answer RQ3.2 and RQ3.3 (see Table 1).

Table 1
Characteristics of study groups determined for RQs

Study group	RQ	Programme	Age				Lessons
			18–24	25–31	32–38	39+	
All students (n = 1634)	RQ1 RQ2	A	30.2%	51.7%	12.8%	5.4%	All lessons
		B	76.1%	12.4%	7.1%	4.4%	
		C	64.2%	22.4%	9.9%	3.4%	
		D	10.0%	28.7%	45.9%	15.4%	
		E	64.9%	18.9%	14.5%	1.8%	
		F	56.1%	20.7%	11.0%	12.2%	
		Total	50.3%	25.8%	16.9%	7.1%	
Student taking three different online lessons together in a programme (n = 95)	RQ3 RQ3.1	E	72.6%	9.5%	16.8%	1.1%	Medical Documentation (MD)
							Anatomy (A)
							Information and Communication Technologies (ICT)
Students taking ICT course (n = 112)	RQ3.2 RQ3.3	E	71.4%	12.5%	15.2%	0.9%	ICT

Note. Programmes: A. Judicial Services Support Personnel DEP (Distance Education Programme); B. Banking and Insurance Programme DEP; C. Computer Programming DEP; D. Divinity Diploma Upgrade DEP; E. Medical Documentation Support Personnel DEP; F. Tourism and Hotel Management DEP.

Course designs

Online courses are structured according to the "Procedures and Principles Regarding Distance Education in Higher Education Institutions" (Council of Higher Education, 2020). Accordingly, various tools are provided for all instructors to create multiple activities such as content transfer, evaluation or assessment, discussion, collaborative work. The students are responsible for learning. In all programmes, the duration of all lessons is 14 weeks. The midterm and final times are the same. Although the final exams in all programmes were supervised paper-and-pencil tests, the midterms differed in each programme (either as homework, an electronic exam or a supervised multiple-choice test). The criteria for the assessment of lesson success used are homework or electronic midterm exam scores (20%) and supervised paper-and-pencil test scores (80%) in associate degree programmes. In bachelor's degree completion programmes, criteria used are supervised midterm scores (30%) and supervised paper-and-pencil test scores (70%).

In the system considered in the research, some courses and programmes are similar and different from each other in terms of content transfer, assessment, discussion and rewarding activities. The content transfer is provided in all the courses covered in the research using similar tools such as sharable content object reference model (SCORM), PDF or presentation (PowerPoint), or live lesson. The instructors are obliged to conduct at least a synchronous live lesson once a week in all courses. Synchronous live lessons are held through web conferences on the date and time planned by the instructors. Each live lesson is recorded by instructors and is accessible for learners who cannot attend synchronously. Structured content (SCORM) containing summary information is prepared for each unit of the courses in the whole system. The contents are prepared and uploaded to the system in cooperation with instructional design and subject matter experts (Table 2).

Table 2
Course designs in study groups

Group	Programme	Lesson	Content transfer	Assessment					Discussion	Reward
				FA	SA midterm			SA final		
				Q	EE	A	SMC	SMC		
Group 1 (*n* = 1634)	A	All	SCORM, PDF or presentation Live lesson	-	17	-	17	17	+	-
	B			-	13	13	-	16	+	-
	C			2	9	6	-	15	+	-
	D			-	-	-	16	16	+	-
	E			1	11	1	-	12	+	+
	F			-	4	12	-	16	+	-
Group 2 (*n* = 95)	E	MD		-	+	-	-	+	+	-
		A		+	-	+	-	+	-	-
		ICT		+	-	+	-	+	+	+
Group 3 (*n* = 112)	E	ICT		+	-	+	-	+	+	+

Note. FA: formative assessment; SA: summative assessment; A: assignment: EE: electronic exam; SMC: supervised multiple choice.

For example, when we compare Programmes D and E for Study Group 1, in terms of assessment activities, Programme D has midterm exams with supervised multiple choice (SMC) in all lessons, Programme E in some lessons with electronic exam (EE; *n* = 4) and assignment (*n* = 12) in some lessons. Discussion activities are used by teachers in both programmes for announcement and informational purposes. On the other hand, Programme E uses rewards while Programme D does not use them at all. Moreover, even when considering a single programme (e.g., F), there are differences in design between courses (e.g., A, MD, ICT).

Under the current situation without any intervention, RQ2 compares students' interactions with different designs regarding the activities in courses with the same students (*n* = 95) enrolled. The duration of these lessons is 14 weeks. While EE is applied for the midterm exam in the MD, A and ICT courses assess the midterm exam with an assignment grade. Although the A course uses discussion and rewarding, the ICT course does not (Table 2).

Data collection tools

Data collection tools consisted of the final exams and the system logs. Ethics committee approval was obtained from Hacettepe University Ethics Boards and Commissions (Ethics approval number: 433-1209). Written informed consent in the LMS was also obtained from all participants before the research. The final exam was a multiple-choice paper-and-pen test in a supervised environment. System logs were collected from the logstore_standard_log table (LogT) in the Moodle database. There were 2,148,443 records in the LogT for one semester. The processed data consists of 615,086 (%28.6) records. The data include a total of 77 actions performed by the learners. These actions include logging in, viewing any Moodle component (e.g., forum, assignment, live lesson, message, SCORM), adding, updating, sending and signing. Among the actions, those with zero variance values were discarded.

Data analysis

The data analysis was performed in two stages: pre-processing the data and applying clustering and sequential pattern mining (Figure 1).

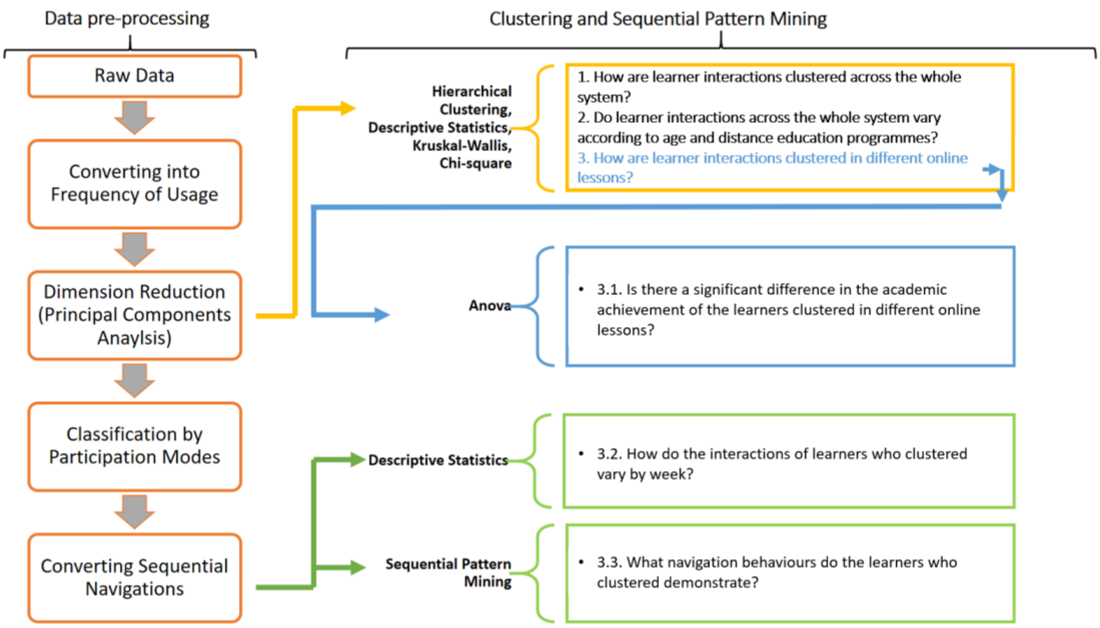


Figure 1. All stages of data analysis

Converting into frequency of usage and dimension reduction

This study utilised SQL Pivot table to convert system logs into the frequency of usage. Then, the system logs of three different study groups (Table 1) were analysed through principal components analysis (PCA) to identify the thematically related variables. Varimax rotation method was used for PCA. Regression scores were recorded by naming variables associated with each other in a single dimension. As a result, the learner interactions across the whole system were grouped under seven dimensions as exam (quiz), assessment, message, live lesson, content and discussion (Tables 3, 4 & 5). The learner interactions in various lessons have been grouped by Study Group 2 (Table 2). The MD course was grouped under three dimensions (live lesson, content and discussion); the A course was grouped under four dimensions (exam (quiz), assessment, content and live lesson); the ICT course was grouped under six dimensions (content, assessment, award, exam (quiz), discussion and live lesson).

