IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 October 2023, accepted 23 October 2023, date of publication 25 October 2023, date of current version 31 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327334

RESEARCH ARTICLE

What Attention Regulation Behaviors Tell Us About Learners in E-Reading?: Adaptive Data-Driven Persona Development and Application Based on Unsupervised Learning

YOON LEE, GOSIA MIGUT, AND MARCUS SPECHT
Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 XE Delft, The Netherlands
Leiden-Delft-Erasmus Centre for Education and Learning, 2628 XE Delft, The Netherlands

Corresponding author: Yoon Lee (y.lee@tudelft.nl)

This work was supported by the Leiden-Delft-Erasmus Centre for Education and Learning (LDE-CEL).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the TU Delft Human Research Ethics Committee under Application No. 1793, and performed in line with the General Data Protection Regulation (GDPR).

ABSTRACT Different individual features of the learner data often work as essential indicators of learning and intervention needs. This work exploits the personas in the design thinking process as the theoretical basis to analyze and cluster learners' learning behavior patterns as groups. To adapt to the learning practice, we develop data-driven personas by clustering learners' features based on factual learning outcomes (i.e., knowledge gain, perceived learning experience, perceived social presence) based on unsupervised learning, a more accessible and objective intervention design strategy for e-reading practices. Using the Chi-square test, we quantitatively evaluate different clusters driven by various unsupervised learning methods on the multimodal SKEP dataset. Furthermore, for a more practical real-life application, we achieved automatic persona prediction based on the attention regulation behaviors of learners. The subject-independent evaluation results indicate the best classification accuracy of 70% for the four-level classification task, differentiating three personas of learners with needs and another without feedback needs. It also shows that time-based sampling on both independent and cumulative learner behaviors works as robust predictors of learner personas, achieving a stable accuracy range of 65%-70% throughout the e-reading with the SVM classifier. Our work inspires the design of a real-time feedback loop for e-learning based on conversational agents.

INDEX TERMS Data-driven persona development, human–robot interaction, instructional design, learning analytics, unsupervised learning.

I. INTRODUCTION

Understanding users is an essential system design requirement for usability and better-perceived services [1]. It is especially well-emphasized for digital product (e.g., software, online courses, eBooks) design since poor user requirement engineering causes a perception gap between users and the practitioners, while users are often veiled with unknown varieties [2]. Likewise, understanding learners' traits and needs has been a critical challenge in e-learning intervention design. Especially, learning and learner necessities in e-learning tend to be more specific due to the physical absence of human educators and peers, while keeping close attention and engagement remains a challenge compared to the traditional on-site learning environment [3]. E-learning is becoming a mainstream education with recent social changes (e.g., COVID-19 [4]) with widespread e-learning platforms, digital devices, various forms of learning interventions,

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

feedback agents, and modalities (e.g., social robot [5]). Those are designed for diverse learning objectives (e.g., formal and informal learning [6]) in e-learning and hybrid education settings [7]. However, e-reading system development approaches for better engagement seem scarce compared to the field's rapid growth and necessities [8].

As a user-centric decision-making tool [1], [2], the concept of "personas" has taken place in various domains, such as healthcare, knowledge management, social media, software development, and games [9] since its first appearance in 1999 [10]. Persona was devised as a practical and iterative [1] interaction design tool [10] in the design thinking process [11]. Persona has been further elaborated as hypothetical "archetypal" representatives [12] with specific needs, goals [13], attitudes, skills, roles, and expected behaviors [1], [14]. Those imaginary presences are believed to deliver certain behavioral traits, perceptions, and beliefs of specific segments of people in the real-world [9]. Persona is meaningful in providing a shared understanding of target users, their needs, and system usage [15]. Recent advancements in big data, data science algorithms, and data infrastructures have made data-driven persona development and analytics more accessible than ever [9], that has been traditionally done by several dozens of experts [16] for months and years [17]. Even though feedback personalization in education has become more accessible with more accurate predictions available through sensors, algorithms, and computing resources, according to our best knowledge, data-driven persona developments and the following learning analytics in education for feedback system development, especially in e-reading, have yet to be attempted.

In this regard, we develop data-driven personas using user modeling techniques based on unsupervised learning and its analysis [1], [12]. Instead of designing the feedback first and fitting learners with somewhat arbitrary criteria, we utilize the factual learning outcomes (i.e., knowledge gain, perceived learning experience, perceived social presence) collected from learners and use them as features for clustering learners, serving as the objective ground truths for analysis. Our data-driven persona approach is especially valuable for instructional designers and practitioners who lack standardized methods for analyzing learners as groups for further learning analytics and intervention design. Even with the same set of learner data and analytical objectives, it is nearly impossible to share the same criteria when developing a persona with somewhat manual and qualitative methods, with different perceptions and experiences of evaluators. Such deviations in decision-making inevitably lead to subjective and inconsistent learner clustering, which hinders timely and adequate intervention provision for learners.

Not merely working on the quantification and diversification of clusters, which has been a focus of early development of quantitative persona [9], this paper strives for deeper insights into learner analysis for e-reading intervention design by connecting the quantified persona model to statistical analysis. We explore utilizing data-driven learner persona to provide valuable insights into who learners are in terms of their categorical divisions, feature compositions, and their needs as a group in one grasp with statistical interpretations [18] and recognize them with classical machine learning classifiers.

From the perspectives of instructional designers, it is also more practical and feasible to understand the semantic and statistical meaning of the core features of groups and design interventions for them than making specific rules for individual features that deliver fragmentary and linear information. Feature-based learner divisions often end up deriving hypothetical learners with flat and stereotypical characteristics, which limits deeper insights about learners. To compensate for the limitation, we suggest the intervention design based on the data-driven persona using learners' factual learning outcomes as major dimensions of learner clustering.

Furthermore, we address a core issue of the utility of the above automatically generated persona categories for the following intervention: predicting the learners' persona categories for robust and timely learning interventions. To this end, we utilize human-labeled video samples from the SKEP dataset [19] to train machine-learning models to achieve the prediction of learners' persona categories based on their real-time and accumulated behaviors. The methods are validated via subject-independent protocols to ensure the generalizability of our method. Our automatic data-driven persona development framework and its prediction can assist in forming a feedback loop for better learning outcomes and experiences [13].

This work follows the procedure of 1) feature engineering on various types and levels of factual learning outcomes, 2) implementation of various unsupervised learning techniques and validation, 3) archetype extraction and data-driven persona development based on quartile analysis, and 4) learner persona prediction based on attention regulation behaviors (see Fig.1). We first utilize the multimodal SKEP dataset with the 25 multimodal features that include various matrices (e.g., pre-post test, Attrikdiff questionnaire, Social Presence questionnaire, and human annotation of six attention regulation behaviors for every second on approximately 40 hours of video data) to understand diverse perspectives of factual learning outcomes, collected from 60 higher education learners. It is a dataset that has been carefully designed and collected to understand learner behaviors and internal attributes in e-reading with emphatic and metacognitive feedback prompts from conversational agents. See [5] for the experimental details.

As suggested in the recent review of [9], we implement and compare various clustering methods on the dataset, such as k-means clustering, hierarchical clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and spectral clustering, that represent various modeling methods (i.e., centroid, hierarchy, density, graph) with various hyperparameters, which have further been cross-validated via Chi-square test. Subsequently, we conduct statistical analysis on each cluster to find distinctive and significant clusters features and draw our insights based on it [15]. Using classical machine learning models, such as AdaBoost, SVM, kNN, and Random Forest classifiers, we develop the behavior-based prediction model for personas on multi-levels as a part of the potential feedback loop for e-reading. All in all, our research questions are as listed as followings.

- RQ1. How can learner features best reflect learners' performances, experiences, and perceptions of conversational agents' interventions in e-reading?
- RQ2. How can unsupervised learning methods be used for learner pattern clustering and validation?
- RQ3. How can we extract valuable archetypes of learners from different clusters and develop data-driven personas based on them?
- RQ4. Can we predict learner personas based on attention regulation behaviors?

To summarize, our contributions are listed below.

1) To our best knowledge, it is the first attempt to extend the data-driven persona development framework to e-reading with conversational agents. Personas provide learner clusters with a more concrete, multi-dimensional synthesis of learner features that represent learner categories differently from the cumbersome manual divisions of learners. Our feature engineering and the clustering result can provide the foundation for future data-driven persona-based learning analytics and intervention design for learners and instructors.

2) Despite its necessity, an extensive comparison among various clustering methods with learner data has yet to be attempted. We implement four unsupervised models with various modeling methods: k-means, hierarchical, DBSCAN, and spatial clustering. We conduct a Chi-square test to find the similarity among clusters derived by different modeling methods to validate clusters suggested by each other. It is a valuable attempt for future researchers' model implementation decisions for unsupervised learning-based clustering.

3) We explore the application of proposed data-driven personas: predicting the learners' persona categories for robust and timely learning interventions. We train machine-learning models to predict learners' persona categories based on their real-time and accumulated attention regulation behaviors. It will provide a foundation for a solid HRI feedback loop design in e-reading, promoting knowledge gain, perceived learning experiences, and perceived social presence of learners. It is beneficial for the following learning analytics and instructional design in e-learning for adaptive feedback implementation.

II. RELATED WORK

In recent years, more technology-enhanced learning and machine learning approaches have taken roles to reveal hidden patterns in learning and help with the intervention design for the education administration and instructional design [20]. This section introduces previous approaches using unsupervised machine learning methods in diverse learning scenarios. The topic will be more specified with the review of data-driven persona development, which will be the focus of this work. At the same time, the section on learning analytics indicators on e-reading will help us derive important learner features and further analysis. Lastly, we develop behavior-based machine learning models to bridge learning analytics and data-driven persona prediction.

A. UNSUPERVISED LEARNING METHOD FOR EDUCATION

In this section, we focus on input features, objectives, and validation methods that have been applied to unsupervised methods in education. Reference [21] focused on individual factors (e.g., gender, age, region, highest education, Index of Multiple Deprivation (IMD) bands, disability) and the data from the previous course (e.g., studied credit, number of clicks), to gauge the student involvement and their achievements in online-learning, using k-means clustering. Reference [22] segmented the students' learning behaviors, utilizing data layers of 22 features (e.g., in information acquisition, solution construction, and solution assessment). It applied a t-test to represent the significance of particular features and a sparse k-means clustering for the feature selection and the final segmentation of learners, respectively. Reference [23] has used k-means clustering with multimodal indicators, such as eye-tracking, physiological, and motion-sensing data, to automatically identify learners' productivity states (e.g., neutral, collaborative, non-collaborative) in collaborative learning. The model has been evaluated through correlation analysis between learner states, task performances, and learning gains. Reference [24] has utilized student posts (i.e., textual dialogues) in MOOC for K-12 education to understand functional similarities of discourses (e.g., questioning, statements, reflections, scaffolding, references) via the k-means clustering, combined with bayesian information criterion. For validation, machine-generated clusters have been compared with human-coded clusters. Reference [25] has divided learners based on their answers to system questions, comparing clusters from hierarchical (i.e., hierarchical clustering) and non-hierarchical (i.e., k-means clustering) clusters. For validation purposes, the within-group and between-group squared sum have been evaluated, indicating that the non-hierarchical method enables more detailed clustering results than the hierarchical method. Reference [26] has used 12 engagement metrics (e.g., number of logins, number of forum reads, number of forum posts, quiz reviews, assignment lateness, assignment submission) to cluster higher education learners with k-means clustering method, aiming at personalized online education. Various values of k have been applied to draw multi-levels of learner engagement clusters. Reference [27] has segmented higher education learners' using the k-means clustering method based on learners' academic performance (e.g., students' entry mode, residential category, scores of courses, age, post-UTME scores, GPA, gender, class of degree) and validated the clusters with a self-organizing map.

