A Multiple Study Investigation of the Evaluation Framework for Learning Analytics: Instrument Validation and the Impact on Learner Performance

Mehmet Kokoç[1*] and Mehmet Kara[2]
[1]School of Applied Sciences, Trabzon University, Turkey // [2]Vocational School of Design, Amasya University, Turkey // kokoc@trabzon.edu.tr // mehmet.kara@amasya.edu.tr
[*]Corresponding author

(Submitted August 19, 2020; Revised October 15, 2020; Accepted November 16, 2020)

ABSTRACT: The purposes of the two studies reported in this research are to adapt and validate the instrument of the Evaluation Framework for Learning Analytics (EFLA) for learners into the Turkish context, and to examine how metacognitive and behavioral factors predict learner performance. Study 1 was conducted with 83 online learners enrolled in a 16-week course delivered through the Moodle learning management system. The findings from the confirmatory factor analysis indicated that a three-factor model of the EFLA for learners provided the best model fit for the collected data. The model is consistent with the factorial structure of the original instrument developed based on the data from the European learners. Study 2 aimed to reveal how the metacognitive and behavioral factors pertaining to the learning analytics dashboard predict learners' academic performance. A total of 63 online learners enrolled in a 14-week online computing course participated in this study. The results from the logistic regression analysis indicated that online learners more frequently interacted with the learning analytics dashboard demonstrated greater academic performance. Furthermore, the dimensions of the EFLA, together with the interaction with the dashboard, significantly predicted learners' academic performance. This multiple-study investigation contributes to the generalizability of the EFLA for learners and highlights the importance of metacognitive and behavioral factors for the impact of learning analytics dashboards on learner performance.

Keywords: Learning analytics, Learning analytics dashboards, Evaluation, Validation, Learning performance

1. Introduction

Data use in various institutions has been adopted as a way of decision making as a result of the vast amount of data produced through online systems. Learning Analytics (LA) is one of these methods used to make decisions on learning design. LA was defined by Long and Siemens (2011) as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (p. 34). It has emerged as an evolving field of research and practice and has a great potential to transform learning and teaching activities based on data-driven practices (Howell et al., 2018; Ifenthaler & Yau, 2020; Wong & Li, 2020). This potential requires educational researchers and practitioners to pay more attention to LA research and practice. The review studies on LA, however, indicate that the studies generally focus on small-scale implementations (Wong & Li, 2020) and had inadequate empirical evidence on their effectiveness (Larrabee Sønderlund, Hughes, & Smith, 2019; Viberg, Hatakka, Bälter, & Mavroudi, 2018). Thus, the projected potential has not been satisfactorily transformed into practice yet (Viberg et al., 2018) and further research is needed in this regard.

Research shows that the challenges are generally faced during the implementation of the LA interventions such as evaluation of effectiveness (Wong & Li, 2020). For this reason, it is highly desired to continuously evaluate and improve LA interventions and tools (Larrabee Sønderlund et al., 2019; Scheffel, Drachsler, Stoyanov, & Specht, 2014). One of the typically used LA interventions is LA dashboards (Jivet, Scheffel, Drachsler, & Specht, 2017; Jivet, Scheffel, Specht, & Drachsler, 2018). In spite of their common use, little attention was paid to their effectiveness or their impact on learner performance (Verbert et al., 2013; Verbert et al., 2014). It was also underlined that there is a need to further evaluate their effectiveness as pedagogical tools (Jivet et al., 2018). Matcha, Gašević, and Pardo (2020) also point out the restrictions in the way they are evaluated and reported. The current literature, therefore, urges the evaluation of LA dashboards as the pedagogical interventions through the robust evaluation instruments.

In this sense, Viberg et al. (2018) suggest spending more effort on the validation of LA tools and methods. Additionally, the inclusion of the stakeholders such as learners and teachers in the evaluation process is also underlined by the relatively recent studies (Howell et al., 2018; Samuelsen, Chen, & Wasson, 2019; Schumacher & Ifenthaler, 2018). Based on the need for a standardized evaluation framework and instrument, Scheffel et al.

(2014) developed quality indicators for LA and this framework was iteratively implemented, evaluated, and improved (Scheffel, Drachsler, & Specht, 2015; Scheffel, 2017; Scheffel et al., 2017). The final version of the Evaluation Framework for Learning Analytics (EFLA 4), included a solid measurement instrument that enables to gather data from both teachers and learners on the effectiveness of a specific LA tool (Scheffel, 2017). Consequently, the main goal of this multiple-study investigation is firstly to adapt the instrument developed and validated based on this framework and then, to use it to evaluate the impact of an LA dashboard on learner performance. A Prescriptive Learning Dashboard (PLD) was used and its impact on learner performance was evaluated in terms of both metacognitive and behavioral levels. In the present study, metacognitive levels covered the dimensions of the validated EFLA 4 for learners, including "data," "awareness & reflection," and "impact," while behavioral level included dashboard use and the time spent on the dashboard. The specific research questions were as follows:

- How valid and reliable is EFLA 4 for learners in the context of Turkey? (Study 1)
- How do metacognitive dimensions of the EFLA4 instrument for learners and interaction with LA dashboards predict learner performance in an online undergraduate course? (Study 2)

1.1. Evaluation framework for learning analytics

Based on the lack of evaluation standards for LA tools, several versions of the EFLA were created and validated. The EFLA studies aimed to reveal quality indicators for LA tools, to develop and use an instrument based on these indicators, and to validate an instrument to evaluate LA tools (Scheffel, 2017). In this regard, four versions of the EFLA were proposed as EFLA1, EFLA2, EFLA3, and EFLA4, each of which was based on the previous version.