Table 3
KMO and Bartlett's test

Kaiser-Meyer-Olkin measure	Bartlett's test of sphericity		
	Approx. chi square	df	Sig.
.835	85999.122	435	.000

Table 4
Total variance explained

Component	Initial eigenvalues			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	10.247	34.155	34.155	6.951	23.170	23.170
2	5.000	16.668	50.823	5.286	17.620	40.791
3	3.443	11.476	62.299	4.640	15.468	56.259
4	2.280	7.599	69.898	3.603	12.009	68.268
5	2.049	6.828	76.727	2.394	7.980	76.248
6	1.698	5.661	82.388	1.842	6.140	82.388

Table 5
Rotated component matrix

Actions	Exam (Quiz)	Assessment	Message	Live lesson	Content	Discussion
awarded badge	0.897	0.152	-0.005	0.108	-0.058	0.022
leveledup user	0.895	0.155	-0.006	0.104	-0.012	0.018
updated course module completion	0.878	0.213	0.054	0.119	-0.035	0.025
viewed attempt summary	0.844	0.138	0.03	0.084	0.403	0.018
started attempt	0.841	0.129	0.027	0.085	0.424	0.016
submitted attempt	0.827	0.125	0.03	0.079	0.445	0.016
reviewed attempt	0.81	0.137	0.039	0.086	0.429	0.02
viewed attempt	0.808	0.175	0.035	0.08	0.423	0.013
accepted statement	0.758	0.163	-0.01	0.077	-0.049	0.03
submitted assessable	0.162	0.961	0.075	0.087	0.064	0.024
viewed submission form	0.222	0.925	0.097	0.129	0.044	0.049
created submission	0.16	0.924	0.071	0.081	0.074	0.01
uploaded assessable	0.253	0.913	0.066	0.101	0.046	0.162
viewed submission status	0.106	0.861	0.202	0.15	0.1	0.076
updated submission	0.26	0.706	0.036	0.103	-0.021	0.05
deleted message	-0.016	0.097	0.939	0.098	0.001	-0.061
viewed message	0.061	0.27	0.884	0.114	0.005	-0.03
sent message	0.082	0.191	0.858	0.117	0.031	-0.001
blocked message contact	-0.001	0.012	0.843	0.06	0.002	0.197
viewed user profile	0.006	-0.033	0.834	0.091	0.001	-0.05
unblocked message contact	-0.018	-0.004	0.824	0.041	-0.001	0.157
view adobeconnect	0.077	0.063	0.025	0.957	0.004	0.007
meeting adobeconnect join	0.042	0.106	0.017	0.83	0.026	-0.011
recording adobeconnect view	0.153	-0.044	-0.042	0.806	-0.026	-0.015
submitted status	0.247	0.055	0.014	0.043	0.932	0.049
submitted scoreraw	0.17	0.086	-0.002	0.061	0.886	0.006
launched SCORM	0.393	0.054	0.008	0.047	0.872	0.019
updated post	-0.013	0.073	0.023	0.027	0.013	0.779
viewed course module instance list	-0.08	0.039	0.157	0.125	0.054	0.757
created discussion	0.286	0.185	0.019	0.032	0.002	0.702

Note. Shading shows item loads collected in one dimension.

Classification by participation modes

Analysing system interactions based on participation mode and frequency of usage would potentially provide more information on learner behaviours. For this, the actions were classified according to participation modes based on the classification in the literature. The classification includes the appropriate actions related to each other due to the PCA in the ICT course in the third study group. Besides, this study considered viewing the component about the award as passive participation and getting a badge upon completing the necessary tasks as active participation.

Sequential navigation behaviours
Different approaches are used to perform data pre-processing in SPA (Munk & Drlik, 2011; Zhou, 2010). Zhou (2010) put forward two approaches: user-based and session-based. These approaches have positive and negative aspects compared to each other. In a session-based process, every user engages with a different number of sessions. For example, let us assume that user A navigated differently in each of the 15 sessions, while user B navigated similarly in each of the five sessions. Accordingly, the order pattern of user B has a high rate in their sessions, while it has a low rate in the group. In the user-based approach, a single series is obtained from each learner's actions over the entire period according to time. Therefore, it is more appropriate to determine the proportions of those who make a specific sequence of actions among a group of learners. In this research, a user-based approach was used while analysing sequential patterns.

Hierarchical clustering and SPA
This study carried out a hierarchical clustering analysis using the factor loadings obtained from the PCA in data pre-processing. The clustering method was conducted based on Ward's method (Murtagh & Legendre, 2014; Sun et al., 2018) and Euclidean distance by using SPSS. After the learners were clustered, the average factor loading for each cluster was calculated and presented as graphics. The primary purpose of this method was to identify the frequently used sequential item sets in a data set (Dráždilová et al., 2010). The sequential patterns that appear more often than others are determined based on support value or support ratio (Campagni et al., 2012; Dráždilová et al., 2010). This study utilised the Weka 3 data mining software in Java and the generalised sequential patterns algorithm to perform SPA.

Findings

How are learner interactions clustered across the whole system?

Learners were clustered according to their regression scores calculated for each student as a result of the PCA. The results showed three different interaction clusters across the whole system by component scores (Table 6).

Table 6
Whole-system interactions according to the interaction clusters

Clusters		Live lesson	Assignment	Exam (Quiz)	Discussion	Content	Message
CS1 (*n* = 979; 59.9%)	Mean	-0.468	-0.304	-0.083	-0.190	-0.044	-0.013
	SD	0.324	0.178	0.143	0.213	0.225	0.079
CS2 (*n* = 383; 23.4%)	Mean	0.132	1.118	0.388	0.599	0.126	0.157
	SD	0.972	1.578	1.999	1.909	2.020	2.047
CS3 (*n* = 272; 16.7%)	Mean	1.499	-0.480	-0.250	-0.160	-0.021	-0.173
	SD	1.114	0.267	0.142	0.250	0.244	0.170
Total (*n* = 1634; 100%)	Mean	0.000	0.000	0.000	0.000	0.000	0.000
	SD	1	1	1	1	1	1

Note. CS = cluster across the whole system.

The homogeneity assumption was taken into consideration to test the significance of the difference between the clusters. As the assumption of homogeneity of variances was not met, the Kruskal-Wallis test was performed. It showed that there is a significant difference between the clusters in terms of all components on the interaction ($\chi2_{\text{live lesson}}(2,1634) = 693.579$, $\chi2_{\text{assignment}}(2,1634) = 527.168$, $\chi2_{\text{exam}}(2,1634) = 378.463$, $\chi2_{\text{discussion}}(2,1634) = 295.298$, $\chi2_{\text{message}}(2,1634) = 527.168$, $p < .01$; $\chi2_{\text{content}}(2,1634) = 8.983$, $p < .05$). The significance of the difference between the groups was calculated with Dunnett C (Table 7).