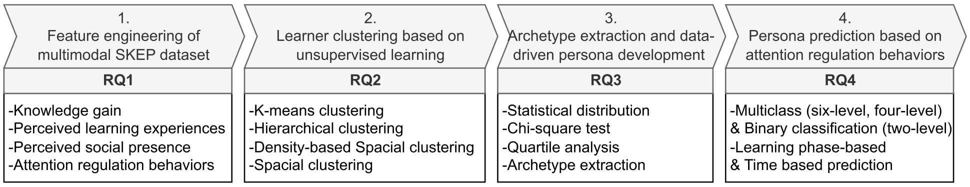


FIGURE 1. Our work covers multimodal data processing, learner categorization based on unsupervised learning methods, archetype extraction based on statistical analysis, and personas prediction based on learner behaviors.

All in all, 1) from available implementation cases, it has been observed that the *k-means clustering method has been dominantly applied*. The only exception was [25], which has applied a hierarchical modeling method (i.e., hierarchical clustering) and a non-hierarchical modeling method (i.e., k-means clustering) to cross-validate each cluster. 2) *Large datasets from online education platforms have often been used* as input for modeling due to the easily accessible data. However, because such a dataset often only conveys rather superficial quantitative log data (e.g., demographics of learners, number of clicks), result analysis has shown its limitation without in-depth insights on the specific topics. It differentiates the application of our SKEP dataset, which has been exclusively designed to understand learner behaviors (i.e., attention regulation behaviors), performances (i.e., knowledge gain), and internal states (i.e., perceived learning experiences, perceived social presence) with metacognitive feedback prompts from conversational agents in e-reading. 3) *There has yet to be a fixed validation method for modeling results* due to the nature of the unsupervised machine learning method, which relies on practitioners' further interpretations of results. Thus, various validation methods (e.g., within-group squared sum, t-test) have been applied based on researchers' needs on model implementations. 4) Though all works have represented learning analytics as outcomes to certain degrees, *there has yet to be an attempt to directly analyze the effect of feedback prompts of conversational agents and connect them with intervention loops*. It supports our attempt to develop an automatic data-driven persona and behavior-based prediction model that expands the feedback loop in e-reading with conversational agents.

B. DATA-DRIVEN PERSONA DEVELOPMENT APPROACHES

Personas have been developed as representative figures that carry diverse user roles (e.g., users' characteristics, needs, and behaviors), profiles (e.g., demographic characteristics, motivation, goals, and personalities of users), segments (e.g., user relationship to the system, fundamental needs, characteristics of groups), and extreme characters (e.g., radical personalities of users), that delivers personal, technical, relationship, opinion information of users [1]. It started as a somewhat manual and qualitative analytics tool until recent years' proliferation of data, computing resources, and machine learning techniques [9]. The data-driven persona has been developed to compensate for the limitations of manual and qualitative persona: 1) high cost with long development duration with high monetary investments, 2) lack of objectivity and rigor due to the subjective criteria, 3) lack of scaling, which often leads to poor adaptation in big-scale data, 4) misrepresentation of clusters due to different insights and expertise of practitioners, and 5) expiration of validity with sample updates [9], [12]. The persona has evolved from the 1) qualitative method and 2) qualitative method with further quantitative validation in the early development. 3) Quantitative personas [28] have taken place with the implementation of unsupervised machine learning techniques, which is often further supported by the qualitative interpretations of practitioners on input indicators and clusters. Thus, the recent challenges of data-driven persona development have mostly come from data quality as the model input and interpretations of unsupervised models (e.g., data quality, data availability, method-specific weaknesses, human and machine biases [9]). The inputs of recent work of data-driven persona have ranged from accessible mouse-click log data to pricey data from surveys, self-reports, interviews, and user observations [14]. Regarding model implementation, a recent review has represented k-means clustering as the most used algorithm, followed by non-negative matrix factorization and hierarchical clustering. Various methods, such as latent semantic analysis [29], principal component analysis [9], and Cohen's Kappa [29], have been applied to best describe the distinctive cluster features and cluster validation using the clusters and new sample sets, respectively. Though no standardized methods exist for cluster validations, the most common data-driven persona validation methods have been calculating the Euclidean distance between the different variables or testing the Chi-square. At the same time, subject experts validated the cluster by reviewing the clusters in a few pieces of literature [12].

To conclude, 1) the framework of *data-driven persona development has yet to be applied to the field of education*, which seems to be especially valuable for instructional design practitioners and researchers by representing learner groups with the synthesis of learner features. 2) *Comparative research among various unsupervised methods has been suggested but did not take place* in the field of data-driven persona development [9], which encourages our attempt to compare modeling methods (i.e., centroid, hierarchy, density, graph-based) and use each other for the cluster validation.

C. LEARNING STYLES AND OUTCOMES

This subsection discusses the endeavors of past research for understanding different learning styles and subsequent learning outcomes. The foundational framework of this section is the literature by Bruyckere et al. [30], representing that all learners exhibit different learning styles connected to different outcomes. In the framework, three challenges of correlating the learning styles and their outcomes have been derived as follows: 1) Most people do not fit one particular learning style [31], that presents the necessity for more multi-dimensional understanding of learning styles, often shown in forms of various learning behaviors and cognitive process. 2) Also, the indicators or measures used to evaluate learning styles often need to be more adequately designed or connected to learning outcomes [32]. It emphasizes the importance of finding valid evaluation methods with an understanding of context and domain knowledge of learning. 3) Lastly, linking myriad learning styles to learning outcomes has been seen as cumbersome since it is often not straightforward to understand [33]. It supports our attempt to "cluster" learners based on their multi-dimensional features for its intuitive and practical understanding for further connection with interventions. In the following section, we specifically look into diverse indicators and measures for understanding attentive e-reading, which is our targeted learning scenario.

D. INDICATORS AND MEASURES OF ATTENTIVE E-READING

This section investigates various indicators to evaluate learners' e-reading with emphatic and metacognitive feedback prompts with conversational agents, especially based on Human-Robot Interaction (HRI). Analytics4Action Evaluation Framework (A4AEF) [34], an evidence-based learning analytics intervention evaluation protocol for online learning, has been applied, that has empathized teaching presence, cognitive presence, emotional presence, and social presence as core components of learning interventions. In the subsection of *knowledge gain*, various feedback strategies from human educators and the existing systems are studied for insights into the feedback for better learning performances [35]. In the subsections of *perceived learning experiences* and *perceived social presence*, we investigate how multimodal feedback from systems is utilized and perceived by learners. In the subsection of *attention regulation behaviors in e-reading*, we investigate observable behavioral cues of learners that can be collectively used with other learning analytics measures to understand learners' attentional states during e-reading practices.

1) KNOWLEDGE GAIN

Knowledge gain is the primary goal of e-reading activities and vice versa; reading has been one of the most fundamental forms of knowledge gain in higher education [5]. In recent years, e-reading has become more commonplace with the rapid digitalization of education and the widely-used smart devices [3]. Reading comprehension, reducing reading times, and increasing meta-cognition have been considered the primary learning objectives in e-reading, based on the ability to sift vital information from others [36]. The knowledge gain evaluation has been conducted diagnostic, formative, and summative [37], with questions about finding global or local information, text organization, identifying main ideas, matching the sequence of events, and conclusions [36]. Several e-reading strategies have been suggested for better knowledge gain: exploring, finding, analyzing, and evaluating the reading material [38]. Furthermore, specific behavioral instructions have been suggested, such as oral reading and revisiting mistakes [39].

Setting up the short-term goal related to the result (i.e., product goal) and the process (i.e., process goal) has also been suggested, known to improve learners' self-efficacy, which positively affects the choice of activities, effort, persistence, and achievement of learners [40]. Observing the process goal, such as correct answers, test scores, and grades, was suggested [40]. As known to negatively affect student motivation, learning capabilities, and skill acquisition [40], resolving self-doubts in the learning process has also been suggested as a relevant feedback role. Regarding human educators' feedback provision pattern in reading, more self-corrections were expected from high performers, while more frequent feedback was given to learners with lower learning achievements [39]. Some human instructors focused more on contextual cues that are more relevant to our work, aiming at overall comprehension. In contrast, some focused on specific cues that are more relevant to the proficiency of certain skills [39].

2) PERCEIVED LEARNING EXPERIENCES

Perceived experience is often interpreted and evaluated as User Experience (UX) in diverse domains. One commonly referred definition of UX is users' perceptions and responses toward specific products and services based on users' usage and anticipations [41]. With the emergence of the AI-based approach, there has been an increasing interest in connecting learners' opinions and emotions to machine learning technologies for finding impacts of distance learning on students and teachers [42]. The increasing roles of conversational agents in everyday activities make the consideration of UX in HRI more important [35], which affects the overall system acceptance [43]. In this section, we look into the UX of the HRI, which is our focus as an intervention medium.

The recent HRI evaluation has been criticized for its questionable validity and reliability of measures [44]. It seems to be partial because that UX can only be understood subjectively through perceived users' internal states [44], which makes the evaluation validation more critical. Another comes from the fact that UX design implementation and evaluation of intervention can only be understood through context, which requires the whole iteration as a package but often takes place separately in most practices [35]. However, since HRI is a relatively young research field, we still need the common theories, methods, models, and tools [35] and dedicated studies for HRI design for specified objectives.

Though HRI evaluation can be partially inspired by the field of Human-Computer Interaction (HCI), HRI needs more specific evaluation methods as opposed to comparatively traditional, passive, and computer-based facets of HCI [41]. In the same line, [35] indicated the necessity of a systematic approach in HRI evaluation to guarantee a positive user experience regarding the system's acceptance, usability, learnability, safety, trust, and credibility. The presence of robot agents also makes the understanding of robots more essential, such as contact with humans (e.g., physical robot, virtual robot), robot functionality (e.g., adaptive function), robot roles (e.g., assistant, companion, partner), and social skills of the robot (e.g., desirable to fundamental level) [44]. Understanding the functions of conversational agents' characteristics (e.g., speaking style, personality) and interaction properties (e.g., human-likeness [43]) is also emphasized, built upon users' interaction needs and their profiles [44]. [35] has focused on the roles of the HRI (i.e., do-goals, be-goals), looking into the psychological need fulfillment, positive affect, and product perception of the robot interaction [45]. Reference [45] suggested pragmatic, hedonic-identity, hedonic-stimulation, and attractiveness as primary qualities of UX evaluation, while [35] suggested diverse qualities, such as relatedness, meaning, stimulation, competence, security, and popularity, as means to measure needs in various activities (e.g., watching, listening, playing). Attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty have been suggested as HRI UX evaluation measurements by [46], while users' reactions and feelings were focused by understanding perceived humanness, eeriness, and attractiveness for the robot acceptance in [47].

3) PERCEIVED SOCIAL PRESENCE

Social presence has been defined as the *sense of being in the company of another living being*, which has been widely investigated in social robot studies [48]. The social presence of HRI has often been understood as part of UX, as forms of dependability [46], perceived humanness [47], relatedness, security [46], and acceptance [43]. Perceived social presence is a significant aspect differentiating HRI from other systems with artificially embodied entities [48], including the HCI systems. As revealed in [5], learners perceive that they recognize, understand, and communicate better with the HRI system with the humanoid compared to the HCI system, leading to knowledge gain, even though both feedback conditions were the same other than the assistance of a robot. Through a meta-review, [49] has revealed that the in-person HRI poses positive effects on the combined outcomes, efficacy, perceptions, and attitudes toward systems, compared to the remote HRI, indicating the significant effects of the "physicality" of in-person HRI interfaces on the learner perceptions. In this regard, we understand the perceived social presence of physical humanoids as our focus in this section, separately from the perceived UX. Perceived presence is known to enhance learner participation, satisfaction, [50], cognition, and critical thinking [51]. Also, the sense of social presence is known to aid learners' physical, emotional, and cognitive health in remote education, which seems especially relevant to the recent online education in the post-pandemic era [48].

