POLITECNICO DI TORINO
Repository ISTITUZIONALE

Towards an integrated framework to measure user engagement with interactive or physical products

Publisher copyright

(Article begins on next page)

29 March 2023

ORIGINAL PAPER

Towards an integrated framework to measure user engagement with interactive or physical products

Ivonne Angelica Castiblanco Jimenez[1] · Juan Sebastian Gomez Acevedo[1] · Federica Marcolin[1] · Enrico Vezzetti[1] · Sandro Moos[1]

Abstract
Building great products or services is not easy; users want products and services that exceed their expectations and evolve with their needs; it is not just about building the right features. Knowing the user engagement (UE) towards a physical, virtual product or service can give valuable information that could be used as feedback for the design, enhancing its chances of success. In the context of user-centered design, UE is the assessment of the user experience characterized by the study of the individual's cognitive, affective, and behavioral response to some stimulus, such as a product, a service, or a website. UE considers not only the users' requirements and wishes but also their perceptions and reactions during and after an interaction with a product, system, or service. Many studies looking to quantify the UE are available. Still, a framework that provides a generic view of the most commonly used methods and metrics to measure UE does not yet exist in the literature. Aiming to understand the UE better, in this research, we developed a conceptual framework summarizing the available metrics and techniques used across different contexts, including good practices of self-report methods and physiological approaches. We expect this study will allow future researchers, developers, and designers to consider the UE as one of the most prominent product/service success indicators and use this guideline to find the more appropriate method, technique, and metric for its measurement.

1 Introduction

User engagement (UE) is an attribute of the user experience characterized by assessing the user's cognitive, affective, and behavioral investment when interacting with a digital system [1]. The UE can be interpreted as the level of involvement and absorption into an activity [2]. It has been widely defined as the process involving users through interactive experiences to create and enhance user-product relationships [3]. In this context, the product can be a physical or virtual object, service, or experience online and offline.

While developing a product, choosing carefully the most suitable methods to know the user's requirements and needs becomes an advantage [4, 5]. In this sense, User-centered design (UCD) is one of the most comprehensive methodologies and philosophies for designing interactive products [6]. It enables the creation of valuable and usable products by significantly involving users [6] in every stage of product development. Following the UCD perspective, the UE considers not only the users' requirements and wishes but also their perceptions and reactions during and after an interaction with a product.

UE is a multifaceted, complex phenomenon with several potential and different measurement approaches [8]. However, it is an abstract construct that manifests differently within the whole spectrum of uses and applications; for this

✉ Ivonne Angelica Castiblanco Jimenez
  ivonne.castiblanco@polito.it

  Juan Sebastian Gomez Acevedo
  s261410@studenti.polito.it

  Federica Marcolin
  federica.marcolin@polito.it

  Enrico Vezzetti
  enrico.vezzetti@polito.it

  Sandro Moos
  sandro.moos@polito.it

[1] Department of Management and Production Engineering, Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy

reason, one of the challenges is to define how the Engagement can be measured so that it can be used in design and evaluation. Studies have shown the necessity for a more context-specific approach when constructing measurement scales regarding the Engagement and the desire to obtain a set of items that allows measuring it in different sectors and natures [7].

Moreover, due to the variability of its application in different contexts and scenarios, an exact definition of "user" within the UE terminology is unclear. According to the various definitions of "user" and "consumer," in this study, we use these terms as synonyms, considering the user as the final consumer of a product, service, or experience. For example, in education, the end-user is the student; in health services is the patient; in the marketing area, the buyer; and in video games, the player.

All methodological approaches have advantages and limitations regarding UE use in specific populations and settings, and the measures may capture interactions subjectively or objectively [8]. As McNeal et al. [9] realized, performing direct UE measurement and establishing associated metrics can be challenging. There are studies conducted in fields like education, health, videogames, marketing, arts, and social media, among others. Still, one of the major gaps in the UE literature is the lack of evidence about how it should be standardized and measured [136].

Many studies have aimed to quantify the UE; however, a framework that provides a global and generic view of the different scales and methods used to measure it does not yet exist in the literature. After performing a literature review analysis, we found that most methods and metrics are classified into two main focuses: self-report or physiological. In this paper, the term "physiological" describes the measurement done using a physical indicator and an associated instrument that depends on biological responses, for instance, measuring the students' engagement using electrodermal activity [2, 9] or heart rate analysis [10], or EEG signals [11, 12].

The methods and tools are widely varied; this opens new opportunities for evaluation and metric selection. Thus, with so many options for measuring UE, there is a need for an updated conceptual framework that summarizes the measurement methods and metrics and classifies them according to their parameters, tools/instruments, and context of use.

This study aims to provide a literature review regarding UE's most commonly used measurements and develop a conceptual framework of the current methods and metrics used for Engagement measurement. Specifically, the study addresses two objectives: first, summarize the leading sectors and measures found in the literature. Second, develop a conceptual framework that outlines the most commonly used methods and metrics, particularly physiological approaches, according to their application context. For this study, we use the definition proposed by Maxwell [138], who described

the conceptual framework purpose as "clarify, explain, and justify methods" and later complemented by Burkholder et al. [139] as a theory or literature review that informs and describes the development of a research question, data selection and collection, analysis, and presentation of findings.

We expect this framework will allow future researchers to start from an initial state of the art, where the methods, techniques, and, whenever possible, the measurement scale used to measure UE would be easily found. Thus, this study's perimeter offers guidance and a general scenario of the most commonly used methods, tools, and techniques.

This paper is divided into the following sections. Section 2 sets the methodology used for the review process. Section 3 reports the literature review analysis, exploring the most frequently used methods according to their parameters. Section 4 includes the results included in the framework, and in Sect. 5, there are conclusions and discussion.

2 Method

Authors such as Schimanski et al. [13] and Dresch et al. [14] have made explicit the importance of the information flow in science research [13, 14]; therefore, for carrying out the literature review, we have followed the approach used by Motyl and Filippi [137]. In Fig. 1, we have depicted the flowchart process used for our literature review.
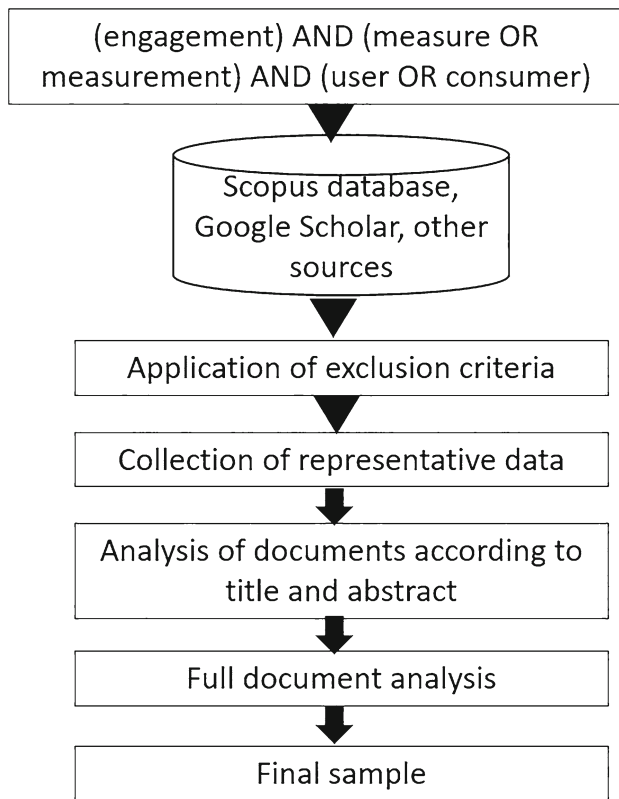
For this research, the question posed is defined as follows, "What are the methods used to measure the user engagement?". Consequently, the search string used is as follows: (engagement) AND (measure OR measurement) AND (user OR consumer). The search was carried out in the Scopus DB and Google Scholar from years 2000 to 2020. Also, we have considered proceedings documents and additional sources such as Microsoft academics and Crossref. The following exclusion criteria were applied: duplicated records; documents not in English; documents written before 2000; and from the other sources, as crossref, only documents with registered doi.