Table 7
Multiple comparisons by Dunnett C

I	J	Dependent	I-J	SE
CS1	CS3	Assignment	0.177*	0.017
		Exam (Quiz)	0.167*	0.010
		Message	0.160*	0.011
CS2	CS1	Live lesson	0.601*	0.051
		Assignment	1.421*	0.081
		Exam (Quiz)	0.471*	0.102
		Discussion	0.789*	0.098
	CS3	Assignment	1.598*	0.082
		Exam (Quiz)	0.638*	0.102
		Discussion	0.759*	0.099
		Message	0.330*	0.105
CS3	CS1	Live lesson	1.968*	0.068
	CS2	Live lesson	1.367*	0.084

Note. CS = cluster across the whole system; *$p < .05$

Table 7 demonstrates that the scores of the CS1 were significantly lower than that of CS2; the scores of CS1 were significantly lower than that of CS3 in the component of live lesson, while CS1 had noticeably higher scores in the components of assignment and exam compared to CS3. CS2 had the highest scores in all components except the component of live lesson; its score was significantly lower than CS3 in the component of live lesson. CS3 had significantly higher scores than CS1 and CS2 in the component of live lesson but its scores in the other components were low. As a result, we named CS1 as user cluster with limited interaction, CS2 as assessment-oriented user cluster and CS3 as live lesson–oriented user cluster (Figure 2).

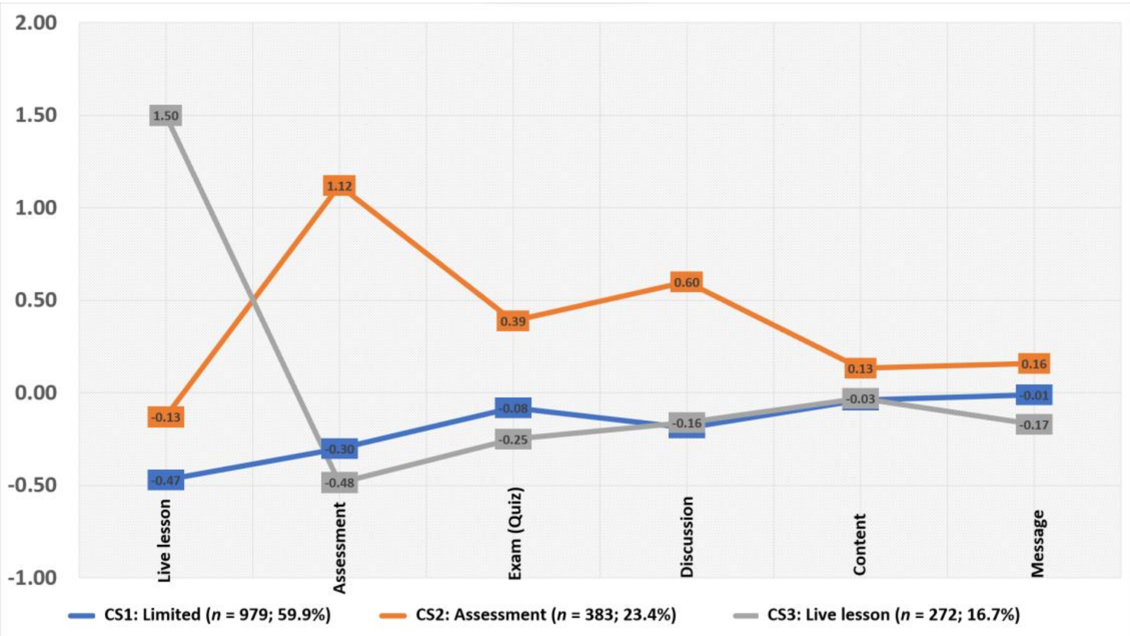


Figure 2. The interaction clusters across the whole system

Do learner interactions vary according to age and distance education programmes?

This research problem has been answered to examine the differences by programme and age in components scores (live lesson, assignment, exam (quiz), discussion, content, message). As the homogeneity assumption was not met for each component, the Kruskal-Wallis test was performed (Table 8).

Table 8
Kruskal-Wallis test statistics

Variables	Components					
	Live lesson	Assignment	Exam (Quiz)	Discussion	Content	Message
Age	55.54*	50.85*	16.45*	18.31*	0.99	50.85*
Programme	35.10*	914.69*	308.21*	163.13*	141.95*	99.24*

Note. $^{*}p < .01$

Table 8 shows that the levels of interaction (excluding content by age) significantly (p = .000) differed according to the different distance education programmes and age (for programme: $\chi2_{\text{live lesson}} = 35.10$, $\chi2_{\text{assignment}} = 914.69$, $\chi2_{\text{exam}} = 308.21$, $\chi2_{\text{discussion}} = 163.13$, $\chi2_{\text{content}} = 141.95$, $\chi2_{\text{message}} = 99.24$, $p = .000$ and for age: $\chi2_{\text{live lesson}} = 55.54$, $\chi2_{\text{assignment}} = 50.85$, $\chi2_{\text{exam}} = 16.45$, $\chi2_{\text{discussion}} = 18.31$, $\chi2_{\text{message}} = 50.85$, $p < .01$; $\chi2_{\text{content}} = 0.99$, $p = .802$).

In order to better understand the differences between age and programme categories, radar graphics were used for both variables (Figure 3). In Figure 3, live lesson interactions of the 32–38 and 39+ age groups differed significantly compared to those in the 18–24 and 25–31 age groups. The assignment and quiz interactions of the 18–24 age group were significantly higher than the other groups. In this context, the learners with advanced age may be live lesson–oriented, while the younger age group may be assessment-oriented. Moreover, younger learners had higher scores in terms of discussion (mean difference: $\bar{x}_{18\text{–}24}\text{–}\bar{x}_{32\text{–}38} = .175$) and messaging (mean difference: $\bar{x}_{18\text{–}24}\text{–}\bar{x}_{39+} = .107$, $p < .01$). But the difference was not as high as in the live lesson, assignment, exam (quiz) components. Age did not show a significant difference in content interaction.

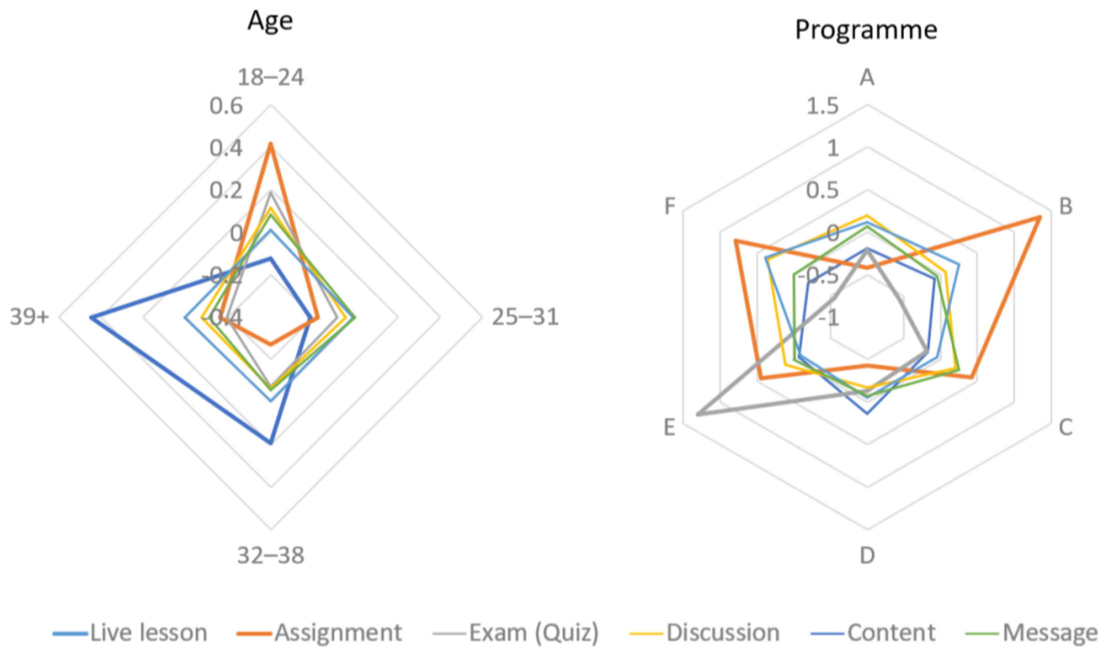


Figure 3. Interactions (scores of components) by age and by programme

The interactions across the whole system are examined according to the programme, while Programmes B, C, E and F had notably higher assignment interactions than other programmes. However, Programme E's exam interactions were notably higher than that of other programmes but not significant. Programme D had higher live lesson interaction compared to other programmes. In this context, assessment-oriented interaction seems to be more dominant in some programmes (e.g., B, F) and live lesson–oriented interaction in other programmes (e.g., D). In addition, it is noteworthy that in Programmes B and F, which had high assignment interactions, the content and discussion were high, and in Programmes E and C, the message and discussion interactions.