As means to evaluate the social presence of social robots, the following measures have been investigated: perceived robot appearances [52], rapport building and relationship dynamics [53], immersion, parasocial interaction, parasocial relationships, physiological responses, social reality, and general social richness [54], salience, perceived actor-hood, co-location/non-mediation, understanding, association, involvement, and medium sociability [55]. Not merely focusing on perception towards the interaction medium itself (i.e., robot), the perception toward the message has also empathized that are relevant to the overall conversational agents [55]: attentional allocation, perceived message understanding, perceived affective understanding, perceived emotional interdependence, and perceived behavioral interdependence [56]. Reference [57] suggested the different effects come from learner groups with varied consciousness, indicating that the higher social presence is associated with the perceived learning and satisfaction in learners with low consciousness. In contrast, the social presence did not affect the perceived learning or satisfaction in the highly-conscious learners. Studies have suggested enhancing the social presence of learners: Using scaffolded and self-reflective topics for better self-disclosure, [58], facilitating small group discussions [59], utilizing the storytelling [60], and providing personalized features in implementation, such as personal profiles, text messages, individualized video feedback, and one-on-one email communication [57].

4) ATTENTION REGULATION BEHAVIORS IN E-READING

Physical reading behaviors have been used as measurements to understand learners' engagement and visual attention in e-reading, having various sensors, such as eye tracker [61], [62], [63], motion sensors [64], [65], webcam [3], [5], [66], 3D-camera [67] and log data layers [68], implemented. However, webcam-based attention feature extraction has rarely been attempted, which could significantly assist the real-world feedback loop design without complicated sensor implementations in various learning scenarios. This work implements the webcam-based e-reading attention recognition framework of [3] for attention regulation behavior annotation. Reference [3] suggested attention regulation behavior as a critical cue where learners are aware of their attention loss and try to regain their focus in e-reading. The behaviors have been movements from eyebrow (e.g., eyebrow raise, bring together), blinks (e.g., blink flurries, prolonged voluntary blinks), mumble (e.g., mumble reading), hand (e.g., touch body, face), and body (e.g., adjust torso, arm, head), as opposed to neutral state without movements mentioned. Such behaviors have been revealed to correlate significantly

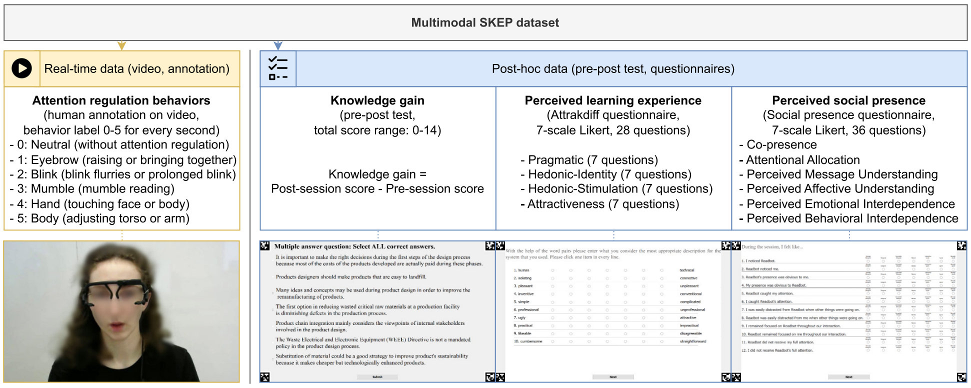


FIGURE 2. Multimodal SKEP dataset for attention regulation behaviors, knowledge gain, perceived learning experience, and perceived social presence in e-learning with a conversational agent.

with self-reported distractions of learners, indicating the behaviors as signs of attention loss and the following attention self-regulation. Video-based deep learning models have been implemented as a good predictor of self-reported distractions [3], knowledge gain, perceived learning experiences, and perceived social presence [5], respectively in the real-world setting [3], as well as in the laboratory-based-setting with the HRI system implemented [5].

III. DATA PRE-PROCESSING AND UNSUPERVISED CLUSTERING BASED ON FACTUAL LEARNING OUTCOMES

This section introduces the dataset and performs data pre-processing to construct features used for unsupervised learning. We represent feature constructions and further conduct feature engineering by comparing the silhouette scores of manually and automatically selected sets of features. We conduct unsupervised training based on various modeling methods (e.g., k-means, hierarchical, DBSCAN, spectral clustering) and validate clusters via the Chi-square test. In the process, we expect to tackle the following research question:

- RQ1. How can learner features best reflect learners' performances, experiences, and perceptions of conversational agents' interventions in e-reading?

A. MULTIMODAL SKEP DATASET

We utilize a multimodal SKEP Dataset (see Fig.2) collected from 60 higher-education learners who use the English language for their daily education [19]. Participants were recruited for an e-reading task on the screen (Age: M = 24.9, SD: 3.92; Gender: 37 males, 23 females). They were given an e-reading system with emphatic and metacognitive feedback from conversational agents through pop-ups and speech from a Furhat Robot [69], a conversational agent in a humanoid robot form.

Before the e-reading, participants were given a pre-test with 14 questions to measure their prior knowledge about the topic as a diagnostic knowledge measurement tool. The e-reading content has had seven subsections with 4,750 words concerning ''Waste management and critical raw materials''. In the process, learners' self-reports from the pre-post test (e.g., knowledge gain) and questionnaires (e.g., perceived learning experience and perceived social presence) were collected. At the end of every subsection of the screen-based e-reading material, pop-up questions were given as formative measurements. At the end of all subsections, another seven questions were given as a summative measurement tool. Additionally, two post-session questionnaires took place to understand learners' perceptions of the learning experiences and their perception of the system as a social presence, respectively: the Attrakdiff questionnaire with 28 questions, which have pragmatic, hedonic-identity, hedonic-stimulation, and attractiveness as its subdimensions, and the Social Presence questionnaire with 36 questions that concerns co-presence, attention allocation, perceived message understanding perceived affective understanding, perceived emotional interdependence, and personal behavioral interdependence.

Also, throughout the experiment, the video data were collected through a webcam, and multiple annotators later annotated learners' behaviors for attention regulation. The video data contains a total of 2,339 minutes, reaching 40 hours. The video samples were segmented every second, and 140,340 frames were annotated into five attention regulation behaviors (e.g., movements from eyebrow, blink, mumble, hand, body) and one neutral label that was further cross-validated. Note that learner data from GUI-based or HRI-based conversational agents from the SKEP dataset were not considered differently in this work. It is because our data-driven persona aims to see learners' perceptions and responses toward the feedback system regardless of the specific type of feedback.

B. MANUAL VS. AUTOMATIC FEATURE SELECTION

Feature vectors representing the best subset of variables are often scrutinized in two different ways: manually and automatically [70]. Manual feature selection is conducted based on a good understanding of the domain and dataset, often criticized for human bias and having deviated results from different evaluators. Automatic feature selection is especially beneficial in high-dimensional data where dimension reduction of data is essential and manual selection cannot achieve the utmost efficiency. However, automatic methods also have limitations, such as information loss and low interpretability in results. To achieve both semantically and scientifically sound results, we conducted the feature selection 1) by manually dividing categorical features into three semantic levels and 2) by conducting the automatic Principal Component Analysis (PCA) based on the percentage of consensus in generalized Procrustes analysis. We compared the silhouette scores of different sets of features, which offers the best distinctions among clusters. We found the best-performing method, which helped us find the optimal feature vectors with the best consistency of data clusters. The silhouette analysis and further applied elbow method are used to understand the number of optimal clusters for future unsupervised training. Note that mean-max normalization was applied to the SKEP dataset to make the subsets identical to avoid potential bias from the different data ranges.

It is of importance to state that the main reason that we adopted PCA as a representative automatic feature selection method has been because it is a linear *feature selection* technique, which provides a clearer understanding of how the original features contribute to the variance in the data, which offers better explainability. It has been required for our work since we wanted to investigate different features among clusters in the context of underlying structures and relationships of features. Though *feature extraction* methods, such as Uniform Manifold Approximation and Projection (UMAP) and t-SNE (t-distributed Stochastic Neighbor Embedding), are also commonly applied methods for the dimensionality reduction, those methods have not been used in our research since they do not retain the original features, thus not retrievable for further analysis nor interpretable, which is unsuitable for our further archetype extraction for persona development.

1) MANUAL FEATURE SELECTION

As can be seen from Table 1 and equation 1, 2, and 3, the SKEP dataset has data with three layers: 1) *low-dimensional features* with three components (e.g., knowledge gain, perceived learning experience, perceived social presence), 2) *mid-dimensional features* with 11 components (e.g., knowledge gain, pragmatic, hedonic-identity, hedonic-stimulation, attractiveness, co-presence, attentional allocation, perceived message understanding, perceived affective understanding, perceived emotional interdependence, perceived behavioral interdependence measures), and 3) *high-dimensional features* with 65 components (e.g., knowledge gain, seven sub-questions of pragmatic, hedonic-identity, hedonic-stimulation, attractiveness measures, six sub-questions of co-presence, attentional allocation, perceived message understanding, perceived affective understanding, perceived emotional interdependence, perceived behavioral interdependence measures). Those are three levels of features with various dimensionality that make semantic sense to most human evaluators based on the information hierarchy. Thus, we used those three levels of dimensional features as manually selected features, which are listed in Table 1.

$$KnowledgeGain = \sum_{i=1}^{N=7} Score_i^{PreSession} + \sum_{i=1}^{N=7} Score_i^{InSession} - \sum_{i=1}^{N=14} Score_i^{PostSession} \tag{1}$$

$$PerceivedLearningExperience = \frac{\sum_{i=1}^{N=7} Score_i^{PragmaticQuality} + \sum_{i=1}^{N=7} Score_i^{Hedonic-Identity} + \sum_{i=1}^{N=7} Score_i^{Hedonic-Simulation} + \sum_{i=1}^{N=7} Score_i^{Attractiveness}}{28} \tag{2}$$

$$PerceivedSocialPresence = \frac{\sum_{i=1}^{N=6} Score_i^{Co-presence} + \sum_{i=1}^{N=6} Score_i^{AttentionalAllocation} + \sum_{i=1}^{N=6} Score_i^{PerceivedMessageUnderstanding} + \sum_{i=1}^{N=6} Score_i^{PerceivedAffectiveUnderstanding} + \sum_{i=1}^{N=6} Score_i^{PerceivedEmotionalInterdependence} + \sum_{i=1}^{N=6} Score_i^{PerceivedBehavioralInterdependence}}{36} \tag{3}$$

2) AUTOMATIC FEATURE SELECTION BASED ON PCA

We conducted the PCA to achieve an automatic feature selection. PCA is often used for unsupervised learning to reduce the data complexity by reducing the noise and the dimensionality of data. By only selecting the principal components that explain the greatest amount of variance, the computation becomes lighter with better clarity in convoluted and multi-directional factors with minimal information loss. The equation below shows that the PCA produces a linear composition of the original components until the d dimensions, from the highest variance in the first element to the lowest variance in the last element. The newly created k is called the principal component, which decides the new dimension of subsets. Note that $k<d$.

$$PC_i = a_1X_1 + a_2X_2 + \cdots + a_dX_{d'}, \tag{4}$$

where X_j is the initial function a_j. j is the ith PC, while a_j is X_j number coefficient.