EFLA 1. The first framework was proposed by Scheffel et al. (2014) in the form of quality indicators for LA. The quality indicators were identified as the result of a two-phase study. The participants of the first phase were the stakeholders involved in the implementation of the LA tools while the participants in the second phase were the subject-field experts of the LA. A total of 103 quality indicators were determined and used to create the framework. The first framework covered five areas of quality indicators as the dimensions (Scheffel et al., 2014): (1) "*Objectives*," (2) "*Learning support*," (3) "*Learning measures and output*," (4) "*Data aspects*," and (5) "*Organizational aspects*" (p. 126).

EFLA 2. In their study, Scheffel et al. (2015) evaluated EFLA 1 and used it as a base to construct an instrument to evaluate LA tools. The study was conducted with the participation of the Learning Analytics Community Exchange (LACE) project (http://www.laceproject.eu) and its partners. The study resulted in EFLA 2, with four dimensions for both teachers and learners including three items in each dimension. The identified dimensions for both learners and teachers were labeled as "data," "awareness," "reflection" and "impact" (Scheffel, 2017, p.54). In this way, the first instrument was developed based on EFLA 1 with diminished dimensions and items. According to Scheffel (2017), EFLA 2 met several requirements in the first version, EFLA 1. These were decreasing the number of dimensions and items, facilitating understanding of the dimensions and items, revising the instrument for the purpose of getting answers from both learners and teachers, and grounding the dimensions and items on a theoretical base.

EFLA 3. The previous version was used and evaluated in another second-phase study with the participation of learners and teachers to test its validity and reliability and to identify the problematic aspects of the previous instrument through principal component analysis and focus group interviews with the experts (Scheffel, 2017). The prior instrument was revised and converted into EFLA 3 based on the quantitative and qualitative findings. Thus, EFLA 3 still consisted of four dimensions for both learners and teachers. However, the statements of the items were revised and the number of the items in the dimensions of awareness and reflection was decreased to two. The statements started with "For this LA tool…" or "This LA tool…" instead of "I …".

EFLA 4. Scheffel et al. (2017) used and evaluated EFLA 3 in their two-phase study conducted in a MOOC environment with the participation of both learners and teachers. As a result of the first phase, two items were removed and the instrument has three dimensions for both learners and teachers: "Data," "Awareness & Reflection," and "Impact" (Scheffel, 2017, p.136). The second phase of the validity and reliability analyses indicated that the final version of the instrument, EFLA 4, is an eight-item valid and reliable instrument to evaluate an LA tool from the perspectives of both learners and teachers. Both data and impact dimensions included two items while awareness and reflection dimension included four items (see Table 1). The obtained Cronbach alpha values for the reliability of the instrument were presented in Table 2 for each dimension.

The continuous implementation, evaluation, and revision of the EFLA frameworks resulted in a valid and reliable research instrument to evaluate LA tools (Scheffel, 2017). Thus, the validated instrument could be used in various research contexts to evaluate the effectiveness of LA tools.

Table 1. Dimensions and items of EFLA 4 for learners and teachers

	Item number	Learners	Teachers
Data	1	"For this LA tool it is clear what data is being collected"	"For this LA tool it is clear what data is being collected"
	2	"For this LA tool it is clear why the data is being collected"	"For this LA tool it is clear why the data is being collected"
Awareness & reflection	3	"This LA tool makes me aware of my current learning situation"	"This LA tool makes me aware of my students' current learning situation"
	4	"This LA tool makes me forecast my possible future learning situation given my (un)changed behaviour"	"This LA tool makes me forecast my students' possible future learning situation given their (un)changed behaviour"
	5	"This LA tool stimulates me to reflect on my past learning behaviour"	"This LA tool stimulates me to reflect on my past teaching Behaviour"
	6	"This LA tool stimulates me to adapt my learning behaviour if necessary"	"This LA tool stimulates me to adapt my teaching behaviour if necessary"
Impact	7	"This LA tool stimulates me to study more efficiently"	"This LA tool stimulates me to teach more efficiently"
	8	"This LA tool stimulates me to study more effectively"	"This LA tool stimulates me to teach more effectively"

Note. Retrieved from Scheffel (2017, p. 136).

Table 2. Reliability values for EFLA 4 for learners and teachers

Dimension	*N*	Cronbach alpha (Scheffel et al., 2017)	
		Learners	Teachers
Data	2	.745	.574
Awareness & reflection	4	.916	.870
Impact	2	.954	.881

1.2. Evaluation of learning analytics dashboards

LA dashboards are one of the most commonly used interventions to make decisions about learning design. Schwendimann et al. (2017) define an LA dashboard as follows: "A learning dashboard is a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations" (p. 37). Based on this definition, data visualization is a key aspect of LA dashboards to inform stakeholders about learning design. LA dashboards work with learner log data, extracting meaning from their data, and visualizing the obtained results (Park & Jo, 2015). Previous studies have indicated their positive effects on learner performance (e.g., Kim, Jo, & Park, 2016; Kokoç & Altun, 2019).