The search protocol in Scopus DB and Google scholar initially matched 157 documents from 2000 to 2009, 229 documents from 2010 to 2015, and 493 from 2016 to 2021. It resulted in a total of 879 for the last 20 years and more than 600 additional documents from other sources. In Table 1, there is a summary of the document's quantity for each year set in the first step of the flowchart.

After the initial match, we applied the exclusion criteria, obtaining 508 documents. Finally, after the abstract and full document analysis, the number of documents considered for this review was 380. After data analysis from the last 20 years, the numbers show an increase in the number of citations in this regard. The most influential articles had more than 300 citations in the last two decades. Particularly, between the

Table 1 Number of documents by year

	(2000–2009)		(2010–2015)		(2016–2021)	
	Papers	Cites/year	Papers	Cites/year	Papers	Cites/year
Google scholar	110	232	132	300	283	357
Scopus	47	112	97	229	210	233

Fig. 1 Literature review flowchart

Table 2 Classification of articles

		# Considered articles
Self-report methods	(Surveys, Interviews, Questionnaires)	57
Physiological methods	Skin Conductance	31
	Heart rate	22
	EEG	36
	Pupillometry	30
	Posture	26
	Respiratory rate	28
	Facial expression	18

tools are Skin conductance, Heart rate, Electroencephalography (EEG), Pupillometry, Posture analysis, Respiratory rate, and Facial expression.

3 Review of user engagement methods and technics

The literature review analysis shows that the UE measurement methods can be classified into two main groups, self-report and physiological. Many examples of good practices are available in the literature; according to these, the self-report methods can be categorized by approach (quantitative or qualitative) and type (interview, questionnaire, or survey). In addition, we have noticed that UE dimensionality was considered in many studies to describe and compare the self-report methods. Regarding the physiological methods, we have found seven approaches, as reported in Table 2.

Still, there is no evidence of a classification compendium for measuring the UE from the literature review. To fill this gap, in the following sections, we will explore the methods and technics most used according to the literature review analysis; our interest is to give an overview of the current UE methods to help researchers and practitioners quickly recognize the available methodologies for analyzing the engagement with a product or service. Table 3 shows an overview of the methods and parameters that will be reviewed in the following sections. Later, in Sect. 4, the conceptual

years 2008 to 2010, the studies made by O'Brien and Toms [18], who developed a multidimensional qualitative scale to test the engagement with software applications; Sprott, Czellar, and Spangenberg [49] who established an eight-item scale to measure the brand engagement and Brockmyer et al. [20] who created a questioner to measure the engagement in video game-playing, impacted the current research on this subject.

After the procedure illustrated in Fig. 1, we classified the documents according to the method used, obtaining a collection of articles for each type (Table 2). Based on the results obtained, we found that most studies applied self-report methods to measure the UE, while others used physiological methods. Over the studies analyzed, we found that the most common methods used to measure UE using physiological

Table 3 Overview of the parameters subdivisions of the methods for measuring user engagement

	Type	Parameter	Description
Methods for measuring user engagement	Self-report methods	Self-report type	Questionnaire, survey, or interview
		Approach	Qualitative or Quantitative
	User engagement dimensionalities	Concept	Type of engagement used or identified
		Dimensionality	Multi or unidimensionality
		Dimensions	Sub-constructs that define a level of interaction
	Physiological methods	Physiological measurements	Skin conductance, Heart rate, Electroencephalography (EEG), Pupillometry, Posture analysis, Respiratory rate, and Facial expressions
		Measurement range period	Periods in which the measurement is performed
		Procedure for data analysis	Process used to measure the UE through the implemented method
		Engagement scales	Scale to measure the engagement used or identified
		Equations	Equations used to calculate or facilitate the UE understanding or classification

framework will contain the analysis results obtained during the reviews carried out for each type of method.

3.1 User engagement self-report methods

Self-reported methods are assessments where users are asked to report their responses directly. Many standard measurements of attitudes, such as Likert scales and semantic differentials, are self-report. Similarly, constructs of interest to researchers, such as behavioral intentions, beliefs, and retrospective reports of behaviors, are often measured using this method [134, 141].

The self-report methods used to analyze the UE can be classified according to type or approach. Regarding the type, the studies typically used questionnaires, interviews, and surveys. Authors such as Lalmas et al. [8], Fredricks and McColskey [16], and Henrie et al. [17] have made significant contributions to the recompilation of self-report instruments.

The self-reports can also be categorized by the approach used (qualitative or quantitative); in this case, the classification is made according to the following criteria: quantitative,

the instrument that includes numerical scales, and qualitative the instruments that do not contain numerical scales in its measurement.

From the literature review, considering these factors, the self-report type, the approach, and the context of the application (sector), we found numerous tools and methods to measure the UE. Many of these tools and techniques were created ad-hoc for each study, so the chances of standardization are low. However, some instruments are more frequently used. For example, the user engagement scale (UES) [18] is one of the most used standard questionnaires since it aims to measure the quality of the user experience. The National Survey Student Engagement (NSSE) measures student participation in Canada and the United States concerning learning and engagement. At the same time, the Student Engagement Questionnaire (SEQ) [19] aims to capture student feedback and perception in a learning environment. In the field of task demand, the National Aeronautics and Space Administration (NASA) uses the NASA-TLX instrument as a subjective workload assessment tool to allow users to perform subjective workload assessments on the operator while