As [71] indicated, 70% of explained variance is common, while [72] applied a total variance ratio greater than 80% to reveal the most critical variables through the PCA. Below, we applied variously explained variances to find the number of features required to achieve specific proportions of explained variance. Note that 55%, 65%, 75%, 85%, and 95% have been applied as the proportions of explained variance (see Fig.3). Feature numbers derived from each proportion of explained variance have been applied for the silhouette analysis in the next section (see Table 2). Note that the number of components in Fig.3 is 60, equivalent to the sample number since *samplenumbers* < *featurenumbers* in this dataset.

TABLE 1. SKEP dataset with low-dimensional, mid-dimensional, and high-dimensional features.

Low-dimensional Features	Mid-dimensional Features	High-dimensional Features
Knowledge Gain	Knowledge Gain	Knowledge Gain
Perceived Learning Experience	-	-
	Pragmatic Quality	7 Sub-questions
	Hedonic-Identity	7 sub-questions
	Hedonic-Stimulation	7 sub-questions
	Attractiveness	7 sub-questions
Perceived Social Presence	-	-
	Co-presence	6 sub-questions
	Attentional Allocation	6 sub-questions
	Perceived Message Understanding	6 sub-questions
	Perceived Affective Understanding	6 sub-questions
	Perceived Emotional Interdependence	6 sub-questions
	Perceived Behavioral Interdependence	6 sub-questions

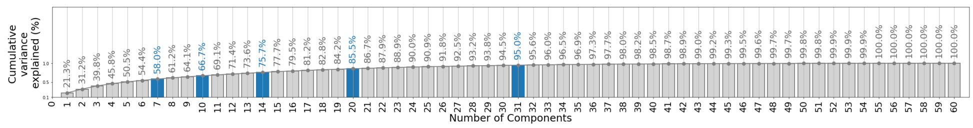


FIGURE 3. Number of principal components for explaining variance. 7, 10, 14, 20, and 31 components were required to explain the 55%, 65%, 75%, 85%, and 95% variances, respectively.

In such a case, the PCA automatically takes the sample numbers as the feature numbers.

3) FEATURE SELECTION METHODS COMPARISION: SILHOUETTE ANALYSIS

We have conducted silhouette analysis on manually selected features and automatically selected features to find the best-distinguished clusters from the feature selection. Note that the silhouette coefficient ranges between -1 and 1, and a score close to 1 indicates the best performance.

In this study, silhouette analysis has been applied for two purposes: 1) measuring the quality of the clusters based on different feature vectors as a part of the feature selection process and 2) getting the first indication of the optimal number of clusters. See Table 2 for the silhouette coefficients derived from manually and automatically selected sets of features. Various cluster numbers have been applied in the analysis process for further insights. The result shows the best silhouette score has been achieved when manually selected low-dimensional data has been applied, indicating the optimal number of clusters as 6. It was ascertained by analyzing the Within-Cluster Sum of Squares (WCSS), a method employed to find the point at which an additional increase in the number of clusters results in a diminishing return regarding the reduction of WCSS. Thus, this work uses knowledge gain, perceived learning experiences, and perceived social presence as three feature vectors for unsupervised model training.

We assume that the PCA did not improve the performance of silhouette analysis, seemingly because the PCA is based on the noise and the corresponding dimension reduction in the dataset. In the PCA process, some essential data structures or features might have been damaged, while all features were restructured as linear data and de-noised. In some cases, the neighboring clusters might have been too close when feature selection was made based on the PCA.

IV. UNSUPERVISED LEARNING FOR LEARNER PATTERN CLUSTERING AND COMPARATIVE ANALYSIS

In this section, we implement different unsupervised learning methods for further comparative analysis, suggested in previous review [9], but has yet to be attempted in data-driven persona development studies. We compare four unsupervised methods with various hyperparameters to evaluate the result consistency among methods as cross-validation. In the process, we tackle the following research question:

- RQ2. How can unsupervised learning methods be used for learner pattern clustering and validation?

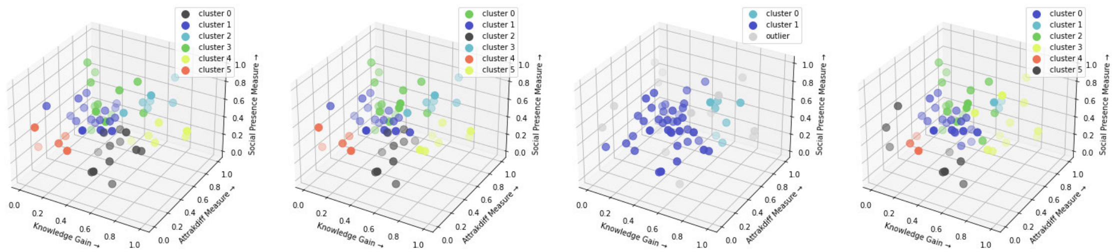
Specifically, we implemented k-means clustering, agglomerative hierarchical clustering, DBSCAN clustering, and spectral clustering methods that represent centroid, hierarchy, density, and graph-based methods, respectively (see Fig.4 for the 3-D visualization of the clusters). With different underlying principles for segmenting clusters, we hypothesized that they could cross-validate each others' robustness of cluster distinctions, which has also been suggested in previous research [25]. This has been done because unsupervised methods do not have ground truths and thus require further validation, while there are no standardized methods for such practices, as articulated in the subsection of "Related Work". Please see the following subsection for the rationale and implementation details that are applied in the cross-validation process.

A. CROSS-VALIDATING CLUSTERS FROM VARIOUS MODELING METHODS VIA CHI-SQUARE TEST

In this section, we apply the Chi-square test to validate the cluster distributions derived from different modeling methods. The chi-square test is a frequently applied method to determine the statistical differences and homogeneity in one or more categories of a contingency. We premised that the clusters are well-defined if homogeneity is found among the models. As can be seen from Table 3, significant p values are observed, indicating a significant relationship between

TABLE 2. Silhouette analysis conducted on manually selected features and automatically selected features as a part of feature selection.

		Manual Feature Selection			Automatic Feature Selection				
		Low-dimensional (3 Features)	Mid-dimensional (11 Features)	High-dimensional (65 Features)	PCA (31 Features)	PCA (20 Features)	PCA (14 Features)	PCA (10 Features)	PCA (7 Features)
Number of Clusters	2	0.679	**0.451**	**0.282**	**0.297**	**0.276**	**0.261**	**0.265**	**0.324**
	3	0.604	0.397	0.194	0.171	0.188	0.201	0.217	0.232
	4	0.641	0.406	0.174	0.114	0.126	0.118	0.129	0.159
	5	0.647	**0.451**	0.149	0.094	0.083	0.101	0.118	0.148
	6	**0.729**	0.367	0.143	0.077	0.076	0.103	0.131	0.125
	7	0.670	0.338	0.116	0.026	0.059	0.085	0.076	0.099
	8	0.531	0.250	0.115	0.020	0.042	0.090	0.070	0.097



(a) k-means. (b) Hierarchical. (c) DBSCAN. (d) Spectral.

FIGURE 4. Visualized results of different unsupervised learning methods: k-means, hierarchical, DBSCAN, spectral clustering methods.

TABLE 3. Chi squared test applied to unsupervised models with different parameters.

		k-means			Hierarchical												DBSCAN						Spectral		
					Single			Average			Complete			Ward			Min Outlier			Max eps					
		Value	df	p	Value	df	p	Value	df	p	Value	df	p	Value	df	p	Value	df	p	Value	df	p	Value	df	p
k-means		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hierarchical	Single	163	25	**<.001***	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Average	163	25	**<.001***	300	25	**<.001***	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Complete	163	25	**<.001***	300	25	**<.001***	300	25	**<.001***	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Ward	163	25	**<.001***	300	25	**<.001***	300	25	**<.001***	300	25	**<.001***	-	-	-	-	-	-	-	-	-	-	-	-
DBSCAN	Min Outlier	37.9	10	**<.001***	52.9	10	**<.001***	52.9	10	**<.001***	52.9	10	**<.001***	52.9	10	**<.001***	-	-	-	-	-	-	-	-	-
	Max eps	77.6	15	**<.001***	128	15	**<.001***	128	15	**<.001***	128	15	**<.001***	128	15	**<.001***	73.4	6	**<.001***	-	-	-	-	-	-
Spectral		163	25	**<.001***	300	25	**<.001***	300	25	**<.001***	300	25	**<.001***	300	25	**<.001***	52.9	10	**<.001***	128	15	**<.001***	-	-	-

The *single* linkage combines clusters with the nearest data pair. The *average* linkage combines clusters with the closest average distance between clusters. The *complete* linkage combines clusters, which have the maximum distance between any two points of different clusters. The *ward* linkage combines clusters with minimized within-cluster variances (=sum-of-squared distance) of all clusters. We apply the Eps of 0.205, where the result shows the least outlier, while the $K > 1$. We apply the Eps of as 0.19, where the Eps has the smallest value before the exponential increase in the number of clusters at 3.

categorical variables built upon different models. Thus, from the following section, we use the cluster distribution derived from the k-means clustering method, considering the model capability of being applied to larger datasets for future research reproduction with increased samples. Note that the Chi-square test has been conducted based on independent categorical cluster inputs (e.g., A, B, C, etc.).

V. DATA-DRIVEN PERSONA DEVELOPMENT AND STATISTICAL INTERPRETATION OF EACH CLUSTER

In this section, we conduct the statistical analysis on each cluster derived from the k-means clustering method. In this section, unique features from each cluster are derived based on quartile analysis to tackle the following research question:

- RQ3. How can we extract valuable archetypes of learners from different clusters and develop data-driven personas based on them?

A. ARCHETYPE EXTRACTION BASED ON QUARTILE ANALYSIS USING THE LOW-DIMENSIONAL DATA

As shown in Table 4, we conducted the statistical analysis, built upon clusters derived from the k-means clustering on factual learning outcomes. We first find the average feature of all learners from all clusters. We further conduct the quartile analysis on each cluster and see where each cluster is located from the whole set by comparing the *mean* value of each cluster, which represents the most typical learner in the cluster, and the *quartile* ranges from all learners.

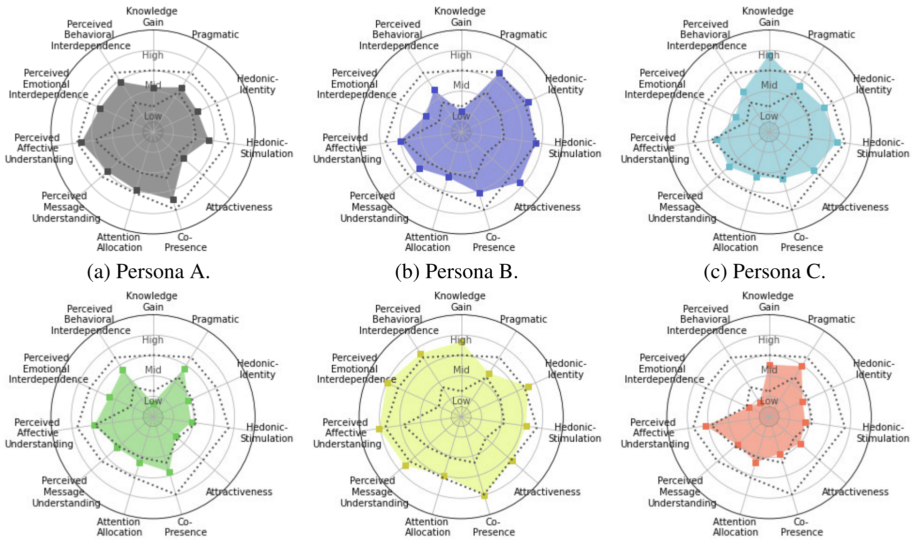
Quartile analysis provides statistically critical information about the center point and the spreads of the data [73]. It shows where a specific learner cluster is located from the overall learner data points. Based on the quartile, we interpreted the learners into three levels: if *mean of cluster<1st quartile (25%) of all learners*, we interpreted it as *low*, which means that learners in the cluster show less tendency of having the specific feature than the average learners. If *1st quartile (25%) of all learners<mean of cluster<2nd quartile (50%) of all learners*, we interpreted it as *Mid*, which means that learners in the cluster are located in the average range of the particular feature (e.g., knowledge gain, perceived learning experience, perceived social presence). If *2nd quartile (50%) of all learners<mean of cluster<3rd quartile (75%) of all learners*, we interpreted it as *high*, which means that learners in the cluster show the strong tendency of having the particular feature than the average learners.