This study confirmed that age and programme might be a factor as one of the reasons for the difference in components related to the interaction. However, there may be some cases that result from course design other than age and programme. For example, although there is no significant difference by the age for Programmes B, C and E (df (2,570); F = .018; p = .982), the profiles of these programmes were not similar in terms of all system interactions (Figure 4).

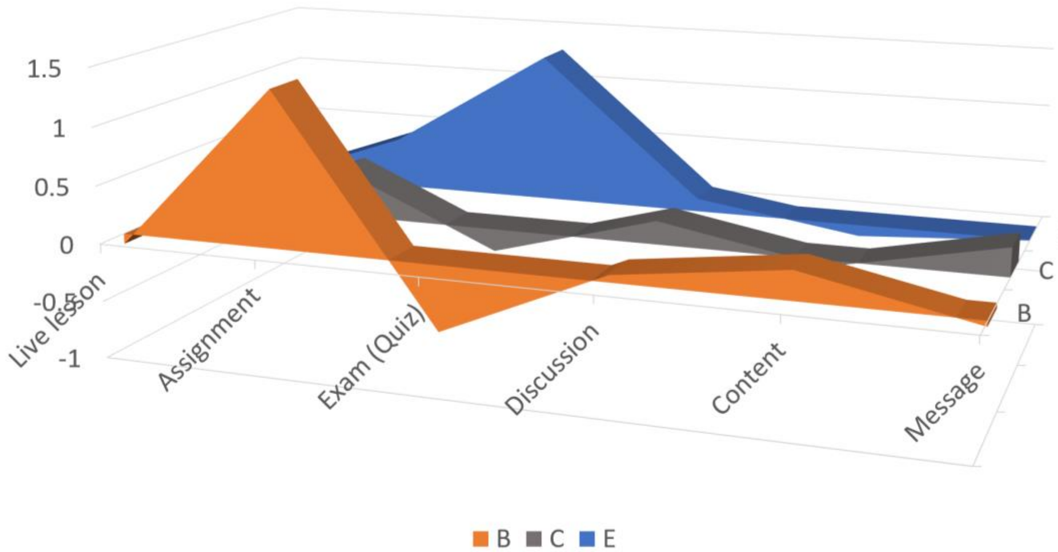


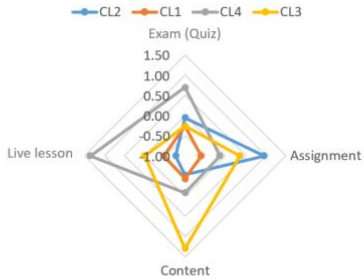
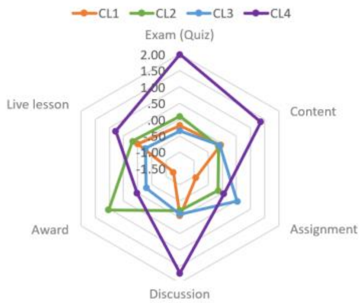
Figure 4. Interactions by Programmes B, C and E

Figure 4 shows that Programmes B, C and E (there is no difference by age) had different profiles in terms of interaction. For example, assignment interaction was higher in Programme B, quiz interaction was higher in E, discussion interaction and messaging had higher in Programme C. Thus, learners' behaviours may be explained partially with the interaction between age and distance programmes. In this context, evaluating the elements related to the course designs in the programme may provide more information. Therefore, RQ3 and RQ3.1 were answered through the second study group (the students taking three different online lessons together in a programme; n = 95).

How are learner interactions clustered in different online lessons?

In the second study group, the learner interactions were grouped on each lesson as follows; under three dimensions as live lesson, content and discussion for MD course (total variance = 75.58); under four dimensions as exam (quiz), assignment, content and live lesson for A course (total variance = 85.08), under six dimensions as content, assignment, award, exam (quiz), discussion and live lesson for ICT (total variance: 88.90). The learners' interactions were clustered separately in the selected lessons (Table 1, Study Group 2). According to hierarchical clustering analysis, learners in the MD course divided into three clusters with different profiles: user cluster with limited interaction (LI); content-oriented (CO) user cluster and live lesson–oriented (LO) user cluster. The learners in the A course divided into four clusters: LI, assessment-oriented (AsO) user cluster, LO and CO. The learners in the ICT course divided into four clusters: LI, AsO, award-oriented (AwO) user cluster and user with enriched interaction (EI) (Table 9).

Table 9
Profiles of clusters (n = 95) in different lessons

Courses	LI	LO	CO	AsO	AwO	EI	Visualisation
MD	CL1 (*n* = 32; 33.7%)	CL2 (*n* = 52; 54.7%)	CL3 (*n* = 11; 11.6%)	-	-	-	
A	CL1 (*n* = 35; 36.8%)	CL4 (*n* = 20; 21.1%)	CL3 (*n* = 23; 24.2%)	CL2 (*n* = 17; 17.9%)	-	-	
ICT	CL1 (*n* = 20; 21.1%)	-	-	CL2 (*n* = 36; 37.9%)	CL3 (*n* = 31; 32.6%)	CL4 (n = 8; 8.4%)	

Note. CL = cluster in various lessons.

In the MD course, the results of the Kruskal-Wallis test demonstrated that the interaction of the clusters with live lessons and content significantly differed ($\chi2_{\text{live lesson}}$ (2,95) = 59.03, *p* = .000; $\chi2_{\text{content}}$(2,95) = 29.72, *p* = .000). The interactions of those in the cluster CL3 ($\bar{x}_{CL3}$ = 2.55, $p < .01$) with content were significantly higher than that of the other two clusters ($\bar{x}_{CL1}$ = -0.35; $\bar{x}_{CL2}$ = -0.33). The interactions of the clusters CL1 and CL2 with content were similar. The interactions of the cluster CL2 with live lesson were significantly higher than that of the clusters CL3 and CL1 ($\bar{x}_{CL1}$ = -0.85; $\bar{x}_{CL2}$ = 0.56; $\bar{x}_{CL3}$ = -0.17). In the A course, the results of the Kruskal-Wallis test demonstrated that the interaction of the clusters with all components significantly differed ($\chi2_{\text{exam}}$(3,95) = 15.77, *p* = .000; $\chi2_{\text{assignment}}$(3,95) = 31.6, *p* = .000; $\chi2_{\text{content}}$(3,95) = 44.14, *p* = .000; $\chi2_{\text{live lesson}}$(3,95) = 49.49, *p* = .000). All clusters except the cluster CL1 had the significantly highest score with at least one component of interaction. The assignment interactions of CL2 were significantly higher than the other clusters, and exam interactions of CL3 were significantly higher than other clusters. The content interactions of CL4 were significantly higher than other clusters; but the assignment interactions of CL4 were higher than CL1 and CL3; and the interactions of CL4 were higher than the clusters CL1 and CL2. In ICT, the results of the test demonstrated that the interaction of the clusters with all components exhibited significant differences ($\chi2_{\text{exam}}$(3,95) = 30.059, *p* = .000; $\chi2_{\text{assignment}}$(3,95) = 55.95, *p* = .000; $\chi2_{\text{content}}$(3,95) = 34.17, *p* = .000; $\chi2_{\text{live lesson}}$(3,95) = 24.52, *p* = .000; $\chi2_{\text{discussion}}$(3,95) = 8.36, *p* = .000; $\chi2_{\text{award}}$(3,95) = 55.96, *p* = .000). CL1's interactions are significantly lower than other clusters. CL2's assignment interactions were significantly higher than the other clusters. CL3's reward interactions were significantly higher than the other clusters. CL4's interactions with exam, live lesson, discussion and content were higher than the other clusters.