Several review studies on LA dashboards revealed their potential to impact learning design and outcomes (e.g., Bodily & Verbert, 2017; Jivet et al., 2018; Matcha et al., 2020; Schwendimann et al., 2017). One of the earliest review studies was conducted by Verbert et al. (2013). They reviewed 15 learning dashboard applications and proposed a process model for their evaluation. The model includes awareness, self-reflection, sense-making, and impact. By characterizing LA dashboards as the interventions enhancing awareness, reflection, and behavioral change, Verbert et al. (2014) categorized them into three groups: (1) the ones used in face-to-face courses, (2) used in face-to-face groups, and (3) used in online or blended courses. Both of these review studies concluded in common that little research was conducted to evaluate the impact of these dashboard interventions. In their review study, Bodily and Verbert (2017) further argued that there is a need for the dashboard intervention studies focusing not only on the impact, but also on the design and development process through needs analysis, analysis of visual design, and learner surveys. A recent review study by Matcha et al. (2020) similarly concluded that most of the interventions evaluated perceived usefulness while a few of them evaluated the impact.

Although acceptance studies are assumed as a requisite (Jivet et al., 2018), several of the studies underlined that the main research focus is required to be on their impact on learners and learning outcomes (Bodily, Ikahihifo, Mackley, & Graham, 2018; Jivet et al., 2018; Matcha et al., 2020; Park & Jo, 2015; Schwendimann et al., 2017). Jivet et al. (2018) advocated that the primary focus of the research on LA dashboards is required to be on learning goals and the dashboards should be evaluated as the pedagogical tools. Likewise, Matcha et al. (2020) criticized that the studies on LA dashboards are seldom based on learning theory and do not provide suggestions for effective learning. Schwendimann et al. (2017) similarly revealed that more than half of the reviewed studies did not specify any pedagogical approach and suggested evaluation of the interventions in a way that they clearly show the impact on learning. In this sense, Jivet et al. (2018) proposed the levels and criteria for the pedagogical evaluation of the LA dashboards. The levels encompassed metacognitive, cognitive, behavioral, emotional, self-regulation, and usability. Based on these levels and criteria, the current study focused on the evaluation of an LA dashboard at the metacognitive and behavioral levels. Metacognitive level included understanding, agreement, and impact on awareness and reflection while behavioral level included the impact on behavior and system usage (Jivet et al., 2018).

2. Study 1: Scale adaptation and validation

The purpose of Study 1 is twofold: (1) to adapt EFLA 4 for learners into Turkish language and culture and (2) to improve its generalizability through the evidence about its psychometric properties from a different context and participants. In this respect, the present study contributes to the relevant literature on LA by providing a valid and reliable instrument to evaluate the effectiveness of an LA tool. The external validity of EFLA 4 is also to be improved. Thus, this standardized instrument would be used in various contexts to implement, evaluate, and improve LA tools.

2.1. Method

2.1.1. Participants

A total of 83 undergraduate students voluntarily participated in the study. 41 of them (49.4%) were females and 42 of them (50.6%) were males. The participants were the pre-service teachers enrolled in the undergraduate program of "Computer Education and Instructional Technology." All participants stated that they previously registered for at least one fully online course.

2.1.2. Context and procedure

The study was conducted within an undergraduate course entitled "Measurement and Evaluation in Education" at a large-scale public university in Turkey. The content of the course covered the fundamentals of measurement and evaluation in education, validity and reliability, measurement instruments, methods, and item analyses. This 16-week course was delivered through Moodle v3.2 LMS in a semester. In this course, the PLD developed by Kokoç and Altun (2019) was used as an open source LA dashboard compatible with Moodle (see Figure 1). EFLA 4 for learners was distributed as a 10-point Likert scale, ranging from "strongly disagree" to "strongly agree."

The PLD visualizes learners' performance and class average performance in terms of eight LA indicators (basic usage, learning objects, and discussion activities) and assessment scores. Also, the PLD provides personalized real-time recommendations to help learners as text messages based on predictive learner success model.

The scale adaptation procedure was conducted by following the steps recommended by Hambleton (2005). The robustness of the dataset was investigated in the first step through data screening. Missing data, outliers, and floor and ceiling effects were examined in this step. Based on this examination, no data were removed from the dataset as there was no missing data or outlier, and floor and ceiling effects were not observed. The scale items were translated into Turkish and their language equivalency and meanings were tested, and the required items were revised based on the feedback from the subject field experts. Then, the factorial structure of the scale was tested for its validity and reliability in Turkish context.

Figure 1. Screenshots of the used PLD interface (Kokoç & Altun, 2019)

2.1.3. Data analysis

Confirmatory Factor Analysis (CFA) was conducted to examine the construct validity of EFLA 4 for learners. It was also compared with the relevant measurement models. The CFA findings were interpreted based on the fit indices (χ^2/SD, Root Mean Square Error of Approximation - RMSEA, Goodness of Fit Index - GFI, Normed Fit Index - NFI, Standardized Root Mean Square Residual - sRMR, Comparative Fit Index - CFI) as recommended by Jöreskog, Olsson, and Wallentin (2016). Convergent and discriminant validity techniques were also used for construct validity. As for the reliability of the scale and items, item-total correlations, Cronbach Alpha, and Composite Reliability coefficients were computed and evaluated.