B. ARCHETYPE EXTRACTION BASED ON QUARTILE ANALYSIS USING THE MID-DIMENSIONAL AND HIGH-DIMENSIONAL DATA: TOP-DOWN APPROACH

We applied the quartile analysis to the mid and high-dimensional data to understand learners based on more

TABLE 4. Statistical analysis conducted on clusters derived from k-means clustering result. Factual learner features, such as knowledge gain, perceived learning experience, and perceived social presence, have mainly been investigated through quartile analysis.

Statistical analysis (k-means)										
		Counts	M	SD	Min	25%	50%	75%	Max	Interpretation
Knowledge Gain	Overall	60	0.431	0.249	0.000	**0.250**	**0.416**	**0.604**	1.000	
	Persona A	19	**0.438**	0.123	0.250	0.375	0.416	0.500	0.666	**Mid**
	Persona B	12	**0.201**	0.164	0.000	0.083	0.116	0.333	0.500	**Low**
	Persona C	8	**0.760**	0.150	0.500	0.729	0.750	0.791	1.000	**High**
	Persona D	8	**0.135**	0.098	0.000	0.062	0.166	0.187	0.250	**Low**
	Persona E	6	**0.736**	0.081	0.666	0.666	0.708	0.812	0.833	**High**
	Persona F	7	**0.511**	0.089	0.333	0.500	0.500	0.583	0.583	**Mid**
Perceived Learning Experience	Overall	60	0.491	0.244	0.000	**0.310**	**0.444**	**0.652**	1.000	
	Persona A	19	**0.386**	0.120	0.129	0.296	0.388	0.444	0.611	**Mid**
	Persona B	12	**0.819**	0.128	0.592	0.787	0.824	0.898	1.000	**High**
	Persona C	8	**0.581**	0.151	0.314	0.541	0.629	0.657	0.759	**Mid**
	Persona D	8	**0.236**	0.007	0.148	0.185	0.194	0.300	0.370	**Low**
	Persona E	6	**0.635**	0.171	0.370	0.574	0.648	0.694	0.888	**Mid**
	Persona F	7	**0.280**	0.189	0.000	0.166	0.259	0.425	0.518	**Low**
Perceived Social Presence	Overall	60	0.482	0.206	0.000	**0.327**	**0.523**	**0.606**	1.000	
	Persona A	19	**0.590**	0.106	0.396	0.551	0.584	0.617	0.811	**Mid**
	Persona B	12	**0.438**	0.208	0.047	0.341	0.433	0.603	0.745	**Mid**
	Persona C	8	**0.358**	0.114	0.160	0.308	0.334	0.433	0.518	**Mid**
	Persona D	8	**0.435**	0.162	0.235	0.320	0.415	0.530	0.726	**Mid**
	Persona E	6	**0.768**	0.132	0.632	0.693	0.726	0.816	1.000	**High**
	Persona F	7	**0.215**	0.122	0.000	0.155	0.264	0.301	0.330	**Low**



(a) Persona A. (b) Persona B. (c) Persona C.

(d) Persona D. (e) Persona E. (f) Persona F.

FIGURE 5. Archetype extraction and visualization based on the mid-dimensional data.

detailed artifacts. While the quartile analysis on the low-dimensional data provides a general understanding of learner clusters, the top-down approach based on the mid and high-dimensional data lets a vivid understanding of learners based on more detailed features. See Fig.5 for the visualized archetype based on the mid-dimensional data. See Table 5 for the detailed archetype descriptions based on the high-dimensional data.

C. DATA-DRIVEN PERSONAS BUILT UPON ARCHETYPES OF DIFFERENT CLUSTERS

1) PERSONA A: ARCHETYPES DERIVED FROM CLUSTER 0

Persona A has been the most common type among all (60 participants), having 19 participants (31.67%) in the same cluster. *Persona A* has shown no significant knowledge gain and perceived learning experiences. In the perceived social presence measure, *persona A* did not show significant variances from the average learners, aside from one sub-measure from perceived message understanding, that "it was easy to understand Readbot (i.e., feedback system with conversational agents)". All in all, *persona A* is the most average type of learner among all participants.

2) PERSONA B: ARCHETYPES DERIVED FROM CLUSTER 1

Persona B has been the second most common type of learner group among all participants, having 12 learners (20.0%) in the same segment. *Persona B* has achieved the second lowest knowledge gain compared to other groups. However, *Persona B* has evaluated the system as partially pragmatic and most attractive among all groups. The feedback from conversational agents has been evaluated as "human", "pleasant", "likable", "appealing", and "motivating" by *persona B*. *Persona B* did not show any significant perceived social presence. *Persona B* is the learner type that perceives the system positively and has a good learning experience. However, it did not lead to good knowledge gain, which is

TABLE 5. Archetype extraction and quartile analysis based on the high-dimensional data.

			Persona A		Persona B		Persona C		Persona D		Persona E		Persona F	
Low-dimensional	Mid-dimensional	High-dimensional	High (Q3)	Low (Q1)	High (Q3)	Low (Q1)	High (Q3)	Low (Q1)	High (Q3)	Low (Q1)	High (Q3)	Low (Q1)	High (Q3)	Low (Q1)
Knowledge Gain			·	·	·	✓	✓	·	·	✓	✓	·	·	·
Perceived Learning Experience	Pragmatic	Q1. Technical (Low)-Human (High)	·	·	✓	·	·	·	·	·	·	·	·	·
		Q8. Impractical (Low)-Practical (High)	·	·	·	·	·	·	·	·	·	✓	·	·
		Q12. Unpredictable (Low)-Predictable (High)	·	·	·	·	·	·	·	·	·	✓	·	·
	Hedonic-Identity	Q13. Cheap (Low)-Premium (High)	·	·	·	·	·	·	·	·	·	·	·	✓
		Q15. Separates me from people (Low)-Brings me closer to people (High)	·	·	·	·	·	·	·	·	·	·	✓	·
	Hedonic-Stimulation	Q4. Conventional (Low)-Inventive (High)	·	·	·	·	·	·	·	✓	·	·	·	·
		Q24. Dull (Low)-Captivating (High)	·	·	·	·	·	·	·	·	·	·	·	✓
		Q27. Ordinary (Low)-Novel (High)	·	·	·	·	·	·	·	·	·	·	·	✓
	Attractiveness	Q3. Unpleasant (Low)-Pleasant (High)	·	·	✓	·	·	·	·	·	·	·	·	·
		Q9. Disagreeable (Low)-Likable (High)	·	·	✓	·	·	·	·	✓	·	·	·	·
		Q21. Repelling (Low)-Appealing (High)	·	·	✓	·	✓	·	·	✓	✓	·	·	·
		Q26. Discouraging (Low)-Motivating (High)	·	·	✓	·	·	·	·	✓	·	·	·	✓
Perceived Social Presence	Co-presence	Q1. I noticed Readbot.	·	·	·	·	·	·	·	·	·	·	·	✓
		Q2. Readbot noticed me.	·	·	·	·	·	✓	·	·	✓	·	·	✓
		Q3. Readbot's presence was obvious to me.	·	·	·	·	·	·	·	·	·	·	·	✓
		Q4. My presence was obvious to Readbot.	·	·	·	·	·	·	·	·	·	·	·	✓
		Q5. Readbot caught my attention.	·	·	·	·	·	✓	·	·	✓	·	·	✓
		Q6. I caught Readbot's attention.	·	·	·	·	·	·	·	·	·	·	·	✓
	Attention Allocation	Q10. Readbot remained focused on me throughout our interaction.	·	·	·	·	·	·	·	·	·	·	·	✓
	Perceived Message Understanding	Q14. Readbot was clear to me.	·	·	·	·	·	·	·	·	✓	·	·	·
		Q15. It was easy to understand Readbot.	·	✓	·	·	·	·	·	·	·	·	·	✓
		Q17. Understanding Readbot was difficult.	·	·	·	·	·	·	·	✓	·	·	·	·
	Perceived Affective Understanding	Q19. I could tell how Readbot felt.	·	·	·	·	·	·	·	·	✓	·	·	·
		Q20. Readbot could tell how I felt.	·	·	·	·	·	·	·	✓	·	·	·	·
		Q22. My emotions were not clear to Readbot.	·	·	·	·	·	·	·	·	·	✓	·	·
		Q23. I could describe Readbot's feelings accurately.	·	·	·	·	·	✓	✓	·	·	·	·	·
		Q24. Readbot could describe my feelings accurately.	·	·	·	·	·	·	·	·	✓	·	·	·
	Perceived Emotional Interdependence	Q25. I was sometimes influenced by Readbot's moods.	·	·	·	·	·	✓	·	·	✓	·	·	·
		Q26. Readbot was sometimes influenced by my moods.	·	·	·	·	·	·	·	·	✓	·	·	✓
		Q27. Readbot's feelings influenced the mood of our interaction.	·	·	·	·	·	✓	·	·	✓	·	·	✓
		Q29. Readbot's attitudes influenced how I felt.	·	·	·	·	·	·	·	·	✓	·	·	✓
		Q30. My attitudes influenced how Readbot felt.	·	·	·	·	·	·	·	·	·	·	·	✓
	Perceived Behavioral Interdependence	Q31. My behavior was often in direct response to Readbot's behavior.	·	·	·	·	·	·	·	·	·	·	·	✓
		Q32. The behavior of Readbot was often in direct response to my behavior.	·	·	·	·	·	·	·	·	·	·	·	✓
		Q33. I gave and took Readbot's actions mutually.	·	·	·	·	·	·	·	·	·	·	·	✓
		Q34. Readbot's gave and took my actions mutually.	·	·	·	·	·	·	·	·	✓	·	·	✓
		Q35. Readbot's behavior was closely tied to my behavior.	·	·	·	·	·	·	·	·	✓	·	·	✓
		Q36. My behavior was closely tied to Readbot's behavior.	·	·	·	·	·	·	·	·	✓	·	·	✓

against of notion that the quality of the learning experience somewhat leads to positive learning outcomes.

3) PERSONA C: ARCHETYPES DERIVED FROM CLUSTER 2

Persona C was derived from eight learners (13.33%). *Persona C* has shown high knowledge gain among all participants and found the system "appealing" in the attractiveness of learning experience evaluation. However, *persona C* has responded generally negatively to the social presence measures, especially in co-presence, perceived emotional interdependence, and perceived affective understanding. *Persona C* has evaluated that "Readbot did not notice me." and "Readbot did not catch my attention.", showing low sense of co-presence. Furthermore, regarding perceived emotional interdependence, *persona C* answered that "I could not describe Readbot's feeling accurately.". *Persona C* has responded that "I was not influenced by Readbot's moods." and "Readbot's mood did not influence the mood of our interaction.", showing the lower perceived emotional interdependence in two sub-measures. All in all, *persona C* is the learner type that performs highly in knowledge gain, regardless of mediocre learning experience and mediocre to low perceived social presence of the system. Person C is a learner type that has trouble relating to conversational agents due to his or her low co-presence with the system. However, the knowledge gain has been achieved highest among all learners groups.