Table 4 Self-report UE measurement

Tool/instrument	Self-report type	Reference	Sector	Approach
Student engagement questionnaire (SEQ)	Questionnaire	Leung and Kember [19]	Education	Qualitative
National survey of student engagement (NSSE)	Survey	Administered by Indiana University School of Education	Education	Qualitative
High school survey on student engagement (HSSSE)	Survey	Center for evaluation and Education Policy, Indiana University	Education	Qualitative
Community college survey of student engagement (CCSSE)	Survey	Center for Community College Student Engagement, University of Texas	Education	Qualitative
NASA-TLX	Questionnaire	National Aeronautics and Space Administration (NASA)	Task demand	Qualitative
Children's engagement questionnaire	Questionnaire	McWilliam [146]	Education	Qualitative
Structured interview matrix (SIM) (For collaboration and engagement)	Interview	O'Sullivan et al. [21]	Community Resilience	Qualitative
Engagement interview protocol (EIP)	Interview	Yeung et al. [22]	Immigration	Qualitative
Dundee stress state questionnaire (DSSQ)	Questionnaire	Matthews et al.[23]	Psychological	Quantitative
Short stress state questionnaire (SSSQ)	Questionnaire	Helton [24]	Task Engagement	Quantitative
Patient and family engagement (PFE)	Survey	Herrin et al. [25]	Healthcare	Quantitative
Shirom-Melamed vigor measure (SMVM)	Questionnaire	Shirom [26]	Work	Quantitative
User engagement scale (UES)	Survey	O'Brien and Toms [18]	Marketing	Quantitative
Utrecht work engagement scale (UWES-9) shortened version	Questionnaire	Schaufeli et al. [27]	Work	Quantitative
Game engagement questionnaire	Questionnaire	Brockmyer et al. [20]	Video games	Quantitative
Affective benefits and costs of communication technology (ABCCT)	Interview	Yarosh et al. [28]	Communication technologies	Quantitative
Athlete engagement questionnaire (AEQ)	Q uestionnaire	Lonsdale et al. [130]	Sports	Quantitative
Student engagement in schools questionnaire (SESQ)	Questionnaire	Hart et al. [29]	Education	Quantitative
Engagement questionnaire (EQ)	Questionnaire	Hannum and Simons [30]	Food	Quantitative
The music use (Muse) questionnaire	Questionnaire	Chin and Rickard [31]	Music	Quantitative
Music engagement questionnaire (MusEQ)	Questionnaire	Vanstone et al.[32]	Music	Quantitative
Student course engagement questionnaire (SCEQ)	Questionnaire	Handelsman et al. [33]	Education	Quantitative
Game user satisfaction scale (GUESS)	Questionnaire	Phan et al. [34]	Video games	Quantitative

working with human–machine interface systems [127]. The Game Engagement Questionnaire [20] identifies the user's psychological engagement when playing video games. Table 4 summarizes a collection of good practices using self-report methods for the UE analysis in different segments.

Many UE studies have used almost exclusively self-report measurements; however, recent advances are making other types of evaluations more feasible. For example, self-reported outcomes can be contrasted to others that do not rely on respondents' reports, as physiological approaches that measure respondents' behaviors, sometimes in a constrained or controlled environment [141].

3.1.1 User engagement dimensionalities

Many researchers have explored the theoretical foundations as a systematic conceptualization of the UE [35], defining different types of sub-constructs or types of engagement [17], known as *User Engagement dimensionality* (UED), another way to compare self-report assessments [16].

UED describes the magnitude that captures a person's idea as a goal-oriented being equipped with a set of processes that guide their behavior in a changing environment. In other words, it is the level of interaction between a user and a product in a specific context; meanwhile, the user develops a perception that may or may not generate an engagement. The dimensions definition depends on the product; moreover, it depends on the potentially engaged user perspective.

From the studies' examination, it is possible to notice considerable differences in the dimensions that constitute the term *engagement* [36]. Two main dimensionality groups emerged, defined as *uni-dimensionality* and *multi-dimensionality*; these groups allow to develop and estimate diverse theoretical and empirical models about how various characteristics of customer relations with a product impact the user actions, behaviors, and intentions.

The uni-dimensionality is commonly expressed [35] as one of the three following engagement aspects: I) Emotional – Affective, II) Cognitive, or III) Behavioural. It retains the characteristic of simplicity and concept uniqueness, without extensive in-depth analysis and dimensions combination, in contrast to the multi-dimensionality that focuses dynamically on the three aspects defined previously. There is no agreement on the elements that should be primarily considered in uni-dimensionality or multi-dimensionality, but both provide helpful insights concerning the UE concept's appropriateness [37]. Furthermore, despite the number of studies supporting the UE's importance, there are still sharp controversies about its definition and the number of dimensions it includes [36, 37].

Furthermore, studies on UE must consider the characterization of the intended measure; in this regard, an important definition comes from Diamantopoulos and Winklhofer [140], who described the difference between reflective and formative measurements. Measurement development can focus on items composing a scale, perceived as reflective indicators of an underlying construct. Or, as an alternative measurement perspective, it could be based on formative indicators that involve creating an index of a weighted sum of variables rather than a scale.

The sub-constructs vary between authors and segments of application. However, we observed that the multi-dimensionality defined by Fredricks et al. [38] as Cognitive, Emotional, and Behavioural, and subsequently reviewed by Brodie et al. [35], Hollebeek [39], among others, is the most popular. Within the last five years, the research carried out by Dessart et al. [36] is particularly relevant for the topic; the study provides a classification of dimensional UE scales in the context of online brand communities. According to the literature review, it is possible to infer that the multi-dimensionality accomplishes better the principles that define the concept of UE because it includes a more in-depth analysis of user perception. Table 5 summarizes some examples of dimensional classifications for different segments based on and adapted from Brodie et al. [35], Hollebeek [39], and Dessart et al. [36].

3.2 Physiological methods

Body language is a more reliable and authentic form of transmitting information than verbal communication [67]. Without realizing it, the body continually sends information about intentions, feelings, and behaviors. Even without verbal expressions, the physiological factors speak for the body and can be very significant.

During the development process, to create a successful product with higher positioning and differentiation [68], it is crucial to choose the most appropriate method to understand the user requirements [4, 5]. For these reasons, as anticipated in Table 3, we have classified the existing physiological methods used to measure the UE level with a product according to: the *type of physiological measurement*, *measurement range period* (time in which the measurement is performed), the *procedure for data analysis* (process used to measure the UE through the implemented method), *engagement scale* (used or identified scale to measure the UE), and the *equations* used to calculate or facilitate the UE understanding and classification.

The reviewed physiological methods are Skin conductance, Heart rate, Electroencephalography (EEG), Pupillometry, Posture analysis, Respiratory rate, and Facial expressions. These methods will be further described in the following sections.

3.2.1 Skin conductance

Electrodermal activity (EDA), also known in the literature as skin conductance response and galvanic skin response (GSR), are terms used for defining the continuous autonomic variation in the electrical properties and characteristics of the skin. It can reveal the humans' physiological arousal, connected to a subject's attention and alertness, and is considered a suitable approach for measuring the UE during different tasks, stimuli, or situations [2, 9].

GSR includes a slow variation or tonic component known as the skin conductance level (SCL) and the skin conductance response (SCR). The tonic component SCL is interrupted by increased skin conductance due to a particular stimulus [9]. Studies have used a variety of sensors to measure the variation of the electrical properties and characteristics of the skin. For example, Morrison et al. [69] and Di Lascio et al. [2] used the Empatica E4 wristband galvanic sensor; this sensor allows to record the skin conductance variation measured in μSiemens in a specific time interval and report the user experience. The task could be split into three main periods to guarantee a good