2.2. Results

2.2.1. Language validity

The language validity of the scale items was provided through the contribution of the seven professors as experts who have a high level of proficiency in reading, writing and speaking both English and Turkish and have satisfactory knowledge about the literature on the construct of the scale. As the first step, the original items were translated into Turkish by a professor of English language teaching and two professors of instructional technology. The draft form including both Turkish and English items was reviewed by another four-member group (two professors of English language teaching, one professor of Instructional Technology, and one professor of open and distance education) for the appropriateness of the translation through a three-point rubric. Based on the expert evaluations, multi-rater kappa coefficients were computed for each item. The coefficients greater than .60, indicating a good level of consistency, was provided among the experts (Fleiss, 1971). The obtained findings revealed a good level of consistency in terms of language appropriateness as the obtained coefficients for the scale items in the draft form were greater than .80. The scale items were then, implemented with eight undergraduate students for face validity. The findings demonstrated that the scale items were understandable and the scale appeared to measure what it claims to, but the instructions of the scale were required to be detailed more. The draft scale was finalized after the required revisions on the scale.

2.2.2. Measurement model

Figure 2 indicates item-construct parameters of EFLA 4 for learners (standardized factor loadings and the correlations among the factors) obtained through a first-order CFA. Item-construct parameters in Figure 2 show that the standardized factor loadings for the three sub-dimensions of the model range from .65 to .96. The t values showed the factor loadings are significant. According to Brown (2015), the factor loadings are required to be greater than .5, and t values are required to be significant. The CFA results showed that all goodness-of-fit indices were acceptable (χ^2/ df = .96, p > .05, RMSEA = .01, sRMR = .03, CFI = .98, NFI = .96, GFI = .95), indicating that this eight-item instrument had a good model fit when examined with the data from the Turkish online learners. The results from the CFA revealed that similar findings were gathered with the item-construct structure of the original scale developed with a European sample. This indicated that the construct validity of the adapted scale was of high quality.

2.2.3. Construct validity and reliability

In the study, the construct validity of the scale was determined through convergent and discriminant validity techniques. Convergent validity refers to the degree to which the variables measuring the same construct are associated with each other and the construct they belong (Raykov & Marcoulides, 2011). To provide convergent validity, item loadings obtained for each construct are required to be greater than .05, and the average variance extracted for each construct is required to be equal or greater than .05 and to be less than Cronbach Alpha and composite reliability values (Fornell & Larcker, 1981; Hair, Black, Babin, & Anderson, 2010). According to Nunnally and Bernstein (1994), Cronbach Alpha values are required to be greater than .07. Table 3 shows the average variance extracted, Cronbach Alpha, and composite reliability values.

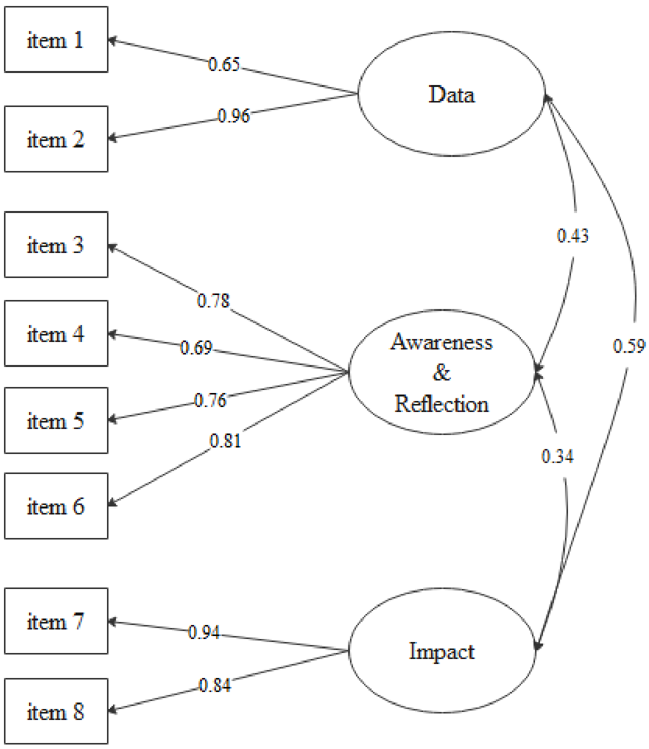

Figure 2. Standardized CFA Solutions for EFLA 4 for learners

Table 3. The average variance values and the reliability coefficients obtained for the constructs

Constructs	Average variance extracted	Composite reliability	Cronbach Alpha (α)
Data	.45	.80	.79
Awareness & reflection	.79	.84	.83
Impact	.58	.88	.89

According to Table 3 the values of the average variance are greater than .05 for the two constructs and less than .05 for one construct. Composite reliability and Cronbach Alpha values obtained for the constructs are greater than .70 as suggested by Nunnally and Bernstein (1994). Although Fornell and Larcker (1981) recommend a value greater than .05 for the average variance explained, they also underline that an obtained composite reliability value greater than .06 for a latent variable is adequate for its convergent validity. Thus, it is concluded that the convergent validity of the scale was obtained as satisfactory and these results indicated the reliability of the total scale is adequate. Discriminant validity refers to the degree to which each latent variable in a measurement model discriminates from others (Farrell, 2010). It requires that the square root of the average variance explained for each construct is not less than the correlation values of each construct with others (Fornell & Larcker, 1981). Table 4 presents the correlations for the constructs in the scale.

Table 4. Correlations among the constructs

Constructs	Data	Awareness & reflection	Impact
Data	$.67^{**}$		
Awareness & reflection	$.43^{*}$	$.89^{**}$	
Impact	$.59^{*}$	$.34^{*}$	$.76^{**}$

Note. $^{**}p < .01$; $^{*}p < .05$.

According to Table 4, correlation coefficients among the constructs are less than the square roots of the average variance values computed for each construct. This finding implies that the discriminant validity of the scale was satisfied. The correlations among the constructs range from .34 to .59 and are moderate and significant.