Is there a significant difference in the academic achievement of learners clustered in different online lessons?

Regarding this sub-problem, we found that the assumption of homogeneity of variances for the final exam grades could be applied in the MD course (Levene = 1.601, df1 = 2, df2 = 92, p = .207), the A course (Levene = 1.666, df1 = 3, df2 = 91, p = .180), and ICT course (Levene = .574, df1 = 3, df2 = 91, p = .633), and we performed the ANOVA test based on the interaction clusters and final exam scores.

The results of the ANOVA test showed that; in the MD course, the academic achievement of the CO and LO users were significantly higher than LI users (X_{CO} = 58.73, X_{LO} = 61.25, X_{LI} = 47.72, F = 19.950, p = .000). However, there was no significant difference between the CO and LO users. In the A course, the academic achievements of CO and LO users were significantly higher than LI and AsO users (X_{CO} = 58.09, X_{LO} = 59.20, X_{AsO} = 37.90, X_{LI} = 26.52, F = 28.669, p = .000). However, there was no significant difference between CO and LO users. In ICT, the academic achievement of AwO and EI users was significantly higher than LI and AsO users (X_{EI} = 58.38, X_{AwO} = 56.86, X_{AsO} = 46.9, X_{LI} = 39, F = 9.216, p = .000). However, there was no significant difference between AwO and EI users.

While analysing learner behaviours across the whole system according to age and programme and analysing the change in academic achievement according to interaction clusters in various courses, we have seen that the components used in course design may impact both behaviour profiles academic achievement. Therefore, analysing ICT course according to interaction clusters, which include more components (e.g., award, assignment, exam (quiz)) in course design, may contribute to observing the effect of components in course design on behavioural profiles in more detail. In this context, RQ3.2 and RQ3.3 were answered through the data of the third study group (the students taking the ICT course (n = 112).

How do the interactions of learners vary by week?

As for this sub-problem, the actions identified by PCA performed in the ICT course were classified by participation modes, similar to the other studies in the relevant literature. Following the identification, calculations were made to determine how frequently the learners performed the actions classified by weeks according to participation modes. A graphic was created to represent the weekly participation mode in terms of the interaction clusters that the learners were in (Figure 5).

Figure 5. Weekly participation mode in terms of the interaction clusters (C: content interaction, T: teacher interaction)

Figure 5 shows each cluster's content and live lesson interactions and participation mode (the number of active or passive participation in the assessment, exam (quiz), discussion and award) by week. LI users according to content, teacher and participation mode were lower than other clusters for all weeks. EI users were higher than other clusters for all weeks. Although interactions of AsO and AwO users were similar between the eighth and 10th weeks, interactions of AwO users were higher than AsO for most of the weeks (Figure 6).

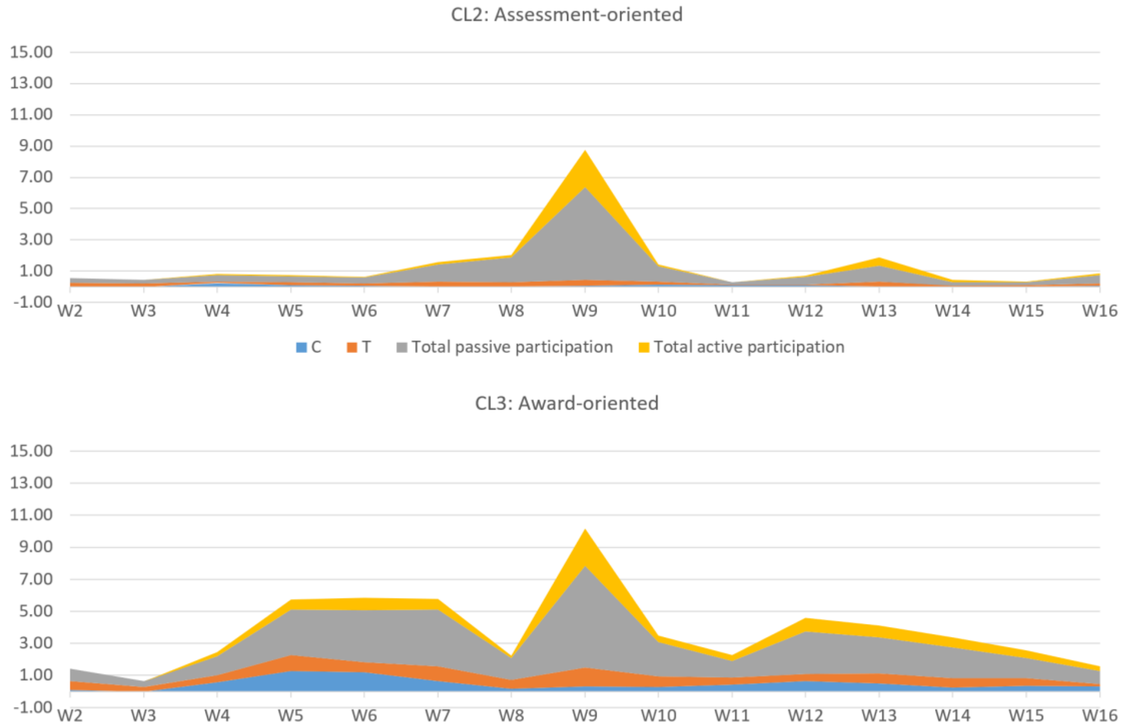


Figure 6. Comparison of AsO and AwO

What navigation behaviours do the learners demonstrate?

SPA demonstrated that LI users had no sequential pattern exhibited by at least half of the learners. More than half of AsO users (support = 0.63, 0.70, 0.85) had similar navigational behaviours (Table 10). For example, 85% of the learners in the cluster (support = 0.85) became aware of the presence of an award in the ninth week; then, they demonstrated active participation in assignments and passive participation in the award. The interactions of this cluster increased only in the ninth week when the learners were given an assignment, but their interactions in the ninth week were not sufficient to receive the award.

Table 10
Sequential patterns of the AsO users

Sequential patterns	Support	Length
{9|AwP}{9|AsP}{9|AsA}{9|AwP}	0.63	4
{9|AwP}{9|AsP}{9|AwP}	0.70	3
{9|AsP}{9|AsA}{9|AwP}	0.85	3

Note. 9: ninth week, AwP: award-passive, AsP: assignment-passive, AwA: award-active, AsA: assignment-active.

More than half of the AwO users (support = 0.53, 0.58, 0.50) had similar navigational behaviours (Table 11). However, the support ratio of the sequential navigation patterns of this cluster were lower than that of the AsO users. Of the AwO, 53% interacted with the teacher in the fifth week, then showed active participation in the assignment in the ninth week and received the award in the same week. Also, it is remarkable that the interactions of 58% of the cluster after interacting with the content in the sixth week, and 50% of the cluster after interacting with the teacher in the eighth week, were similar to the ninth week.