Table 5 UE Dimensionality sub-constructs

| Author(s)	Segment	Concept	Dimensionality	Dimensions/sub-constructs
Achterberg et al. [40]	Psychology	Social engagement	Uni	Social behavior
Jennings and Stoker [41]	Sociology	Civic engagement	Multi	Media attentiveness, trust, political involvement
Fredricks et al. [38]	Education	School engagement	Multi	Cognitive, emotional, behavioural
Algesheimer et al. [42]	Marketing	Brand community engagement	Uni	Motivational
Whelan and Wohlfeil, [43]	Marketing	Brand engagement	Uni	Behavioral
London et al. [44]	Education	Student engagement	Multi	Academic investment, motivation investment, commitment, psychological connection, comfort, sense of belonging
Bejerholm and Eklund [45]	Psychology	Occupational engagement	Multi	Daily rhythm of activity and rest, variety and range of occupations, place, social environment, social interplay, interpretation, extent of meaningful occupations, routines, initiating performance
Bowden [46]	Marketing	Consumer engagement process	Multi	Commitment, development of trust, involvement, eventually affective commitment
Higgins and Scholer [47]	Marketing	Consumer engagement	Multi	Cognitive, emotional, behavioural
Calder et al. [48]	Marketing	Online engagement	Multi	Experiential, social
Sprott et al. [49]	Marketing	Brand engagement in self-concept	Uni	Emotional/affective
Gambetti and Graffigna [50]	Marketing	Engagement	Multi	Soft (relational), pragmatic (Managerial)
Mollen and Wilson [51]	Marketing	Consumer engagement	Multi	Active sustained cognitive attention, experimental value, instrumental value
van Doorn et al. [52]	Marketing	Customer engagement behaviour	Uni	Behavioral
Phillips and McQuarrie [53]	Marketing	Advertising engagement	Multi	Engage to: act, identify, feel, transport, immerse
Verhoef et al. [54]	Marketing	Customer engagement	Uni	Behavioral
Fredricks and McColskey [16]	Education	School engagement	Multi	Cognitive, emotional, behavioural
Brodie et al. [35]	Marketing	Customer engagement	Multi	Cognitive, emotional, behavioural
Abdul-Ghani et al. [55]	Marketing	Engagement	Multi	Utilitarian, hedonic, social benefits
Wang et al. [56]	Education	School engagement	Multi	First-order dimensions: A) Attention, compliance. B) Belonging, valuing. C) Regulation, strategy second-order dimensions: Behavioural (A), Emotional (B), Cognitive (C)
Hollebeek [39]	Marketing	Consumer brand engagement	Multi	Cognitive, Emotional, Behavioural
Vivek et al. [57]	Marketing	Consumer engagement	Multi	Cognitive, Emotional, Behavioural, Social
Gambetti et al. [58]	Marketing	Consumer brand engagement	Multi	Experiential, social
Gummerus et al. [59]	Marketing	Customer engagement	Uni	Behavioral

Table 5 (continued)

Author(s)	Segment	Concept	Dimensionality	Dimensions/sub-constructs
Cambra et al. [60]	Marketing	Customer engagement	Multi	Commitment, loyalty, word-of-mouth
Vivek et al. [61]	Marketing	Consumer engagement	Multi	Cognitive, emotional, behavioural, social
Hollebeek and Chen [62]	Marketing	Brand engagement	Multi	Cognitive, emotional, behavioural
Hollebeek et al. [63]	Marketing	Consumer brand engagement	Multi	Cognitive, emotional, behavioural
Baldus et al.[64]	Marketing	Online brand community engagement	Multi	Brand influence, brand passion, connecting, helping, like-minded discussion, reward (utilitarian and hedonic), seeking assistance, self-expression, up-to-date information, validation
Dessart et al. [36]	Marketing	Consumer engagement	Multi	First-order level: A) Enthusiasm, Enjoyment. B) Attention, Absorption. C) Sharing, Learning, Endorsing Second-order level: Affective (A), Cognitive (B), Behavioural (c)
Hollebeek et al. [65]	Musical	Musical engagement	Multi	Social identity, transportive experience, affect-inducing experience
Violante et al. [66]	Marketing	Consumer engagement	Multi	First-order level: A) Enthusiasm, Enjoyment. B) Attention, Absorption. C) Sharing, Learning, Endorsing Second-order level: Affective (A), Cognitive (B), Behavioural (c)

Table 6 Sensor location examples

Mota and Picard [105]	Shirehjini et al. [108]		Bibbo et al. [106]	
Posture classification				
Leaning forward	Sitting on the edge of the seat	S1	All	P1
Leaning forward left	Sitting upright	S2	OverLapRightBack	P2
Leaning forward right	Slumping back	S3	OverLapLeftBack	P3
Sitting upright	Loose	S4	OverLapRight	P4
Leaning back	Leaning left	S5	OverLapLeft	P5
Leaning back left	Leaning right	S6	SeatAll	P6
Leaning back right	Right leg on left	S7	SeatFront	P7
Sitting on the edge of the seat	Left leg on the right	S8	SeatFrontBackUp	P8
Slumping back				
Pressure sensors location				
Two matrices of pressure sensors are located at the seat-pan and backrest. 64 integrated pressure sensors	Eight sensors are embedded in the seat and back of the chair		Chair equipped with eight analog tactile pressure sensors, four in the seat and four in the back	
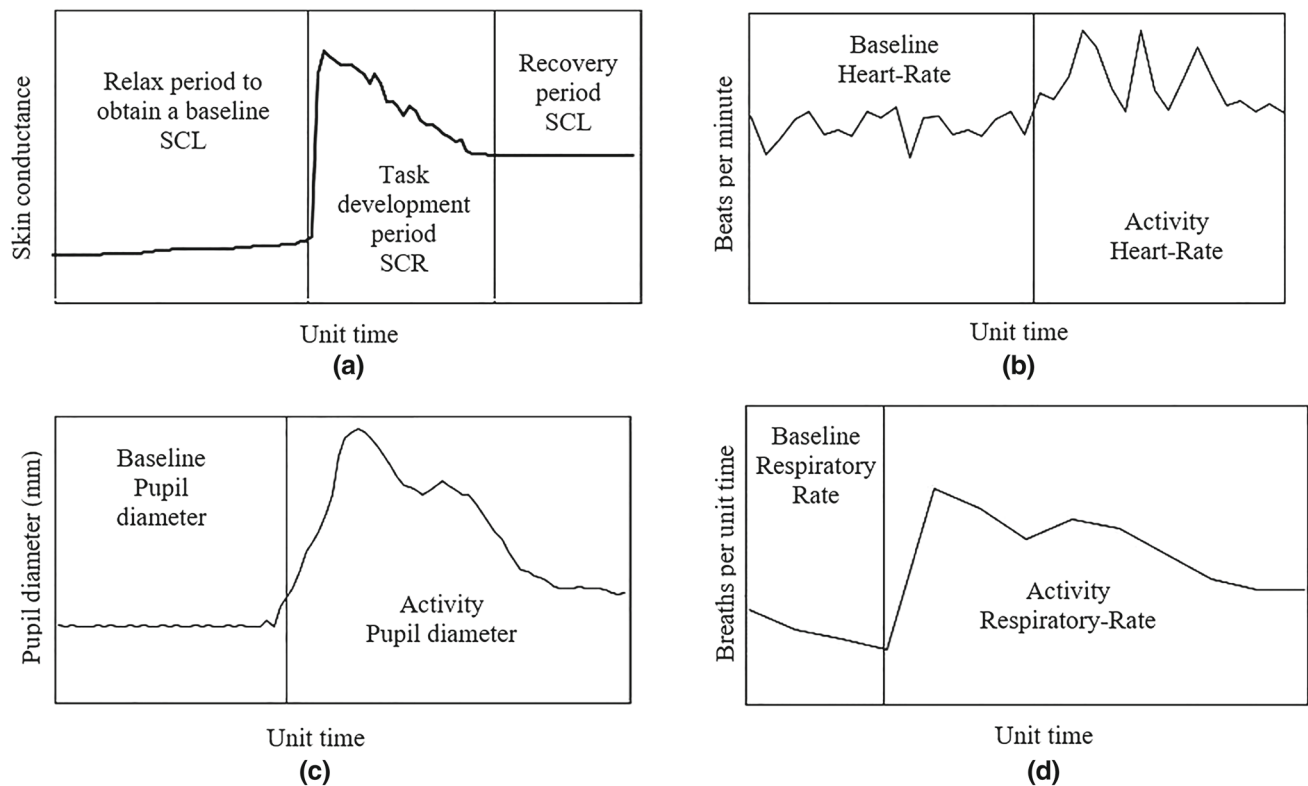
Fig. 2 Physiological time intervals. **a** Skin conductance variation. **b** Heart rate data graph. **c** Pupil diameter data graph. **d** Respiratory rate data graph

measure of the skin conductance variation [2, 69–72], as can be seen in Fig. 2(a):