4) PERSONA D: ARCHETYPES DERIVED FROM CLUSTER 3

Persona D has derived from eight learners (13.33%). *Persona D* has achieved the lowest knowledge gain among all participant groups. Also, *persona D* has evaluated the perceived learning experience among all participant groups, especially in attractiveness and pragmatic value of the system, perceiving the system as "disagreeable", "repelling", "discouraging", and "conventional", respectively. In perceived social presence, *persona D* has provided answers within the mid-range. However, in some perceived affective understanding sub-measures, indicating that "Understanding Readbot was difficult." and "Readbot could not tell how I felt.", while perceiving that, " I could describe Readbot's feelings accurately.". Overall, *persona D* is regarded as the learner type who performs poorly in knowledge gain based on a poor perceived learning experience with the system. *Persona D* seemed discouraged and repelled by the system that did not understand him or herself, likely in awareness that the feedback was not based on their responses (i.e., intelligent system), having no difference from the conventional one-way feedback system. In that regard, It seems that a better interaction design based on an intelligent system might bring a better-perceived learning experience and subsequent improvements in knowledge gain for *persona D*.

5) PERSONA E: ARCHETYPES DERIVED FROM CLUSTER 4

Persona E has been built based on data from six learners (10.0%). *Persona E* has recorded the second-highest knowledge gain among all learner groups. *Persona E*'s evaluation of his or her learning experience was average. However, *Persona E* has evaluated the pragmatic value of the conversational agents poorly, perceiving the feedback as 'impractical'' and "unpredictable". However, *persona E* evaluated the system as "appealing". The most distinctive feature of *persona E* came from its generally high perceived social presence, which has not been found in other groups. The tendency has shown more obvious in assessing perceived emotional interdependence, reporting their perceptions as "I was sometimes influenced by Readbot's mood.", "Readbot was sometimes influenced by my mood.", "Readbot's feelings influenced the mood of our interaction.", and "Readbot's attitudes influenced how I felt.". Accordingly, *persona E* has shown high perceived behavioral interdependence, perceiving that "Readbot's gave and took my actions mutually.", "Readbot's behavior was closely tied to my behavior.", and "My behavior was closely tied to Readbot's behavior.". Moreover, in the co-presence sub-measures, *persona E* responded that "Readbot noticed me." and "Readbot caught my attention.". *Persona E* has also reported that "I could tell how Readbot tells.", "Readbot could describe my feelings accurately.", and "Readbot was clear to me.", showing high perceived message understanding and per affective understanding compared to other groups of participants.

6) PERSONA F: ARCHETYPES DERIVED FROM CLUSTER 5

Persona F has been developed based on seven learners (11.67%). *Persona F* did now show any significant knowledge gain compared to other groups of participants. The general perceived learning experience and social presence have been the lowest. *Persona F* has evaluated the system as "cheap", "dull", and "ordinary", in Hedonic-Identity and Hedonic-Stimulation measures. In the attractiveness sub-measures, *Persona F* found the system "discouraging".

In terms of perceived social presence, *Persona F*'s responses toward co-presence and perceived behavioral interdependence were all negative, indicating that "I did not notice Readbot.", "Readbot did not notice me.", "Readbot's presence was not obvious to me.", "My presence was not obvious to Readbot.", "Readbot did not catch my attention.", "I did not catch Readbot's attention.", and "My behavior was not in direct response to Readbot's behavior.", "The behavior of Readbot was not in direct response to my behavior.", "I did not give and take Readbot's actions mutually.", "Readbot's did not give and take my actions mutually.", "Readbot's behavior was not closely tied to my behavior.", and "My behavior was not closely tied to Readbot's behavior.", respectively. Also, *persona F*'s perceived emotional interdependence was also low, responding that "Readbot was not influenced by my mood.", "Readbot's feelings did not influence the mood of our interaction.", "Readbot's feelings did not influence the mood of our interaction.", "Readbot's attitudes did not influence how I felt.", and "My attitudes did not influence how Readbot felt.". Low attention allocation and perceived message understanding sub-measures from *persona F* have shown that "Readbot did not remain focused on me throughout our interaction." and "It was not easy to understand Readbot.".

Overall, *Persona F* did not consider conversational agents as beings with identity or being good hedonic stimuli to e-reading. At the same time, poorly perceived co-presence seemed to lead to *persona F*'s low perceived emotional interdependence and behavioral interdependence, subsequently. Interestingly, low perceived learning experience and social presence did not negatively impact the knowledge gain of *persona F*. However, it also indicates room for improvement in knowledge gain if guaranteed a better-perceived learning experience and social presence of conversational agents.

VI. AUTOMATIC PERSONA PREDICTIONS BASED ON ATTENTION REGULATION BEHAVIORS

Attention regulation behaviors are proven to be a robust predictor of learners' attention in e-reading [3]. In this section, we study if persona prediction can also be achieved using attention regulation behaviors. We implement multiple classification models to classify different patterns of personas via the attention regulation behaviors of learners. We utilized the cross-subject evaluation protocol in all classification tasks. We chose the classical 70-30 protocol of dividing 60 samples into 40 for training and 20 for testing. We also scrutinize which part of the video sample can provide the best clues for persona prediction by introducing comparative learning phase-based and time-based prediction approaches. Also, we compared two different sampling methods of instant and cumulative learner behavior labels. In the process, the research question below is answered:

- RQ4. Can we predict learner personas based on attention regulation behaviors?

A. LEARNING PHASED-BASED AND TIME DURATION-BASED PERSONA PREDICTION

This section implements four classical machine learning classifiers: AdaBoost, Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Random Forest. We predict learner personas based on their behaviors shown during 1) various phases of learning and time points based on 2) instant and cumulative behavioral data points. 3) we train on instant and cumulative data so that our work can contribute to the real-time feedback loop by investigating behavioral clues for predicting various personas. Note that such attempts have been derived from common technical challenges associated with real-time recognition: real-time recognition of learner states and closing the feedback loop.

B. SIX-CLASS CLUSTER PREDICTION (MULTICLASS CLASSIFICATION TASK)

Six-class persona prediction via attention regulation behavior has been conducted to differentiate all six personas (A-F) derived in the previous data-driven persona development section. As seen from Table 6, The best performances have been 45% of accuracy, using SVM and kNN applied to cumulative behaviors shown in various learning phases; the same performance has been achieved in the kNN and Random Forest, using the time duration-based method in 25%∼50% of reading duration. It is a significantly higher

TABLE 6. Six-class persona prediction based on the learning phased-based & time duration-based learner behavior data.

		Learning phase-based							Time duration-based			
Instant		Subtopic 1	Subtopic 2	Subtopic 3	Subtopic 4	Subtopic 5	Subtopic 6	Subtopic 7	∼25%	25%-50%	50%-70%	75%∼
	Random Guess	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	AdaBoost	0.15	0.30	0.25	0.20	**0.35**	0.20	0.30	0.30	0.30	0.30	**0.35**
	SVM	0.15	0.25	0.25	0.15	**0.40**	0.20	0.25	0.30	**0.40**	0.20	0.35
	kNN	0.30	0.25	0.30	0.20	**0.40**	0.20	0.30	0.30	0.35	0.30	0.30
	Random Forest	0.35	0.25	0.30	0.20	**0.40**	0.20	0.30	0.35	0.35	0.20	0.30
Cumulative		Subtopic 1	Subtopic 1-2	Subtopic 1-3	Subtopic 1-4	Subtopic 1-5	Subtopic 1-6	Subtopic 1-7	∼25%	∼50%	∼75%	∼100%
	Random Guess	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	AdaBoost	0.30	0.20	0.20	0.30	**0.40**	**0.40**	0.15	0.30	0.30	0.30	0.25
	SVM	0.15	0.20	0.30	0.30	0.25	**0.45**	0.10	0.35	0.40	0.20	0.35
	kNN	0.15	0.30	0.10	0.30	0.25	**0.45**	0.10	0.30	**0.45**	0.25	0.30
	Random Forest	0.15	0.25	0.15	0.30	0.30	0.30	0.10	0.30	**0.45**	0.30	0.35

The best and the second best performances are **bolded** and underlined, respectively.

performance than the random guess of 17%. There has been a general tendency that behavior from instant behavior data from subtopic 5 data has achieved better accuracy than the other part of instant data. Likewise, cumulative behavior data shown throughout subtopics 1-6 has derived the best result, with 45% as the best accuracy.

C. FOUR-CLASS PERSONA PREDICTION (MULTICLASS CLASSIFICATION TASK)

We further conducted the four-class persona prediction (see Table 7). We selected three personas that we found to have the feedback necessity among six personas: two personas with low knowledge gain with a low and high perceived learning experience, respectively (i.e., Persona B, Persona D), which indicates low learning performances. Another cluster was with the low learning experiences and social presence (Persona F), which suggests a potential to improve system perceptions and the following knowledge gain improvements with the future feedback loop implementation. We made the task to classify those three personas from others (Persona A, Persona C, Persona E), which have been combined as one label in the training process.

A four-class persona prediction is an economical approach to classify learners with learning needs, compared to the six-level persona prediction for all learner personas. As seen from Table 7, the time duration-based model has achieved the best accuracy both with learners' instant and cumulative behavioral data points. The result shows the highest classification accuracy of 70% via the SVM, kNN, and Random Forest classifiers with the instant behavior data and the SVM with the cumulative behavior data. It is a considerable improvement from the 6-class persona prediction of 45% as the best prediction result and observers' random guesses, which has an accuracy of 25%. The SVM classifier has shown relatively stable and robust performances in both instant and cumulative data in the time duration-based method, proving the most appropriate classifier for real-time feedback loop development. Model training on instant behavior data has shown generally higher accuracy than training on cumulative data.

Once the model is implemented as part of the real-time feedback loop, the time duration-based model using the SVM model based on both instant and cumulative video samples can work as a stable and robust method among all combinations from attempted cases, achieving the lowest 65% and the highest 70% accuracy.

All in all, our experimental result with various machine learning models and diverse sampling methods tackled the practical challenges of model implementations. The result represents that our real-time persona recognition, predicated on learner behaviors, is a valid approach within e-reading environments that can further be facilitated by real-time feedback with conversational agents. The major benefits of automatic persona analysis and recognition are two-fold.

1) More efficient and effective learning analytics and feedback implementation is possible based on the automatic learning analytics and persona recognition enabled by the machine, combined with human judgments and further intervention design for e-reading.

2) In real-time intervention design, the content and timing vary depending on learners' learning styles, perceptions, and interaction with the system. In this context, our automatic persona recognition provides more timely and customized feedback to diverse learners with different learning needs, greatly benefiting educational research and practices.

VII. LIMITATIONS AND FUTURE WORK

A. FEATURE ENGINEERING STILL REQUIRES HIGH-LEVEL HUMAN JUDGMENTS

As revealed in the feature selection process, evaluating multi-dimensional learner features requires a deep understanding of the domain and the specific dataset. Especially learning analytics and cluster classification greatly depend on how feature vectors and structures are designed. Therefore, feature engineering for different learning domains and tasks in future employment requires expertise with a deep understanding of the field and the data. It emphasizes the importance of more iterative and context-based data collection and learning analytics in a loop. Please be informed that our work is aimed at developing a supplementary tool to help human decisions in the design thinking process for educational support and intervention design, not replacing human decisions. We believe such a complementary relationship between the roles of humans and machines in the iterative decision-making process ensures more efficient and effective decision-making and better applicability of our framework to various context and domain-specific learning scenarios.