3. Study 2: EFLA for learners and relation to learner performance

Study 2 aims to investigate the influence of the factors in EFLA 4 for learners and interaction with the dashboard on learners' academic performance. There are various studies in the literature on the influence of the LA dashboards on learner performance. The EFLA instrument for learners developed by Scheffel et al. (2017) measures the metacognitive competencies in relation to LA dashboards. Furthermore, previous studies indicated that behavioral indicators reflecting interaction with LA dashboards can play a key role in evaluating LA dashboards (Kokoç & Altun, 2019; Matcha et al., 2020). In study 2, it was investigated which evaluation constructs together with learner interaction with the dashboard influence learners' academic achievement.

3.1. Method

3.1.1. Participants

The data were collected from 63 out of 66 undergraduate students enrolled in an online course. Their ages ranged from 21 to 26 and the average age was 23.02 (*SD* = 1.75). Three students were excluded from further analysis due to their drop-out from the online course. Of this sample, there were 26 females (41.3%) and 37 males (58.7%). They had taken at least one online/blended course at the university level before the study. The data were collected in the fall term of the 2018 academic year. They are the students of a computer engineering department of a public university and voluntarily participated in the study. All of them were assured as to the confidentiality of their interaction data in the online learning environment.

3.1.2. Context

The study was conducted within a 14-week course entitled "Operating Systems" offered in the third year of a Computer Engineering program at a large university in Turkey. The course was asynchronously delivered four hours per week through Moodle. Several weekly online resources were available such as interactive materials and videos in the learning environment. Weekly assignments and collaborative learning tasks were assigned to the students to reflect upon their knowledge. The PLD was embedded in the Moodle as an LA dashboard. Interaction data reflecting all clicking events of the students were recorded automatically.

3.1.3. Data collection and analysis

The data about the learners' evaluation of the PLD were collected by using the "EFLA 4 for learners" instrument, adapted to Turkish in Study 1. The instrument was distributed to learners as paper-and-pencil, and as a 10-point Likert scale, ranging from "strongly disagree" to "strongly agree." Learners' completion of the scale took about three minutes. The student-generated trace data derived from Moodle time-stamped logs were used to explore learner interaction with the PLD. The interaction data were extracted from the Moodle database by using MySQL queries. The raw data were examined through preprocessing and prepared for the data analysis. Firstly, the raw data were transformed into metrics. Logging traces with timestamps, whenever a student opens and closes the PLD, were stored in a Moodle database table titled "pld_log" which was created by the researchers. Time-related data were logged in a PT format. The time format, PT, does not correspond to a numerical type available in the analysis. Therefore, time traces were automatically converted to a UNIX Time format using a formula created in Excel. To find the time spent viewing the PLD, the time difference between the opening time of the event and the closing or transition of another page was calculated. The time difference reflecting "*time spent on the PLD*" was transformed into seconds. The meaning of the "*total view of the PLD*" is that a student opened and displayed the PLD. In addition, as recommended by Kokoç and Altun (2019), the PLD automatic log-out times was set to 120 seconds as the threshold to prevent bias emerged by fake usage.

Three interaction variables were explored as online behavior indicators reflecting learner interaction with the PLD. In total, the study consisted of six independent variables representing metacognitive and behavioral aspects of evaluating LA dashboards as shown in Table 5.

Table 5. Descriptions of Study 2 variables

Aspects	Variables	Items/Description
EFLA 4 for Learners (Metacognitive Aspects)	Data	Item 1: "For this LA tool it is clear what data is being collected" Item 2: "For this LA tool it is clear why the data is being collected"
	Awareness & Reflection	Item 3: "This LA tool makes me aware of my current learning situation" Item 4: "This LA tool makes me forecast my possible future learning situation given my (un)changed behavior" Item 5: "This LA tool stimulates me to reflect on my past learning behavior" Item 6: "This LA tool stimulates me to adapt my learning behavior if necessary"
	Impact	Item 7: "This LA tool stimulates me to study more efficiently" Item 8: "This LA tool stimulates me to study more effectively"
Interaction with the PLD (Behavioral Aspects)	Total number of the PLD views	Total number of times the PLD was opened
	Time spent on the PLD	Total duration of time spent viewing the PLD (minutes)
	Viewing regularity	(1) Not regularly viewing the PLD at least once every week (2) Regularly viewing the PLD at least once every week

The collected data were analyzed based on the research question. The general overview of the dataset was firstly examined through descriptive statistics. No missing value was identified in the dataset. Pearson's correlation analysis was conducted to check the significance of the relationships among the continuous variables. Binary logistic regression was conducted to investigate a predictive model considering the cause-effect relationship between the variables. Logistic regression analysis is a multivariate statistical technique used to compute the probability of the effects of independent variables (predictors) on dependent variables as well as identifying the risk factors (Field, 2013). Logistic regression was used in the current study as it does not require normality and common covariance assumptions, and the categorical variables might be used in the predictive model both as dependent and independent variables. The dependent variable used in the analysis includes two categories: (0) failing the course (unsuccessful), (1) passing the course (successful). The continuous independent variables were the constructs within EFLA 4 for learners, the total number of the PLD views, and the time spent on the PLD. The categorical independent variable was viewing regularity.