- Relax period to obtain a baseline recording
- Task development period
- Recovery period

After filtering the data, it is possible to plot the variations (on the vertical axis, the Skin conductance measured in μSiemens, and on the horizontal axis, the unit time). In this order, it is possible to identify the three periods. It is critical to note that every user will have specific data, and every plot will have different variations according to the type of activity measured; consequently, it is difficult to generalize the level of engagement.

For that reason, and because the GSR signal is highly individual-dependent and can vary a lot from user to user [73, 76], many studies [2, 71–75] carried out normalization of the skin conductance data, comparing the users' skin conductance during the task development against their baseline recording. In order to counteract this dependency, the following normalization is typically used (Eq. 1):

$$\Phi ix = \frac{SCL(ix) - SCL(\min)}{SCL(max) - SCL(\min)} \tag{1}$$

Equation 1—Normalization of the skin conductance data. Lykken and Venables [131].

- Φix Indicated the normalized value
- $SCL(ix)$ Indicates the current value
- $SCL(\min)$ Indicates the minimum value
- $SCL(max)$ Indicates the maximum value

After normalizing the entire signals (rather than just the tonic component), it is possible to generate a general comparable data graph, as shown in Fig. 3.

Finally, Di Lascio et al. [2] stated that the level of UE in skin conductance could be measured using the normalized graph by dividing the different levels of skin conductance into five groups:

- Very high engagement between 0,8 and 1
- High engagement between 0,6 and 0,8
- Normal engagement between 0,4 and 0,6
- Low engagement between 0,2 and 0,4
- Very low engagement between 0 and 0,2
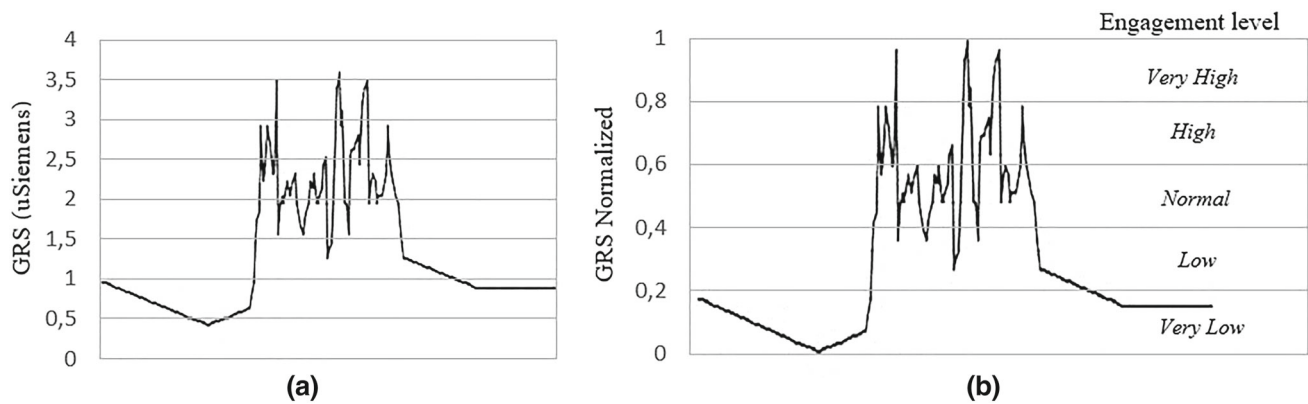
Fig. 3 **a** Original skin conductance data graph. **b** Standardized data graph

3.2.2 Heart rate

Heart rate (HR), known as pulse, is the number of times a person's heart contracts per minute; these contracts vary from person to person according to age, weight, body size, heart conditions, medication, and performed activities [129].

The HR is highly susceptible to variations due to the emotions and reactions of the subject; for that reason, different authors have used this physiological characteristic as a measure of UE. Previous studies have identified and confirmed that the HR and the level of engagement to a task relate to the HR variation during an activity development; this, because a person's heartbeat reflects emotional levels and reactions [10, 77–79, 134]. Furthermore, the HR tends to increase when the user is experiencing strong emotions and tends to decrease in an immersive environment [80, 81]**.**

The method to identify the engagement of a user using the HR [77, 81–83] can be generalized by separating the measure results into two stages (similar to GSR), the first to set the HR baseline and the second to measure the corresponding HR during the development of an activity as can be observed in Fig. 2(b). After data filtering, the HR can be analyzed graphically in two ways: the first one analyses the HR using a graph where the vertical axis corresponds to the beats per minute (BPM), and in the horizontal axis, the unit time; in this way, it is possible to identify the increases and decreases peaks in the HR measured periods. The second method calculates the HR variability (HRV) according to the alterations in the time interval between consecutive heartbeats in milliseconds. To measure the HR, the authors have used a wide range of trackers, and the selection depended on the type of study performed and the required accuracy.

Authors like Darnell and Krieg [10] proposed a linear regression of the recorded data to establish the variations throughout the activity execution, concluding that there is an HR decrease in comparison to the initial one, except for the fortuitous stimuli that can be due to factors internal or external to the test. Richardson et al. [78] proposed a data normalization to remove any baseline difference in the user data and verify the variation more concisely.

Consequently, depending on the type of activity, the engagement can be represented either by an HR increase or decrease, resulting in high variations in the general population's results. For example, Rooney et al. [132] identified a drop in the HR in a short film immersion but increased attention. At the same time, during lectures, Darnell and Krieg [10] concluded that there are increasing and decreasing HR variations that can characterize attention changes. Due to this discrepancy, we can conclude that the UE level will depend on the variation of the HR with respect to the baseline identified in the first measurement step; the interpretation made based on this indicator must consider that the results will vary from subject to subject, and according to the application context. More reflective environments intended to engage the user will tend to have a stable HR variation. In contrast, active environments with many stimuli will have a more significant HR variation concerning the baseline.