Table 11
Sequential patterns of the AwO users

Sequential patterns	Support	Length
{5|T}{9|AsP}{9|AsA}{9|AwA}	0.53	4
{6|C}{9|AsP}{9|AsA}{9|AwA}	0.58	4
{8|T}{9|AsP}{9|AsA}{9|AwA}	0.50	5

Note. 5, 6, 8 and 9: week, T: teacher interaction, C: content interaction, AsP: assignment-passive, AsA: assignment-active, AwA: award-active.

It should be noted that the system interactions of the AwO users were their active participation in the sixth, ninth, or 12th week (see Figure 6). However, it is striking that the navigational behaviours of this cluster were limited to the ninth week. The reason may be that the minimum support value for SPA was set to 0.5. This caused the analysis to yield the navigational patterns demonstrated by only half or more than half of the learners in the cluster. It can be argued that the learners in the AwO users may have navigational patterns that had a lower support ratio but differed within the group.

Three-quarters of the EI users (support = 0.75) exhibited similar navigational behaviours. As seen in Table 12, the learners in this cluster began their interaction with the teacher or content in the fifth week, then interacted with content and received the award in the same week and showed passive participation and then active participation in the exam. They interacted with the teacher in the seventh week; in the ninth week (midterm exam), they showed passive and active participation in the exam and interacted with the teacher again. In the 10th week, they showed passive participation in the award. As expected, the sequential navigation patterns of the EI users immensely varied. Furthermore, it was found that the learners in this cluster showed active participation in awards given their navigations in the fifth week.

Table 12
Sequential patterns of the EI users

Sequential patterns	Support
{5|C}{5|C}{5|AwA}{5|EP}{5|EA}{7|T}{9|AsP}{9|AsA}{9|AwP}{9|T}{10|AwP}	0.75
{5|C}{5|C}{5|AwA}{5|EP}{5|EA}{7|AwP}{9|AsP}{9|AsA}{9|AwP}{9|T}{10|AwP}	0.75
{5|T}{5|C}{5|EP}{5|AwA}{5|EP}{5|EA}{7|T}{9|AsP}{9|AsA}{9|T}{10|AwP}	0.75

Note. 5, 7, 9 and 10: week, T: teacher, C: content, EP: exam-passive, EA: exam-active, AsP: assignment-passive, AsA: assignment-active, AwP: award-passive, AwA: award-active.

In short, the findings obtained from the analysis of the sequential navigation patterns by participation mode revealed the features of the interaction clusters (LI, AsO, AwO, EI) in more detail. For example, AsO users only navigated in the weeks when the midterm exam was held as a project, AwO showed active participation (they received an award) in award unlike AsO users and EI users had a more varied navigation pattern than others. However, it can be claimed that the AwO may have more specific navigation patterns that differed within the group.

Results and discussion

This study first examined the system interactions of the learners in an online learning environment across the whole system and in various lessons (MD, A, ICT). It ascertained three interaction clusters for the analysis of the frequency of usage across the whole system: LI, AsO and LO. The interactions of the LI with the components in the system (live lesson, assignment, exam, discussion, content, message) are significantly lower than the other groups. In the AsO cluster, interactions with assignments, exams and discussion components are also significantly higher than in other groups. In the LO cluster, interactions only with the live lesson component are significantly higher than in other clusters. In this study, learners' profiles across the whole system shape according to the usage of the components. However, is this result from a preference for learners? Or are there other factors that force the learner to do this? At this point, this study discussed findings depending on the individual characteristics of the learner (e.g., age), features of the programmes (e.g., content, the practice of lectures) and the interactions of these dimensions with each other.

When considering the features of the programmes, usage of activities varies significantly according to the distance education programmes. For example, while AsO interaction is higher in some programmes (e.g., Banking and Insurance), LO interaction seems higher in some programmes (e.g., Divinity Diploma Upgrade). This result is discussed first in the context of the reasons arising from the interaction between the programme features and the system. For example, most of the courses (75%–81.25%) in Programme B (Banking and Insurance) and Programme F (Tourism and Hotel Management), midterm exams are conducted according to the assignment activity. Therefore, it may be expected that the interaction of these two programmes (B & F) with assessment is higher than other programmes. We also determined that the content and discussion interactions in Programmes B and F with high assignment interactions were high. In this context, evaluating the midterm exams according to the assignment may have positively affected the interaction with other components in the system.

This study confirmed that age is an element as one of the reasons for the interaction difference with system components. However, age is also not sufficient to explain the differences in some situations. For example, the learners of advanced age (e.g., Programme D) showed LO interaction, while younger learners (e.g., Programmes B & F) showed AsO interaction. However, in Programme B, C and E, which are similar age groups, B was more prominent in assignment interactions, E was more pronounced in exam interactions and C was more prominent in discussion and messaging interactions. This result is consistent with Bravo-Agapito et al.'s (2021) findings. In their study, there was a difference between the characteristics of the two groups having a lower age (e.g., groups 1 and 2). While group 1 (e.g., few interactions, low grades) was mainly composed of students of Computer Science, Journalism and Psychology, group 2 (e.g., high interaction with assignments and questionnaires, high grades) were composed mainly of Computer Science, Criminology and Psychology students. These results may be discussed in two ways.

Behavioural patterns of learners may be a result of the interaction of age and assessment. When predicteing the final exam and the outcome of five assessments that offered at different times, Rizvi et al. (2019) visualised that the age is about 6%–7% important for the final performance and about 20% important for some assessments (2, 3, 4). In our study, midterm exams are primarily evaluated as assignments in Programmes B and C, as electronic exams in Programme E. Therefore, in the current study, the assignments given in the programmes (similar in terms of age groups) may have directed the students differently. For example, in Programme C, assignments may have directed students to discuss.

On the other hand, the contextual characteristics of each distance education programme may reveal its own distance education design. Programme D (Divinity Diploma Upgrade) learners work as religious officials. It may be said that religious officials are prone to verbal communication and the course contents of the theology programme are text-based. Therefore, the lectures in this programme may be carried out mainly with live lessons. There may also be differences in the way courses are conducted based on disciplines. For example, Kálmán et al. (2020) revealed that university teachers apply different teaching approaches (e.g., knowledge-focused approach or practice-focused approach) according to discipline (e.g., soft or hard).

Analysis of learners' interactions in various online lessons (MD, A, ICT) yielded similar clusters but also identified different ones for some of the lessons. Remarkably, in the A course, AsO users were in either LI cluster, CO cluster, or LO cluster in the MD course. On the other hand, when the reward is activated in the design of the ICT course, the behaviour profile of a group of students is AwO, and they behave in a different profile than the behaviour profiles in other courses. In previous studies (Bravo-Agapito et al., 2021; Cerezo et al. 2016, Soffer et al., 2019), researchers do not emphasise enough that besides determining learners' behavioural patterns, behavioural profiles may originate from the components in course design. For example, Machajewski et al. (2019, p. 1) identified "three latent classes of courses were characterised as holistic tool use (28%), complementary tool use (51%), and content repository (21%)." Therefore, these descriptions show there are different needs in various courses for LMS tools. The current study showed that interaction clusters might depend on the components present in the design of lessons. Therefore, this study pointed out that learners may participate in some activities more often depending on their needs when provided with different activities during a lesson. This comment seems to be supported by researching academic achievement by interaction clusters.

After clustering in various courses (MD, A and ICT), how academic achievement changes according to clusters was examined. In the MD course, CO and LO users were more successful than LI users. In the A course, CO and LO users were more successful than LI and AsO users. In ICT, AwO and EI users were more successful than LI and AsO users. However, it was found that there was no significant difference between the clusters (e.g., CO and LO, AwO and EI) that were more successful than LI and AsO in each course. Therefore, it can be said that learners who interact more with any component (e.g., live lesson or content) according to their needs are more successful than LI learners. However, Çebi and Güyer (2020) found that the cluster that uses more support materials in the design of the course does not have a significant difference in terms of academic success compared to those who use the materials less and that the more students who work on the primary course materials (video, example, forum intensive use students) are more successful than other. In the current study, users who have sufficient interaction with any of the components related to content transfer (e.g., content, live lesson) seem to be more successful.