TABLE 7. Four-class persona prediction based on the learning phased-based & time duration-based learner behavior data.

		Learning phase-based							Time duration-based			
Instant		Subtopic 1	Subtopic 2	Subtopic 3	Subtopic 4	Subtopic 5	Subtopic 6	Subtopic 7	∼25%	25%-50%	50%-70%	75%∼
	Random Guess	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	AdaBoost	0.20	0.20	0.30	0.30	0.30	0.15	0.20	0.50	0.55	**0.60**	0.55
	SVM	0.35	0.40	0.35	0.35	0.35	0.40	0.40	**0.70**	0.65	**0.70**	**0.70**
	kNN	0.35	0.25	0.35	0.45	0.45	0.60	0.25	0.65	**0.70**	0.55	0.65
	Random Forest	0.20	0.25	0.35	0.35	0.4	0.35	0.20	**0.60**	0.45	**0.60**	**0.60**
Cumulative		Subtopic 1	Subtopic 1-2	Subtopic 1-3	Subtopic 1-4	Subtopic 1-5	Subtopic 1-6	Subtopic 1-7	∼25%	∼50%	∼75%	∼100%
	Random Guess	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	AdaBoost	0.20	0.40	0.10	0.40	0.10	0.20	0.35	0.50	**0.55**	0.50	**0.55**
	SVM	0.35	0.40	0.40	0.40	0.40	0.40	0.40	**0.70**	0.65	**0.70**	**0.70**
	kNN	0.35	0.40	0.30	0.35	0.30	0.35	0.35	**0.65**	0.60	**0.65**	**0.65**
	Random Forest	0.30	0.35	0.25	0.25	0.30	0.30	0.30	0.60	**0.65**	0.55	0.55

The best and the second best performances are **bolded** and underlined, respectively.

B. COMBINING EXPERT ANNOTATION AND *k*-MEANS CLUSTERING MIGHT PROVIDE MORE VALUABLE INSIGHTS

The k-means clustering method first chooses a random point and forms a cluster from that point until the last sample. Thus, the quality of the randomly-chosen first data point affects the clustering result, which might affect subsequent statistical analysis results. To overcome such methodological limitations, we suggest involving experts in deciding the centroids of each cluster for k-means clustering. By specifying the centroids rather than starting from random data points, the model can significantly reduce the possibility of selecting an outlier as the first centroid point and having misleading clusters that do not appropriately represent the learner groups.

C. FEEDBACK IMPLEMENTATION FOR DIFFERENT CLUSTER NEEDS REMAINS A CHALLENGE

We aimed at the data-driven persona development to build a foundation for a feedback loop in e-reading. Though we built up an architecture for automatic cluster generation, analysis, and persona prediction based on learners' behavior labels, we still need to implement specific interventions for personas at needs and close the feedback loop. Thus, intervention design and implementation in e-reading is our next research focus for the multimodal feedback "loop" design in e-reading.

VIII. CONCLUSION

In this work, we implemented a framework of data-driven persona to a multimodal SKEP dataset, which contains various data layers that reflect learners' attention and perception of their e-reading with feedback from conversational agents. We clustered learners based on their knowledge gain, perceived learning experience, and social presence using various unsupervised learning methods to find the feedback necessities of different learner segments. The Chi-square test has compared and validated machine-generated personas from different modeling methods. In the process, feature selection methods (e.g., manual, automatic) and different hyperparameters have been compared. We conducted the statistical quartile analysis on each cluster based on clusters derived from the k-means clustering method. We extracted each cluster's archetypes that make the cluster distinctive from each other and defined six personas. Furthermore, learners' different attention regulation behaviors were used to predict learner personas. In the process, diverse data points, such as instant and cumulative learner behavior labels, have been explored as one dimension while having the learning phase and time duration as another. Various classical classification models, such as AdaBoost, SVM, kNN, and Random Forest, have been applied to perform the 6-level and 4-level classification tasks. The result indicates that 4-level classification for finding personas with feedback needs, achieving 65-70% accuracy based on the SVM classifier on the time duration sampling method, showing the potential for the real-time feedback loop design. Overall, we aimed to build the architecture for further feedback prompts in e-reading. Our automatic data-driven persona development and prediction can contribute as a practical and effective learning analytics tool for real-time intervention design, greatly assisting researchers and instructional designers in the field.

REFERENCES

[1] P. T. A. Junior and L. V. L. Filgueiras, "User modeling with personas," in *Proc. Latin Amer. Conf. Hum.-Comput. Interact.*, 2005, pp. 277–282.

[2] M. Aoyama, "Persona-scenario-goal methodology for user-centered requirements engineering," in *Proc. 15th IEEE Int. Requirements Eng. Conf. (RE)*, Oct. 2007, pp. 185–194.

[3] Y. Lee, H. Chen, G. Zhao, and M. Specht, "WEDAR: Webcam-based attention analysis via attention regulator behavior recognition with a novel E-reading dataset," in *Proc. Int. Conf. MULTIMODAL Interact.*, Nov. 2022, pp. 319–328.

[4] A. M. Maatuk, E. K. Elberkawi, S. Aljawarneh, H. Rashaideh, and H. Alharbi, "The COVID-19 pandemic and E-learning: Challenges and opportunities from the perspective of students and instructors," *J. Comput. Higher Educ.*, vol. 34, no. 1, pp. 21–38, Apr. 2022.

[5] Y. Lee and M. Specht, "Can we empower attentive E-reading with a social robot? An introductory study with a novel multimodal dataset and deep learning approaches," in *Proc. LAK23: 13th Int. Learn. Analytics Knowl. Conf.* New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 1–16.

[6] J. Nasir, U. Norman, W. Johal, J. K. Olsen, S. Shahmoradi, and P. Dillenbourg, "Robot analytics: What do human–robot interaction traces tell us about learning?" in *Proc. 28th IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, 2019, pp. 1–7.

[7] Y. Lee, H. Chen, E. Tan, S. Praharaj, and M. Specht, "FLOWer: FeedBack loop for group work supporter," in *Proc. Int. Learn. Analytics Knowl. Conf. (LAK Demo Session)*, 2020, p. 1.

[8] S. G. Ghalebandi and A. Noorhidawati, "Engaging children with pleasure reading: The E-reading experience," *J. Educ. Comput. Res.*, vol. 56, no. 8, pp. 1213–1237, Jan. 2019.

[9] J. Salminen, K. Guan, S.-G. Jung, and B. J. Jansen, "A survey of 15 years of data-driven persona development," *Int. J. Human–Computer Interact.*, vol. 37, no. 18, pp. 1685–1708, Nov. 2021.