Odds ratios (ORs) with 95% Confidence Intervals (CIs) were computed to reveal the probabilities of the sustained attention for each independent variable. The Omnibus test with the model coefficients was used to test the relationships between the combinations of the dependent and independent variables. Nagelkerke R^2 coefficient was used to determine how the independent variables explain the variance of the dependent variables. Hosmer-Lemeshow test was used to investigate the goodness of the model-data fit. During the data analysis, .05 was adopted as the level of significance with the two-tailed tests. Finally, the data collection and analyses were conducted in compliance with the ethical guidelines.

3.2. Results

The correlations between the EFLA instrument scores and the interactions with the PLD in Moodle were presented in Table 6. The table shows that there was a significant correlation between data and impact dimension ($r = .26$, $p < .05$). Also, a significant correlation was found between impact and awareness & reflection dimension ($r = .28$, $p < .05$). Time spent on the PLD was found as significantly correlated with the total view of the PLD ($r = .40$, $p < .05$).

Table 6. Descriptives and correlation analysis results

	Data	Awareness & reflection	Impact	Total view of the PLD	Time spent on the PLD
Data	1				
Awareness & reflection	.18	1			
Impact	$.26^{*}$	$.28^{*}$	1		
Total number of the PLD views	.03	.13	.10	1	
Time spent on the PLD	.13	.07	.23	$.40^{**}$	1
Mean	7.65	7.40	7.23	58.3	99.8
SD	2.25	1.92	2.21	17.5	32.2
Minimum	1	2	2	40	29
Maximum	10	10	10	101	217

Note. $^{**}p < .01$; $^{*}p < .05$.

As for the unique contribution of the study variables in predicting course success (failing the course = 0, passing the course = 1), logistic regression analysis was employed. The independent variables including the EFLA constructs and interaction behaviors with the PLD were included in the logistic regression analysis as the possible predictor variables in the regression model. Table 7 shows the results of the logistics regression analysis.

Table 7. Logistic regression results predicting learner performance

Independent variables		95% CI for Odds Ratio (OR)			
		b (SE)	OR	Lower	Upper
EFLA for Learners (Metacognitive Aspects)	Data	.27(.19)	1.31	0.90	1.92
	Awareness & Reflection	$.65(.25)^{**}$	1.92	1.16	3.18
	Impact	$.67(.27)^{**}$	1.95	1.14	3.31
Interaction (Behavioral Aspects)	Total view of the PLD	$.07(.04)^{*}$	1.08	1.01	1.15
	Time spent on the PLD	.01(.02)	1.01	0.98	1.04
	Viewing regularity of the PLD	-.75(.90)	0.47	0.08	2.77

Note. Reference category for regularity is "1"; $^{**}p < .01$; $^{*}p < .05$.

The results showed that the logistic regression model is statistically significant (-2 log L = 44.17, chi-square = 42.77, $p < .01$). The results from the Hosmer-Lemeshow test revealed that the model has an acceptable fit and the data-model fit is satisfactory (Chi-square = 2.01, $p > .05$). According to the Nagelkerke R^2 values, all of the independent variables account for 66% of the variance in the dependent variable. The classification table shows that the total ratio of the correct classification of the model is 82.5%. In other words, 82.5% of students were correctly classified. Three variables were statistically significant predictors of group membership: Awareness & reflection ($b = .65$, $p < .01$), impact ($b = .67$, $p < .01$), and total view of the PLD ($b = .07$, $p < .05$). It was concluded that the online students who have high awareness & reflection score (OR = 1.92, 95% CI = 1.16-3.18), high impact score (OR = 1.95, 95% CI = 1.14-3.31), more frequently interact with the PLD (OR = 1.08, 95% CI = 1.01-1.15) will more likely successful in the course. As indicated by the odds ratios, membership in the successful group was 1.92, 1.95, and 1.08 times more likely for every one-unit increase in the scores of awareness & reflection and impact dimensions, and total view of the PLD, respectively.

4. Discussion, conclusion, and limitations

This multiple study investigation aimed to adapt and validate the instrument of EFLA 4 for learners into the Turkish context (Study 1), and to examine how metacognitive and behavioral factors predict learners' academic performance (Study 2). The main aim was to explore whether the dimensions of the EFLA for learners and interaction variables were predictive of the extent to which learners completed the online course successfully.

In Study 1, EFLA 4 instrument was adapted to Turkish context. Study 1 appears to be the first research that validates an instrument in the context of evaluating LA dashboards. The results of Study 1 provided strong support for the psychometric qualities of the three-factor, eight-item EFLA 4, including construct validity and reliability. The reliability statistics showed that the instrument had a good level of internal consistency. It was revealed that the Turkish version of EFLA 4 instrument demonstrates the dimensions of the original scale: (1) Data, (2) awareness and reflection, and (3) impact. The CFA results confirmed the three-component structure of EFLA 4 (Scheffel, 2017). Thus, it was revealed that the Turkish version of EFLA 4 is a valid and reliable instrument to evaluate LA dashboards. As a measurement tool, the adapted version of EFLA 4 instrument has been one of the first attempts to thoroughly evaluate LA dashboards in the Turkish higher education context. In the literature, most LA studies have used EFLA 4 to evaluate and compare LA dashboards based on student perceptions (e.g., Broos et al., 2018; Toisoul, 2017). These studies indicate that this instrument is a valuable tool to measure and compare the impact of LA dashboards on educational practices in online and/or blended learning contexts. Thus, it is hoped that adapting the instrument into different cultures could be useful for LA researchers and learning designers from various countries as well as enhancing its generalizability.