One of the findings is that AsO was not significantly successful relative to LI. The reason may be that AsO interacts intensely with the assessment activity (e.g., assignment or homework) only during the assessment weeks. In other words, LI users during the term may have been in intensive interaction with the assignment in the assessment week. Indeed, other findings of the research support this interpretation. For example, when examining the learners' interactions clustered by weeks, this study determined that AsO was higher than LI between the eighth and 10th weeks (midterm weeks), and no difference was observed in the other weeks. SPA confirmed that 85% of AsO users demonstrated active participation in assignments but did not receive awards. The interactions of this cluster increased only in the ninth week when the learners were given an assignment, but their interactions in the ninth week were not sufficient to receive the award. So, the final performance of the AsO users may have been poor, as they had low levels of system interaction throughout the term.

Furthermore, Saa et al. (2019) reported that in 25% of the studies examined by a systematic review, the results of the learner's e-learning activities affect performance. Therefore, in our study, the academic performance of the clusters with higher interaction with the activities may be expected higher. For example, the more the students engage in e-learning activities (e.g., accessing online material, solving online quizzes, and uploading assignments into the e-learning system), the more likely the students achieve higher grades and improve their overall performance. However, in our study, although the EI users had high content and teacher interaction levels, there was no significant difference in terms of academic achievement between AwO and EI clusters. This result may be interpreted by the equivalency of interaction (Anderson, 2003). Anderson stated that learning activities should be built by evaluating strategic amounts of each type of interaction (student-teacher, student-student, student-content). Therefore, the reward mechanism in the ICT course may have enabled the AwO users who do not have as rich of interaction as EI to fulfill the minimum amount of interaction in terms of academic achievement.

For the last part of the study, to obtain more detailed information about learner behaviours, navigation behaviours were analysed based on participation mode according to interaction clusters through SPA. While participation mode did not lead to a difference among interaction clusters in the assessment activities, it contributed to a better understanding of the behaviours exhibited by interaction clusters. Moreover, it was ascertained that participation mode was a decisive factor for clusters about the award. For instance, AsO's navigation patterns concentrated on the midterm week as expected. During midterm, AsO participated in the assignment passively and actively (by participation mode). In other words, AsO did not take any action other than completing their homework and uploading it to the system (e.g., pattern: {9|AsP}-{9|AsA}-{9|AwP}; support: 0.85). AwO interacted with the content or the teacher in the fifth, sixth and eighth weeks, completed their homework in the ninth week and received the award (e.g., pattern: {6|C}-{9|AsO}-{9|AsA}-{9|AwA}; support: 0.58).

It was remarkable that in SPA, navigation behaviours may be shaped by the components present in the design of the lessons. For example, the AsO users navigated when given an assignment; the AwO users navigated in the weeks when they were given assignments and awards, whereas the EI users navigated during the whole term and in all activities. Thus, it may be inferred that including activities that help learners passing the lesson and receiving an award in the design of a lesson may encourage learner interactions. Therefore, online lessons may be designed in such a way that learners are assigned more tasks every week and benefit according to their interactions. This study reported that system facilities were not utilised by default in all lessons. Therefore, the assignments or award mechanism may not need to be used for some courses. On the other hand, some users with greatly limited interaction failed to demonstrate any specific navigation pattern regardless of system facilities. Therefore, improvements to the design of a lesson alone are not sufficient.

Conclusions and suggestions

This study has shown interrelation of interaction, sequential patterns and academic achievement across the whole system and in various courses to evaluate a distance education system in terms of age, programme and course design as multidimensional. Consequently, it is difficult to determine or standardise intervention unless the system, programme and course design features are standard. While examining the system interactions of learners, it should not be ignored that the behaviour patterns of learners are the result of the interaction of learner characteristics (e.g., age), features programme and course design (e.g., system components used). In the ICT course, assessment and award activities seem to have positive contributions to both academic performance and interaction with other system components. However, the effects of learning design activities on performance could not be revealed more clearly (Holmes et al., 2019). Alternatively, "it is difficult to deploy a predictive model which is not customised for the target learning environment" (Hung et al., 2019, p.152). Therefore, performance is the result of the interaction of each dimension with each other in distance education. If we are looking for appropriate intervention, the focus should not be data only when making data-based decisions in a distance education system. The contextual characteristics of this data should also be taken into account.

Limitations and future research

In this study, there are limitations in several aspects in the discovery of components related to course design. First, the efficient working time of the components was also neglected. The results may change related to the content component. In future research, variables related to the quality of interaction allow for a deeper understanding of the learner. Second, the analysis used in this research cannot verify a cause-and-effect relationship between academic achievement and activities. In this context, empirical studies showing the effect of course design on academic achievement may be needed.

While examining learners' interactions in various lessons, log data of small learner groups (n = 95 or n = 112) were studied. Although the sample selection process (Table 1) shows the current problem related to course design in data-driven decision-making, it contains limitations in terms of generalisability. Therefore, more generalisable findings can be revealed by working on large lesson samples using similar course designs.

Moreover, the course design is a multidimensional structure. It was not possible to examine the same student group in the same course with two different course designs as the current study reveals the existing situation in the LMS. So, RQ2 does not compare course designs against an experimental design. It reflects the interactions of the same students in different courses, with different designs in terms of content, assessment, discussion, rewarding. Therefore, while discussing students' interactions with course design, we primarily took care not to make a causal interpretation. Future research should focus on more substantial implications for how course designs affect learner interactions.

While examining the navigation behaviours, we categorised learners only by interaction clusters. In future research, after learners are categorised to other variables (e.g., age, discipline, performance) and interaction patterns, SPA could be applied according to the newly created categories. Alternatively, the navigation patterns can be clustered without categorising learners according to some variables.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*, 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *The International Review of Research in Open and Distributed Learning*, *4*(2), 1–14. https://doi.org/10.19173/irrodl.v4i2.149

Botelho, A., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S., & Beck, J. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, *12*(2), 158–170. https://doi.org/10.1109/TLT.2019.2912162

Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate students' academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, *115*, Article 106595. https://doi.org/10.1016/j.chb.2020.106595

Campagni, R., Merlini, D., & Sprugnoli, R. (2012, September 28). *Sequential patterns analysis in a student database* [Workshop presentation]. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Bristol, United Kingdom.

Cantabella, M., López, B., Caballero, A., & Muñoz, A. (2018). Analysis and evaluation of lecturers' activity in learning management systems: Subjective and objective perceptions. *Interactive Learning Environments*, *26*(7), 911–923. https://doi.org/10.1080/10494820.2017.1421561

Çebi, A., & Güyer, T. (2020). Students' interaction patterns in different online learning activities and their relationship with motivation, self-regulated learning strategy and learning performance. *Education and Information Technologies*, *25*, 3975–3993. https://doi.org/10.1007/s10639-020-10151-1

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, *96*, 42–54. https://doi.org/10.1016/j.compedu.2016.02.006

Choudhury, S., & Pattnaik, S. (2020). Emerging themes in e-learning: A review from the stakeholders' perspective. *Computers & Education*, *144*, Article 103657. https://doi.org/10.1016/j.compedu.2019.103657

Cohen, A., Shimony, U., Nachmias, R., & Soffer, T. (2019). Active learners' characterization in MOOC forums and their generated knowledge. *British Journal of Educational Technology*, *50*(1), 177–198. https://doi.org/10.1111/bjet.12670