3.2.3 Electroencephalography

Electroencephalography (EEG) is a technique for recording and interpreting the brain's electrical activity. According to the user's state, the brain's nerve cells generate electrical impulses that fluctuate rhythmically in different patterns [128]. The electrical EEG signal is composed of different frequencies produced by neuronal electrical activity; these frequencies, known as bands, are classified as Delta(δ), Theta (θ), Alfa (α), Beta (β), and Gamma(γ). Each type of band reflects specific and different cognitive processing skills in particular brain areas [84]. Furthermore, studies have validated that EEG can provide metrics for determining task engagement and arousal [11].
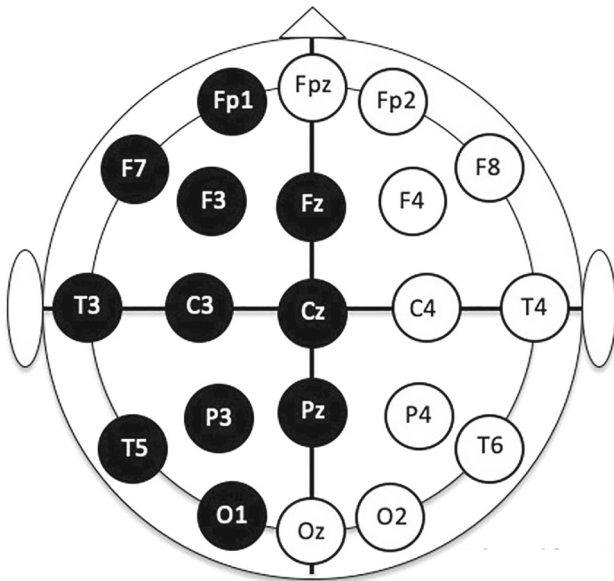
Fig. 4 10–20 system electrode placement

As well as GSD and HR, the EEG measurements regarding UE can be divided into two phases [11, 84, 85]; the first consists of baseline data to control the initial subject response. The second phase contains the electrical brain activity measurement while carrying out a task. For EEG measurements, a range of instruments are available in the market like the neuroSky MindWave, OpenBCI 3D printed devices, traditional Ag/AgCl electrodes, Neuron spectrum 1, and one of the most used in the last years, Emotiv EPOC [11, 86].

For the location of the electrodes, the studies typically used the 10–20 electrode placement system, as McMahan et al. [11], while using the Emotiv EPOC device, as well as, Nuamah and Seong [87]. In this system, each electrode has one or two letters to identify the lobe or brain area as follows: Pre-frontal (FP), Frontal (F), Temporal (T), Parietal (P), Occipital (O), Central (C), and includes the Z (zero) reference (FPz, Fz, Cz, Oz), see Fig. 4 for reference.

Many studies of EEG have identified different engagement indexes. Kang et al. [88] used the Neural Engagement Index (NEI), and Yamada [89] defined the frontal middle theta rhythm as a valuable index of attention. In the study carried out by Pope et al. [12] and confirmed later by Freeman et al. [85], the authors developed an EEG-based engagement index (Eq. 2), considered to be the most effective among different studies [11, 84, 86, 90, 91], this because this index reflects task engagement in a more precise and validated way. Furthermore, studies have shown that EEG-based indexes are reliably linked to various levels of decision-making tasks.

$$\text{EEG Index} = \frac{\text{Beta}(\beta)}{\text{Alpha}(\alpha) + \text{Theta}(\theta)} \qquad (2)$$

Equation 2. EEG index.

The study done by Pope et al. [12] identified the frequency bands as Theta (4–8 Hz), Alpha (8–13 Hz), and Beta (13–22 Hz). The theta waves describe a lower mental activity (resting state), while alpha waves are related to lower mental alertness, this one appears in a sleep–wake cycle. In contrast, Beta waves describe a higher mental activity, system activation, and alert state of the brain; in other words, Beta waves are related to an increase in brain activity during a task and allow the identification of attention changes due to external stimuli, indicating cognitive processes and changes in brain states while performing a task [91].

The EEG engagement index, in Eq. 2, identifies the UE level by comparing the initial user state with the stimulus responses; if it increases, a higher engagement is expected [11, 12, 84]. In this way, it is possible to identify the user engagement or disengagement with an activity, considering the baseline period versus the EEG data while performing a task.

3.2.4 Pupillometry

The pupil diameter analysis has been recognized for many years as a psychophysiological arousal index [92, 93]; pupil dilatation is related to a certain level of engagement during the performance of an activity [94] and a function of processing load or mental effort required to perform a cognitive task [95].

According to several authors [94–99], for the measurement, two periods during the test can be established to confront the data, one to define the baseline of the pupil diameter and the other for the measurement during the activity development. It is possible to plot the total variation of the pupil diameter or just the variation concerning the baseline data as depicted in Fig. 2.c. Also, it is essential to consider that the baseline pupil diameter gradually reduces when the user begins participating in a task [95, 100].

The measurement time for each period depends on the study. For example, authors like Jepma and Nieuwenhuis [99] for the first period identified that the pupil data during the 0,5 s immediately preceding the task could not be included as a baseline period because the user presented an anticipatory increase in the pupil diameter while Murphy et al. [101] used the mean pupil diameter over the 500 ms pre-stimulus as a baseline.

In the last years, authors have used different specialized instruments, like the Tobii eye-tracker 2150, brain vision analyzer, iView X MRI-SV eye-tracker, among others. To eliminate artifacts and blinks, authors [98–100] commonly used a linear interpolation algorithm.

The data analysis is correlated with the time-on-task and difficulty [93, 95, 96, 98], showing that the pupil diameter is more significant when a difficult task is presented, reaching

a peak that specifies an engagement point. Kassem et al. [102] suggest that the presence of arousal stimuli induces a dilatation of the pupil, confirming a rise around the moment of the stimuli independent of the baseline [103].

Podder et al. [104] proposed an engagement behavior function (EB) according to a Graph-Based visual saliency (GBVS) while presenting a video. This method is divided into two steps: in the first one, this model forms activation maps on individual feature channels and then normalizes them in a way that highlights conspicuity based on human visual perception and attention (Eq. 3) [104].

$$EB = f\left(PS_L, PS_R, F, \frac{1}{D_A}, \frac{1}{D_G}, \frac{1}{D_C}, \frac{1}{D_E}\right) \quad (3)$$

Equation 3. Behavior function.

In the function, all the variables have the same weight to create a ranking between the subjects in the study, proving the worst and best level of performance that can represent the concentration/engagement level. The variables' descriptions are the following:

- The left pupil size (PS_L)
- Right pupil size (PS_R)
- Blinking pattern (F)
- Distance between an individual's average gaze location to all participants' average gaze location as obtained by the eye tracker data (D_A)
- Distance between an individual's average gaze location to the average gaze location as predicted by the GBVS model (D_G)
- Distance between an individual's average gaze location to the center of the videos (D_C)
- Distance between an individual's average gaze location to the most significant points according to the video experts (D_E)
- Pupil sizes are measured in mm, blinking patterns in percentage, and gaze distances in pixels.

It is possible to conclude that the level of engagement depends on the pupil diameter variation in relation to the baseline. Additionally, Gilzenrat et al. [98] verified that increases in baseline pupil diameter would be associated with decreases in task utility and disengagement from the task, although reduced baseline diameter (but increases in task-evoked dilatations) would be related to task engagement.

The findings of Gilzenrat et al. [98] are considered a good base for analyzing the results. Furthermore, Eq. 3 is a valid approximation of the UE level between users of the same study group.