In Study 2, we investigated the impact of the dimensions in EFLA 4 for learners and interaction with the PLD on learners' academic performance. The results of Study 2 have led to two issues worthy of further discussion. The first is that the score of "data" dimension did not predict the learner performance although the other EFLA 4 dimensions, "awareness & reflection" and "impact," are the predictors of learner performance. This result may be explained by the relationship between the theory of self-regulated learning and LA. LA plays a key role in helping online learners develop their self-regulated learning skills (Broadbent, Panadero, Lodge, & Barba, 2020). According to the LA process model with four stages based on self-regulated learning (awareness, reflection, sense-making, and impact), LA dashboards visualize learning traces to support awareness, reflection, and sense-making of learners about their learning process (Verbert et al., 2014). If online learners gain insights on their performance through visualized information and reflect them on their learning process, they can change learning behaviors to achieve intended learning goals. Similarly, previous studies indicated that promoting learners' online self-regulation skills via interventions provided by LA dashboards will lead to improved learning performance (Araka, Maina, Gitonga, & Oboko, 2020; Jansen et al., 2019). Thus, our results are in line with those studies focused on the relation between self-regulated learning and LA. The second issue from Study 2 is to discuss the characteristics of PLD with the obtained findings. The PLD used in this study visualizes learners' online behavioral indicators and provides recommendations based on a predictive model. It additionally visualizes the criterion value about a specific indicator expected by a teacher, individual performance of learners, and the average performance of the class on the same graph. It, therefore, has a relatively goal-oriented structure. The results obtained in this study might be the result of these characteristics of the used PLD.

Consistent with the previous studies (Broos et al., 2020; Kokoç & Altun, 2019), Study 2 found that total view of the LA dashboards was a predictor of learning performance and the learners who interacted with the LA dashboard more frequently were also successful in the course more likely. This evidence suggests that students' actual LA dashboard usage is critical in evaluating the impact of LA dashboards. Surprisingly, time spent on the LA dashboard and viewing regularity of the PLD was not found as the predictors of learner performance. This result implies that successful learners pay more attention to the recommendations and interventions offered by the LA dashboard whenever they need them. Another possible explanation for this result is that learners' awareness of the learning process might be enhanced. A learner, who received feedback about his/her learning performance via the PLD, would change his/her learning strategy to improve his/her achievement based on the provided real-time recommendations. For this reason, these results support the conclusion that LA dashboards offering personalized recommendations and interventions influence learner achievement (Ifenthaler & Yau, 2020). In addition, the results should be interpreted with caution because this study is only concerned with one class and one LA dashboard titled PLD. It is important to bear in mind that the results could be different when other LA dashboards are examined in different targets.

The results of Study 2 may make a significant contribution to the institution-wide adoption of LA dashboards and systems. There are some challenges for the implementation of LA initiatives to overcome in higher education (Leitner, Ebner, & Ebner, 2019). Although studies on the design and implementation of LA systems are increasing in the literature, the use of them by large-scale educational institutes worldwide is still at an early stage (Colvin, Dawson, Wade, & Gašević, 2017; Gašević, Tsai, Dawson, & Pardo, 2019). Therefore, it can be inferred that the EFLA instrument for learners is useful for providing evidence about the effects of LA dashboards to educational decision-makers and instructors to facilitate their institutional adoption process and system at higher education institutions. Our results on predicting learner performance based on the EFLA dimensions and interaction with the PLD support this inference. Thus, combining interaction data and self-report data in evaluating the impact of LA dashboards and systems may provide higher education institutions with more valid results and insightful information.

In conclusion, this multiple-study investigation contributes to the generalizability of the EFLA for learners and highlights the importance of both metacognitive and behavioral factors for the impact of LA dashboards on learner performance. In this regard, the current study aimed to address the call for further research on the impact of LA interventions as pedagogical tools on learning (e.g., Jivet et al., 2018; Matcha et al., 2020) and to contribute to the literature by revealing the influence of both metacognitive and behavioral factors on learning performance. The study suggests that evaluation of LA dashboards in terms of metacognitive and behavioral indicators can help learning designers and instructors examine their impact on learners' performance and improve their online learning experience. It is hoped that the insights gained from this study may be of assistance to researchers and learning designers for evaluating the impact of LA dashboards. Even if it is not addressed in the study, it should be kept in mind that qualitative feedback of students may also be important in the evaluation of LA dashboards (Yoo & Jin, 2020).

There are several limitations to this study, which provide directions for future studies. The first limitation is the small sample size. Study 2 was carried out in one online course at a higher education institution. In future studies, large-scale and longitudinal studies can be conducted in various courses with the participation of more learners to test the prediction model of LA dashboards. Further LA studies are needed to confirm the prediction model in other online learning contexts. The second limitation is that learners' emotional and cognitive individual differences were not included in the prediction model. In future studies, students' emotional status when using LA dashboards can be followed using facial expression recognition. Whether students' different cognitive traits and self-regulation levels have a role in evaluating the impact of LA dashboards can be examined in future studies. The final limitation is that this study was conducted in a context where a single LA dashboard was used. Future studies might focus on LA evaluation for learners and instructors in the online learning contexts where multiple LA dashboards were used. The findings from the evaluation of other LA dashboards in various contexts would contribute to the generalizability of the findings of this study.