Council of Higher Education. (2020). Yükseköğretim Kurumlarında Uzaktan Öğretime İlişkin Usul Ve Esaslar [Procedures and principles regarding distance education in higher education institutions]. https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Uzaktan_ogretim/yuksekogretim_kurumlarinda_uzaktan_ogretime_iliskin_usul_ve_esaslar.pdf

Dráždilová P., Obadi G., Slaninová K., Al-Dubaee S., Martinovič J., Snášel V. (2010) Computational intelligence methods for data analysis and mining of elearning activities. In F. Xhafa, S. Caballé, A. Abraham, T. Daradoumis, A. A. Juan-Perez (Eds.), *Studies in Computational Intelligence: vol. 273. Computational intelligence for technology enhanced learning* (pp. 195–224). Springer. https://doi.org/10.1007/978-3-642-11224-9_9

Gavilanes-Sagnay, F., Loza-Aguirre, E., Riofrío-Luzcando, D., & Segura-Morales, M. (2018). A systematic literature review of indicators for the understanding of interactions in virtual learning environments. In *Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence* (pp. 596–600). IEEE. https://doi.org/10.1109/CSCI46756.2018.00120

Gkontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2019). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 1–16. https://doi.org/10.1080/10494820.2019.1709209

Han, F., & Ellis, R. (2020). Combining self-reported and observational measures to assess university student academic performance in blended course designs. *Australasian Journal of Educational Technology*, *36*(6), 1–14. https://doi.org/10.14742/ajet.6369

Holmes, W., Nguyen, Q., Zhang, J., Mavrikis, M., & Rienties, B. (2019). Learning analytics for learning design in online distance learning. *Distance Education*, *40*(3), 309–329. https://doi.org/10.1080/01587919.2019.1637716

Hung, J.L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving predictive modeling for at-risk student identification: A multistage approach. *IEEE Transactions on Learning Technologies*, *12*(2), 148–157. https://doi.org/10.1109/TLT.2019.2911072

Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Kahan, T., Soffer, T., & Nachmias, R. (2017). Types of participant behavior in a massive open online course. *The International Review of Research in Open and Distributed Learning*, *18*(6), 1–18. https://doi.org/10.19173/irrodl.v18i6.3087

Kálmán, O., Tynjälä, P., & Skaniakos, T. (2020). Patterns of university teachers' approaches to teaching, professional development and perceived departmental cultures. *Teaching in Higher Education*, *25*(5), 595–614. https://doi.org/10.1080/13562517.2019.1586667

Lin, F. C., Chen, C. M., & Wang, W. F. (2017). Learning process analysis based on sequential pattern mining and lag sequential analysis in a web-based inquiry science environment. In T. Matsuo, N. Fukuta, M. Mori, K. Hashimoto, & S. Hirokawa (Eds.), *Proceedings of the IIAI International Congress on Advanced Applied Informatics* (pp. 655–660). IEEE. https://doi.org/10.1109/IIAI-AAI.2017.57

Machajewski, S., Steffen, A., Fuerte, E. R., & Rivera, E. (2019). Patterns in faculty learning management system use. *TechTrends*, *63*(5), 543–549. https://doi.org/10.1007/s11528-018-0327-0

Munk, M., & Drlik, M. (2011). Influence of different session timeouts thresholds on results of sequence rule analysis in educational data mining. In H. Cherifi, J. M. Zain, & E. El-Qawasmeh (Eds.), *Communications in computer and information science: vol. 166*. *Proceedings of the Digital Information and Communication Technology and Its Applications Conference* (pp. 60–74). Springer. https://doi.org/10.1007/978-3-642-21984-9_6

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Nguyen, Q., Huptych, M., & Rienties, B. (2018). Using temporal analytics to detect inconsistencies between learning design and students' behaviours. *Journal of Learning Analytics*, *5*(3), 120–135. https://doi.org/10.18608/jla.2018.53.8

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, F., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*, *76*, 703–714. https://doi.org/10.1016/j.chb.2017.03.028

Rienties, B., & Jones, A. (2019). Evidence-based learning: Futures. Using learning design and learning analytics to empower teachers to meet students' diverse needs. In R. Ferguson, A. Jones, & E. Scanlon (Eds.), *Educational visions: The lessons from 40 years of innovation* (pp. 109–125). Ubiquity Press. https://doi.org/10.5334/bcg.g

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning: A decision tree based approach. *Computers & Education*, *137*, 32–47. https://doi.org/10.1016/j.compedu.2019.04.001

Saa, A. A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, *24*(4), 567–598. https://doi.org/10.1007/s10758-019-09408-7

Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., & Montenegro, M. (2018). Centralized student performance prediction in large courses based on low-cost variables in an institutional context. *The Internet and Higher Education*, *37*, 76–89. https://doi.org/10.1016/j.iheduc.2018.02.002

Shih, W. C. (2018, July). Mining sequential patterns to explore users. In S. Reisman, S. I. Ahamed, C. Demartini, T. Conte, L. Liu, W. Claycomb, M. Nakamura, E. Tovar, S. Cimato, C. Lung, H. Takakura, J. Yang, T. Akiyama, Z. Zhang, & K. Hasan (Eds.), *Proceedings of the IEEE 42nd Annual Computer Software and Applications Conference* (pp. 126–129). IEEE. https://doi.org/10.1109/COMPSAC.2018.10216

Soffer, T., Kahan, T., & Livne, E. (2017). E-assessment of online academic courses via students' activities and perceptions. *Studies in Educational Evaluation*, *54*, 83–93. https://doi.org/10.1016/j.stueduc.2016.10.001

Soffer, T., Kahan, T., & Nachmias, R. (2019). Patterns of students' utilization of flexibility in online academic courses and their relation to course achievement. *The International Review of Research in Open and Distributed Learning*, *20*(3), 202–220. https://doi.org/10.19173/irrodl.v20i4.3949

Sun, J. C. Y., Lin, C. T., & Chou, C. (2018). Applying learning analytics to explore the effects of motivation on online students' reading behavioral patterns. *The International Review of Research in Open and Distributed Learning*, *19*(2), 209–227. https://doi.org/10.19173/irrodl.v19i2.2853

Van den Beemt, A., Buijs, J., & van der Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. *The International Review of Research in Open and Distributed Learning*, *19*(5), 37–60. https://doi.org/10.19173/irrodl.v19i5.3748

Wang, Y. L., Wen, L. Y. M., Chen, T. S., & Chen, R. C. (2012). Using sequential pattern mining to analyze the behavior on the WELS. In X. Qu & Y. Yang (Eds.), *Communications in computer and information—Proceedings of the Information and Business Intelligence Conference* (part 1, pp. 95–101). Springer. https://doi.org/10.1007/978-3-642-29084-8_15

Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education*, *140*, Article 103595. https://doi.org/10.1016/j.compedu.2019.103595

Zhang, J., Burgos, D., & Dawson, S. (2019). Advancing open, flexible and distance learning through learning analytics. *Distance Education*, *40*(3), 303–308. https://doi.org/10.1080/01587919.2019.1656151

Zhong, S.H., Li, Y., Liu, Y., & Wang, Z. (2017). A computational investigation of learning behaviors in MOOCs. *Computer Applications in Engineering Education*, *25*(5), 693–705. https://doi.org/10.1002/cae.21830

Zhou, M. (2010). Data mining and student e-learning profiles. In *Proceedings of the International Conference on E-Business and E-Government* (pp. 5405–5408). IEEE. https://doi.org/10.1109/ICEE.2010.1352

Corresponding author: Denizer Yildirim, dyildirim@ankara.edu.tr

Please cite as: Yildirim, D., & Usluel, Y., (2022). Interrelated analysis of interaction, sequential patterns and academic achievement in online learning. *Australasian Journal of Educational Technology*, *38*(2), 181-200. https://doi.org/10.14742/ajet.7360