3.2.5 Posture

Posture is the most natural element to observe and interpret within nonverbal communication [105]. Posture is a primarily involuntary signal that can be involved in the communication process. Although it seems to go unnoticed, the brain receives and processes the information, often unconsciously, and responds automatically to a stimulus; for this reason, it is crucial to measure how the posture of a subject can indicate whether a user is engaged or not.

According to the literature, various systems for measuring posture have been developed, usually indicated as Body Pressure Measurement System (BPMS) and Sitting Posture Monitoring System (SPMS) [106, 107]; both systems look to identify the position of the user according to a reference system.

The most popular method to measure engagement evaluates the subject's posture (while solving a task seated) using pressure sensors located in the chair. Mota and Picard [105] and D'Mello and Graesser [107] used two pressure sensor matrices located at the seat-pan and in the seat's backrest. Shirehjini et al. [108] decided to embed the sensors in the chair, having a non-invasive system allowing the subject to adopt a normal seated posture reproducing real-life conditions; as well as Bibbo et al. [106] have done, using a chair equipped with eight analog tactile pressure sensors. The difference between authors is the location where the sensors are placed in the chair; Table 6 contains some examples of sensor locations.

A graphic example of these locations (Fig. 5) is the one used by Bibbo et al. [106], a study that shows adequate results.

Regardless of the pressure sensor used, different authors [105–108] conclude that a high level of UE exists when there is increased pressure on the chair seat with minimal movement and when the user tends to lean forward on the chair trying to get a better focus in the task [106]. On the other hand, a low UE level (disengagement) is identified when there are rapid changes in the pressure on the chair's seat and when the users have a significant distance between their face and the screen or activity [107, 109].

D'Mello and Graesser [107] used the BPMS produced by Tekscan to examine the pressure map during an emotional stimulus (called frame), finding three significant pressure-related functions: the average pressure (μ), the average coverage (C), and the average coverage change in a pressure mat ($a_{coverage}$). These functions are represented in Eq. 4. R is the number of rows in the pressure matrix, C is the number of columns, and P_{ij} the pressure of a sensing element in row

Fig. 5 Posture classification
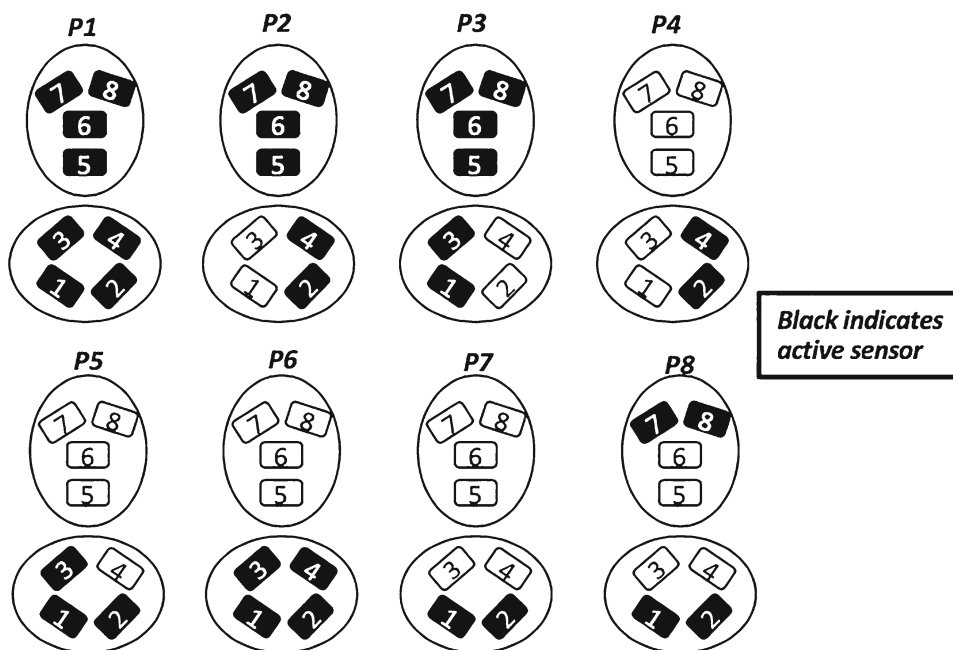
i and column j [107].

$$\mu = \frac{1}{R \times C} \sum_{i=1}^{R} \sum_{j=1}^{c} P_{ij}$$

$$C = \frac{1}{R \times C} \sum_{i=1}^{R} \sum_{j=1}^{c} X_{ij}$$

$$a_{coverage} = \frac{1}{N} \sum_{i=1}^{N} |C_t - C_{t+1}|$$

$$X_{ij} = 1, \ if \ P_{ij} > 1$$

$$X_{ij} = 0, \ if \ P_{ij} = 0 \qquad (4)$$

Equation 4 Pressure related functions

Equation 4 allows the BPMS system to efficiently identify pressure features to facilitate the measurements and better understand the results. In conclusion, regardless of the system used, the changes in the subject posture measured with pressure sensors can indicate the level of UE when performing a task. Specifically, the user is engaged when they tend to lean forward, exerting higher pressure on the seat.

3.2.6 Respiratory rate

Respiration is a way to consciously influence our physical and emotional responses, reflect feelings, relate to the environment, and impact the body's general state. It is a factor that reflects emotional response [110]. Its variability is an indicator of mental stress and sustained attention to a particular task [111]; therefore, it is considered a physiological factor that can measure the level of the UE.

According to several authors [111–114], and following the same procedure used in other physiological methods, the users' respiratory data is generalized in two stages, the first to set the baseline, the second during the development of the activity; in this way, it is possible to contrast the initial data with its variation (see Fig. 3d).

There are different types of instruments for measuring respiratory rate. For example, Webster and Colrain [113] used a mouth breathing mask. Wenger et al.[114] used a respiratory pneumotachograph to measure the respiratory volume. Gomez et al. [115] and Vlemincx et al. [111, 116] used a respiratory inductive plethysmograph to measure the rate and volume of each breath; and Vlemincx and Luminet [117] used two respiratory belts around the chest and the abdomen, besides making use of a breathing face mask.

According to the research, respiratory data, unlike the other physiological methods for measuring the UE, has not been a leading research method [115]. Only a few authors have analyzed this physiological factor and its relations with UE when performing an activity.

Vlemincx et al. [116] proposed some main respiratory parameters (Eq. 5) to measure respiratory variability, breath by breath, including Respiratory volume (Vi), Respiration Rate (RR) and Minute ventilation (MV).

$$RR = \frac{60}{\text{Total breath time}} \quad MV = \frac{RR}{Vi} \qquad (5)$$

Equation 5. Respiration Rate RR and Minute Ventilation MV.

Existing studies [110, 111, 114–116] demonstrate that random breathing variability increases during mental stress, difficulty, or excitement; and decreases with task-related attention. Also, the respiratory intensity during different emotional states varies according to the type and duration of the stimulus.

3.2.7 Facial expressions

Facial expressions are the changes that occur in our faces as a response to inner emotional states. In this sense, the UE analysis using facial expressions (no verbal communication type) aims to identify facial manifestations and their direct relationship with emotions. The facial expression analysis encodes and interprets the facial muscle movements to determine the internal emotional reactions. These reactions can be a rich source of social signals conveying the user's focus, attention, intention, motivation, and emotion [118].