References

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—Systematic review (2008–2018). *Research and Practice in Technology Enhanced Learning, 15*(1), 1-21. doi:10.1186/s41039-020-00129-5

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education, 30*(3), 572-598. doi:10.1007/s12528-018-9186-0

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies, 10*(4), 405-418. doi:10.1109/TLT.2017.2740172

Broadbent, J., Panadero, E., Lodge, J. M., & de Barba, P. (2020). Technologies to enhance self-regulated learning in online and computer-mediated learning environments. In M. J. Bishop, E. Boling, J. Elen, & V. Svihla (Eds.), *Handbook of Research in Educational Communications and Technology: Learning Design* (pp. 37-52). Cham: Springer International Publishing. doi:10.1007/978-3-030-36119-8_3

Broos, T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2020). Learning dashboards at scale: Early warning and overall first year experience. *Assessment & Evaluation in Higher Education, 45*(6), 855-874. doi:10.1080/02602938.2019.1689546

Broos, T., Verbert, K., Langie, G., Van Soom, C., & De Laet, T. (2018, March). Multi-institutional positioning test feedback dashboard for aspiring students: Lessons learnt from a case study in flanders. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 51-55). New York, NY: ACM. doi:10.1145/3170358.3170419

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Colvin, C., Dawson, S., Wade, A., & Gašević, D. (2017). Addressing the challenges of institutional adoption. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (1st ed., pp. 281-289). doi:10.18608/hla17

Farrell, A. M. (2010). Insufficient discriminant validity: A Comment on Bove, Pervan, Beatty and Shiu (2009). *Journal of Business Research, 63*(3), 324-327. doi:10.1016/j.jbusres.2009.05.003

Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Thousand Oaks, CA: Sage publications.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378-382. doi:10.1037/h0031619

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. doi:10.2307/3151312

Gašević, D., Tsai, Y.-S., Dawson, S., & Pardo, A. (2019). How do we start? An Approach to learning analytics adoption in higher education. *The International Journal of Information and Learning Technology, 36*(4), 342-353. doi:10.1108/IJILT-02-2019-0024

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). New Jersey, NJ: Pearson.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Howell, J. A., Roberts, L. D., Seaman, K., & Gibson, D. C. (2018). Are we on our way to becoming a "helicopter university"? Academics' views on learning analytics. *Technology, Knowledge and Learning, 23*(1), 1–20. doi:10.1007/s10758-017-9329-9

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), 1-30. doi:10.1007/s11423-020-09788-z

Jansen, R. S., Van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A Meta-analysis. *Educational Research Review, 28*, 1-20. doi:10.1016/j.edurev.2019.100292

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *European Conference on Technology Enhanced Learning* (pp. 82-96). Springer, Cham. doi:10.1007/978-3-319-66610-5_7

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 31-40). New York, NY: ACM. doi:10.1145/3170358.3170421

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). *Multivariate analysis with LISREL*. Basel, Switzerland: Springer.

Kim, J., Jo, I. H., & Park, Y. (2016). Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review, 17*(1), 13-24. doi:10.1007/s12564-015-9403-8

Kokoç, M., & Altun, A. (2019). Effects of learner interaction with learning dashboards on academic performance in an e-learning environment. *Behaviour & Information Technology*, 1-15. doi:10.1080/0144929X.2019.1680731

Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The Efficacy of learning analytics interventions in higher education: A Systematic review. *British Journal of Educational Technology, 50*(5), 2594-2618. doi:10.1111/bjet.12720

Leitner, P., Ebner, M., & Ebner, M. (2019). Learning analytics challenges to overcome in higher education institutions. In D. Ifenthaler, D. K. Mah, & J. Y.-K. Yau (Eds.), *Utilizing learning analytics to support study success* (pp. 91-104). Cham, Switzerland: Springer. doi:10.1007/978-3-319-64792-0_6

Long, P. & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40.

Matcha, W., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A Self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226-245. doi:10.1109/TLT.2019.2916802

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science, 21*(1), 110-133. doi:10.3217/jucs-021-01-0110

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.

Samuelsen, J., Chen, W., & Wasson, B. (2019). Integrating multiple data sources for learning analytics—Review of literature. *Research and Practice in Technology Enhanced Learning, 14*(11), 1-20. doi:10.1186/s41039-019-0105-4

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*(4), 117-132.

Scheffel, M., Drachsler, H., & Specht, M. (2015). Developing an evaluation framework of quality indicators for learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 16-20). New York, NY: ACM.

Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., & Specht, M. (2017). The Proof of the pudding: Examining validity and reliability of the evaluation framework for learning analytics. In *European Conference on Technology Enhanced Learning* (pp. 194-208). Tallinn, Estonia: Springer. doi:10.1007/978-3-319-66610-5_15

Scheffel, M. (2017). *The Evaluation framework for learning analytics*. Heerlen, Netherlands: Open Universiteit. Retrieved from https://core.ac.uk/download/pdf/95759389.pdf

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior,78*, 397–407. doi:10.1016/j.chb.2017.06.030

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving learning at a glance: A Systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies, 10*(1), 30-41. doi:10.1109/TLT.2016.2599522

Toisoul, C. (2017). *Design, implementation and evaluation of two widgets for a MOOC's learning analytics dashboard* [Unpublished master's thesis] Open University of the Netherlands.

Wong, B. T. M., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education, 7*(1), 7-28. doi:10.1007/s40692-019-00143-7

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500-1509. doi:10.1177/0002764213479363

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing, 18*(6), 1499-1514. doi:10.1007/s00779-013-0751-2

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98-110. doi:10.1016/j.chb.2018.07.027

Yoo, M., & Jin, S. (2020). Development and evaluation of learning analytics dashboards to support online discussion activities. *Educational Technology & Society, 23*(2), 1–18.