[10] A. Cooper, "The inmates are running the asylum," in *Software-Ergonomie'99*. Cham, Switzerland: Springer, 1999, p. 17.
[11] D. Chasanidou, A. A. Gasparini, and E. Lee, "Design thinking methods and tools for innovation," in *Design, User Experience, and Usability: Design Discourse*. Los Angeles, CA, USA: Springer, 2015, pp. 12–23.
[12] J. Salminen, K. Guan, S.-G. Jung, S. A. Chowdhury, and B. J. Jansen, "A literature review of quantitative persona creation," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–14.
[13] J. Brickey, S. Walczak, and T. Burgess, "A comparative analysis of persona clustering methods," in *Proc. 16th Americas Conf. Inf. Syst.*, 2010, pp. 1–9.
[14] X. Zhang, H.-F. Brown, and A. Shankar, "Data-driven personas: Constructing archetypal users with clickstreams and user telemetry," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2016, pp. 5350–5359.
[15] M. Mesgari, C. Okoli, and A. O. de Guinea, "Affordance-based user personas: A mixed-method approach to persona development," in *Proc. 21st Americas Conf. Inf. Syst.*, 2015, pp. 1–5.
[16] J. E. Nieters, S. Ivaturi, and I. Ahmed, "Making personas memorable," in *Proc. CHI Extended Abstr. Human Factors Comput. Syst.*, 2007, pp. 1817–1824.
[17] J. Pruitt and J. Grudin, "Personas: Practice and theory," in *Proc. Conf. Designing User Experiences*, Jun. 2003, pp. 1–15.
[18] J. McGinn and N. Kotamraju, "Data-driven persona development," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2008, pp. 1521–1524.
[19] Y. Lee and M. Specht. (2023). *Multimodal SKEP dataset for Attention Regulation Behaviors, Knowledge Gain, Perceived Learning Experience, and Perceived Social Presence in e-Learning With a Conversational Agent*. [Online]. Available: https://data.4tu.nl/datasets/4c9de645-ca88-4b45-8fc7-2fc325f191dc/1
[20] B. Namatherdhala, N. Mazher, and G. K. Sriram, "A comprehensive overview of artificial intelligence tends in education," *Int. Res. J. Modernization Eng. Technol. Sci.*, vol. 4, no. 7, pp. 1–4, 2022.
[21] S. Liu and M. d'Aquin, "Unsupervised learning for understanding student achievement in a distance learning setting," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2017, pp. 1373–1377.
[22] N. Zhang, G. Biswas, and Y. Dong, "Characterizing students' learning behaviors using unsupervised learning methods," in *Proc. Int. Conf. Artif. Intell. Educ.* Cham, Switzerland: Springer, 2017, pp. 430–441.
[23] K. Huang, T. Bryant, and B. Schneider, "Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data," in *Proc. Int. Conf. Educ. Data Mining*, 2019, pp. 1–6.
[24] A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth, "Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach," in *Proc. 5th Int. Conf. Learn. Analytics Knowl.*, Mar. 2015, pp. 146–150.
[25] O. R. Battaglia, B. D. Paola, and C. Fazio, "A new approach to investigate students' behavior by using cluster analysis as an unsupervised methodology in the field of education," *Appl. Math.*, vol. 7, no. 15, pp. 1649–1673, 2016.
[26] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Student engagement level in an E-learning environment: Clustering using K-means," *Amer. J. Distance Educ.*, vol. 34, no. 2, pp. 137–156, Apr. 2020.
[27] U. G. Inyang, U. A. Umoh, I. C. Nnaemeka, and S. A. Robinson, "Unsupervised characterization and visualization of students' academic performance features," *Comput. Inf. Sci.*, vol. 12, no. 2, pp. 103–116, 2019.
[28] S. Mulder and Z. Yaar, *The User is Always Right: A Practical Guide to Creating and Using Personas for the Web*. San Francisco, CA, USA: New Riders, 2006.
[29] T. Miaskiewicz, T. Sumner, and K. A. Kozar, "A latent semantic analysis methodology for the identification and creation of personas," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2008, pp. 1501–1510.
[30] P. De Bruyckere, P. A. Kirschner, and C. D. Hulshof, *Urban Myths About Learning and Education*. New York, NY, USA: Academic Press, 2015.
[31] P. A. Kirschner and J. J. G. van Merriënboer, "Do learners really know best? Urban legends in education," *Educ. Psychologist*, vol. 48, no. 3, pp. 169–183, 2013.
[32] M. V. J. Veenman, F. J. Prins, and J. Verheij, "Learning styles: Self-reports versus thinking-aloud measures," *Brit. J. Educ. Psychol.*, vol. 73, no. 3, pp. 357–372, Sep. 2003.
[33] F. Coffield, D. Moseley, E. Hall, K. Ecclestone, F. Coffield, D. Moseley, E. Hall, and K. Ecclestone, "Learning styles and pedagogy in post-16 learning: A systematic and critical review," Learn. Skills Res. Centre, London, U.K., Tech. Rep., 2004.
[34] B. Rienties, A. Boroowa, S. Cross, C. Kubiak, K. Mayles, and S. Murphy, "Analytics4action evaluation framework: A review of evidence-based learning analytics interventions at the open university UK," *J. Interact. Media Educ.*, vol. 2016, no. 1, pp. 1–11, 2016.
[35] E. Prati, M. Peruzzini, M. Pellicciari, and R. Raffaeli, "How to include user eXperience in the design of human–robot interaction," *Robot. Comput.-Integr. Manuf.*, vol. 68, Apr. 2021, Art. no. 102072.
[36] M. F. S. Shaughnessy, "An interview with Miriam Schcolnik: Reading, E-reading and writing and their assessment," *Eur. Sci. J. ESJ*, vol. 16, no. 28, pp. 1–22, Oct. 2020.
[37] N. D. Gaybullaevna, "Formative assessment of students'knowledge—As a means of improving the quality of education," *Sci. Rep. Bukhara State Univ.*, vol. 5, no. 3, pp. 144–155, Jun. 2021.
[38] N. Roy, M. V. Torre, U. Gadiraju, D. Maxwell, and C. Hauff, "Note the highlight: Incorporating active reading tools in a search as learning environment," in *Proc. Conf. Human Inf. Interact. Retr.*, Mar. 2021, pp. 229–238.
[39] J. V. Hoffman and C. L. Kugle, "A study of theoretical orientation to reading and its relationship to teacher verbal FeedBack during reading instruction," *J. Classroom Interact.*, vol. 18, no. 1, pp. 2–7, 1982.
[40] D. H. Schunk and J. M. Rice, "Learning goals and progress FeedBack during reading comprehension instruction," *J. Reading Behav.*, vol. 23, no. 3, pp. 351–364, Sep. 1991.
[41] B. Alenljung, J. Lindblom, R. Andreasson, and T. Ziemke, "User experience in social human–robot interaction," in *Rapid Automation: Concepts, Methodologies, Tools, and Applications*. Hershey, PA, USA: IGI Global, 2019, pp. 1468–1490.
[42] I. Lasri, A. Riadsolh, and M. Elbelkacemi, "Self-attention-based bi-LSTM model for sentiment analysis on tweets about distance learning in higher education," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 12, p. 119—141, 2023.
[43] E. S. Shourmasti, R. Colomo-Palacios, H. Holone, and S. Demi, "User experience in social robots," *Sensors*, vol. 21, no. 15, p. 5052, Jul. 2021.
[44] J. Li, M. Galley, C. Brockett, G. P. Spithourakis, J. Gao, and B. Dolan, "A persona-based neural conversation model," 2016, *arXiv:1603.06155*.
[45] M. Hassenzahl, A. Wiklund-Engblom, A. Bengs, S. Hägglund, and S. Diefenbach, "Experience-oriented and product-oriented evaluation: Psychological need fulfillment, positive affect, and product perception," *Int. J. Hum.-Comput. Interact.*, vol. 31, no. 8, pp. 530–544, Aug. 2015.
[46] J. Gerłowska, U. Skrobas, K. Grabowska-Aleksandrowicz, A. Korchut, S. Szklener, D. Szczęśniak-Stańczyk, D. Tzovaras, and K. Rejdak, "Assessment of perceived attractiveness, usability, and societal impact of a multimodal robotic assistant for aging patients with memory impairments," *Frontiers Neurol.*, vol. 9, p. 392, Jun. 2018.
[47] M. Destephe, M. Brandao, T. Kishi, M. Zecca, K. Hashimoto, and A. Takanishi, "Walking in the uncanny valley: Importance of the attractiveness on the acceptance of a robot as a working partner," *Frontiers Psychol.*, vol. 6, p. 204, Feb. 2015.
[48] S. Singh, E. D. Olson, and C.-H. K. Tsai, "Use of service robots in an event setting: Understanding the role of social presence, eeriness, and identity threat," *J. Hospitality Tourism Manage.*, vol. 49, pp. 528–537, Dec. 2021.
[49] N. Liang and G. Nejat, "A meta-analysis on remote HRI and in-person HRI: What is a socially assistive robot to do?" *Sensors*, vol. 22, no. 19, p. 7155, 2022.
[50] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer, "Assessing social presence in asynchronous text-based computer conferencing," *J. Distance Educ./Revue de l'ducation Distance*, vol. 14, no. 2, pp. 50–71, 1999.
[51] S. C. Cobb, "Social presence and online learning: A current view from a research perspective," *J. Interact. Online Learn.*, vol. 8, no. 3, pp. 1–14, 2009.
[52] D. Belanche, L. V. Casaló, J. Schepers, and C. Flavián, "Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The humanness-value-loyalty model," *Psychol. Marketing*, vol. 38, no. 12, pp. 2357–2376, Dec. 2021.
[53] J. M. Kory-Westlund and C. Breazeal, "Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning," *Frontiers Robot. AI*, vol. 6, p. 54, Jul. 2019.
[54] M. Lombard, T. B. Ditton, D. Crane, B. Davis, G. Gil-Egui, K. Horvath, J. Rossman, and S. Park, "Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument," in *Proc. 3rd Int. Workshop Presence*, Delft, The Netherlands, vol. 240, 2000, pp. 2–4.

[55] J. J. Cummings and E. E. Wertz, "Capturing social presence: Concept explication through an empirical analysis of social presence measures," *J. Comput.-Mediated Commun.*, vol. 28, no. 1, Nov. 2022, Art. no. zmac027.
[56] I. Leite, C. Martinho, A. Pereira, and A. Paiva, "As time goes by: Long-term evaluation of social presence in robotic companions," in *Proc. RO-MAN 18th IEEE Int. Symp. Robot Hum. Interact. Commun.*, Sep. 2009, pp. 669–674.
[57] S. A. Andel, T. de Vreede, P. E. Spector, B. Padmanabhan, V. K. Singh, and G.-J. de Vreede, "Do social features help in video-centric online learning platforms? A social presence perspective," *Comput. Hum. Behav.*, vol. 113, Dec. 2020, Art. no. 106505.
[58] C. A. Jones-Roberts, "Increasing social presence online: Five strategies for instructors," *FDLA J.*, vol. 3, no. 1, p. 8, 2018.
[59] M. Akcaoglu and E. Lee, "Increasing social presence in online learning through small group discussions," *Int. Rev. Res. Open Distrib. Learn.*, vol. 17, no. 3, pp. 1–17, May 2016.
[60] J. E. Kendall and K. E. Kendall, "Enhancing online executive education using storytelling: An approach to strengthening online social presence," *Decis. Sci. J. Innov. Educ.*, vol. 15, no. 1, pp. 62–81, Jan. 2017.
[61] J. Smallwood, K. S. Brown, C. Tipper, B. Giesbrecht, M. S. Franklin, M. D. Mrazek, J. M. Carlson, and J. W. Schooler, "Pupillometric evidence for the decoupling of attention from perceptual input during offline thought," *PLoS ONE*, vol. 6, no. 3, Mar. 2011, Art. no. e18298.
[62] S. Hutt, K. Krasich, C. Mills, N. Bosch, S. White, J. R. Brockmole, and S. K. D'Mello, "Automated gaze-based mind wandering detection during computerized learning in classrooms," *User Model. User-Adapted Interact.*, vol. 29, no. 4, pp. 821–867, Sep. 2019.
[63] B. J. Jongkees and L. S. Colzato, "Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review," *Neurosci. Biobehavioral Rev.*, vol. 71, pp. 58–82, Dec. 2016.
[64] S. D'Mello and A. Graesser, "Mining bodily patterns of affective experience during learning," in *Educational Data Mining*. PA, USA: Citeseer, 2010.
[65] S. D'Mello, T. Jackson, S. Craig, B. Morgan, P. Chipman, H. White, N. Person, B. Kort, R. E. Kaliouby, and R. Picard, "Autotutor detects and responds to learners affective and cognitive states," in *Proc. Workshop Emotional Cognit. Issues Int. Conf. Intell. Tutoring Syst.*, 2008, pp. 306–308.
[66] J. Li, G. Ngai, H. V. Leong, and S. C. F. Chan, "Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics," *ACM SIGAPP Appl. Comput. Rev.*, vol. 16, no. 3, pp. 37–49, Nov. 2016.
[67] A. Veronese, M. Racca, R. S. Pieters, and V. Kyrki, "Probabilistic mapping of human visual attention from head pose estimation," *Frontiers Robot. AI*, vol. 4, p. 53, Oct. 2017.
[68] J. M. Harley, F. Bouchet, M. S. Hussain, R. Azevedo, and R. Calvo, "A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system," *Comput. Hum. Behav.*, vol. 48, pp. 615–625, Jul. 2015.
[69] S. A. Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: A back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems*, Dresden, Germany: Springer, 2012, pp. 114–130.
[70] S. Alelyani, J. Tang, and H. Liu, "Feature selection for clustering: A review," in *Data Clustering*. Oxfordshire, U.K.: Taylor & Francis, 2018, pp. 29–60.
[71] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202.
[72] M. Alkhayrat, M. Aljnidi, and K. Aljoumaa, "A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA," *J. Big Data*, vol. 7, no. 1, pp. 1–23, Dec. 2020.
[73] D. Luo, X. Wan, J. Liu, and T. Tong, "Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range," *Stat. Methods Med. Res.*, vol. 27, no. 6, pp. 1785–1805, Jun. 2018.

YOON LEE received the bachelor's degree in industrial design from Hongik University, Seoul, South Korea, in 2015, the master's degree in industrial design engineering from the Delft University of Technology, The Netherlands, in 2019. She is currently pursuing the Ph.D. degree with the Faculty of Electrical Engineering, Mathematics, and Computer Science, and she is affiliated with the Leiden-Delft-Erasmus Centre for Education and Learning (LDE-CEL). Her master's thesis uses a machine learning approach to explore a soundscape with the Internet of Things (IoT) design for intensive care units (ICU) with the Erasmus University Medical Center.

From 2015 to 2017, she started her career as a Research Engineer with Hyundai Motor Group. She participated as a Design Lead for a hardware and software development project for mother machines, a government research project that resulted in notable achievements, including productions, design patents, and winning design awards. Her research interests include deep learning, behavior-based attention analysis, educational feedback design, and human–robot interaction (HRI), focusing on developing feedback loop systems by bridging computer vision and multimodal interactions.

GOSIA MIGUT received the Ph.D. degree in explainable machine learning from the University of Amsterdam. She is currently an Assistant Professor of computing education with the Delft University of Technology. Her research interests include innovating university education, in particular better teaching machine learning and machine learning applications to enhance the learning and teaching experience.

MARCUS SPECHT received the Diploma degree in psychology and the Ph.D. degree in adaptive information technology from the University of Trier, Trier, Germany, in 1995 and 1998, respectively.

Since 2001, he has been the Head of the Department of Mobile Knowledge, Fraunhofer Institute for Applied Information Technology (FIT), Sankt Augustin, Germany. From 2005 to 2018, he was a Professor of learning technologies with the Open University of the Netherlands, Heerlen, The Netherlands, and the Head of the Learning Innovation Laboratory. He is currently a Professor of digital education with the Delft University of Technology, Delft, The Netherlands, and the Director of the Leiden-Delft-Erasmus Centre for Education and Learning. His research interests include mobile and contextualized learning technologies and social and immersive media for learning.

Prof. Specht is a member of ACM, SIKS, and ICO Research Schools, The Netherlands. He is also an Apple Distinguished Educator. He was the President of the International Association of Mobile Learning, from 2013 to 2015.