Different studies have identified the importance of the time scale for recognizing facial expressions. Nezami et al. [119] used in their research three time-scale types: an entire video projection, a 10-s video clip, and single images. This time scale allowed them to categorize the facial expressions manually using labels and annotations to analyze and classify them according to a scale to rate the user status. This process is challenging due to inconsistencies between labelers/annotators, hence the need to automate this kind of process [120].

The literature review shows different computer-vision techniques to automate the UE analysis with a product. Several methods/architectures make it easier to analyze facial recognition, such as the Boostbf, Support Vector Machine (SVM), and the CERT Toolbox [120]. One of the most used is the Convolutional Neural Network (CNN) [119, 121, 122], a deep learning algorithm trained to recognize and classify facial expressions according to a preestablished facial data set.

The engagement scale parameters using facial expressions vary according to the study. For example, Whitehill et al. [120] defined it as I) Not engaged at all, II) Nominally engaged, III) Engaged in a task, and IV) Very engaged, Nezami et al. [119] defined it as I) Engaged and II) Disengaged, and Ramya et al. [123] defined it as I) Like, II) Dislike and III) Favourite. The main difference between studies that use CNN is related to the dataset used to train the network. For example, Nonis et al. [121] used the Bosphorus public database, Nezami et al. [119] used the dataset from the facial expression recognition challenge 2013, and Zhang et al. [122] combined temporal information with spatial information by applying CNN's to recognize facial expressions.

The study by Nonis et al. [121] shows adequate results regarding the relationship between facial expression and engagement of a user with a product. The study used an Intel® RealSenseTM SR300 and MobileNetV2 instruments and trained the network using the Bosphorus public database, then classified the engagement into three classes according to Russell's circumplex model of affect as follows: I) Deactivation: low engagement level with relaxed and calm expressions, II) Average: medium engagement level with happy, contented, serene and elated expressions, and III) Activation: high engagement level with alert and excitement expressions.

Olivetti et al. [125] also applied this classification while analyzing the user engagement during the interactions with a virtual environment using a Support Vector Machine (SVM) method to classify the engagement level according to the analysis of the facial expressions. Likewise, Violante et al. [126] used an SVM, and Russell's circumplex model of affect, generating three classes to define the inner users' requirements: I) Deactivation, II) Pleasure and III) Arousal.

It is possible to conclude that the analysis and recognition of facial expressions is an effective physiological method to identify the UE level with a product, mainly because of the adequate results found using CNN trained with a facial dataset. The classification of the UE used by Nonis et al. [121] can be used as a reference to identify the user status.

Finally, Table 7 shows a general guideline of the physiological methods discussed in this section, including the parameters of range period measurement, metrics, the procedure for data analysis, engagement scale, and equations.

4 Results

User engagement within a User-Centered Design context has been validated as an approach that integrates design visions with appropriate responses to user feedback and needs. Moreover, fields like manufacturing, engineering, virtual reality, and any technical discipline must consider the user's feelings and reactions towards an expected output during the first design steps; in other words, a wrong decision during the design process would affect the manufacturing operations directly. The design of a product and its formal specification must be combined with behavioral and organizational theories to develop an understanding of the user needs, contexts, and possible solutions [15].

We have realized that no document in the literature provides a generic view of the different methods and metrics used to measure user engagement. Thus, we aim to offer insights and guidelines for any design discipline that intends to implement UCD methods and apply user engagement as feedback for its analysis.

The results of this conceptual framework, summarized in Table 8, came from three analyses: the first one concerned the self-report methods, the second related to the conceptual

Table 7 Physiological methods for measuring UE

Physiological method	Measurement range periods	Metrics (concerning coordinate axis) Vertical	Horizontal	Procedure for data analysis	Engagement scale	Equations
Skin conductance	1) Relax period (Baseline) 2) Task development period 2) Recovery period	µ.Siemens	Unit time or period	Normalization (Φix) of the skin conductance data compared to the user skin conductance during the task development against their baseline and recovery recording	According to the normalized data [2] • Very high engagement Φix between 0,8 and 1 • High engagement Φix between 0,6 and 0,8 • Normal engagement Φix between 0,4 and 0,6 • Low engagement Φix between 0,2 and 0,4 • Very low engagement Φix between 0 and 0,2	$\Phi ix = \frac{SCL(ix)-SCL(\min)}{SCL(max)-SCL(\min)}$ Where: −Φix Indicated the normalized value −SCL(ix) Indicates the current value −SCL(min) Indicates the minimum value −SCL(max) Indicates the maximum value Taken from Lykken and Venables [131]
Heart Rate	1) Rest Heart Rate (Baseline) 2) Task development period	BPM or HRV		Compare the user's Heart rate during the task development against their baseline recording	The engagement can be represented by a heart rate increase or decrease depending on the activity	
Electroencephalography	1) Subject regulated measure period (Baseline) 2) Task development period	EEG engagement index		Extract frequencies for each band, calculate each period's EEG index and compare the index	The EEG index identifies the engagement level when it increases compared to the baseline index; a higher EEG index means a higher engagement level	$EEGIndex = \frac{Beta(\beta)}{Alpha(\alpha)+Theta(\theta)}$ Frequency bands: Theta (4–8 Hz), Alpha (8–13 Hz), and Beta (13–22 Hz) [12]
Pupillometry	1) Initial Pupil Diameter period (Baseline) 2) Task development period	Pupil diameter or pupil diameter dilatation (mm)		Compare the user's Pupil diameter or dilatation during the task development against their baseline recording	There is no consensus about a scale that indicates a level of UE using the pupil diameter analysis because the level of engagement depends on the variation of the pupil diameter for the baseline identified in the first measurement step, so the interpretation is made on this fact and varies from study to study. However, Gilzenrat et al. [98] verified that increases in baseline pupil diameter would be associated with decreased task utility and disengagement from the task. In contrast, reduced baseline diameter (but increases in task-evoked dilatations) would be related to task engagement It is possible to calculate an engagement behavior function (EB) [104]. It means that a higher EB function indicates a higher engagement ranking	$EB = f(PS_L, PS_R, F, \frac{1}{D_A}, \frac{1}{D_G}, \frac{1}{D_C}, \frac{1}{D_E})$

Table 7 (continued)

Physiological method	Measurement range periods	Metrics (concerning coordinate axis)		Procedure for data analysis	Engagement scale	Equations
		Vertical	Horizontal			
						PS_L and PS_R are the left and right pupil size measured in mm, F is the blinking pattern. D_A is the distance between an individual's average gaze location to all participants' average gaze location as obtained by the eye tracker data. D_G is the distance between an individual's average gaze location to the average gaze location as predicted by the GBVS model, D_C is the distance between an individual's average gaze location to the center of the videos, and D_E according to the experts, the distance between an individual's average gaze location to the most significant point. Taken from Podder et al. [104] In the function, all the variables have the same weight to create a ranking between the subject in the study that allows demonstrating the worst and best level of performance that could represent the concentration/engagement level according to the study carried out
Posture	1) Initial Posture period (Baseline) 2) Task development period	Pressure level		Compare the initial user posture during the task development against their baseline posture	It is possible for mat pressures using the BPMS to calculate the average pressure (μ), the average coverage (C), and the average coverage change in a pressure mat ($a_{coverage}$) [107] When the user tends to lean forward, exerting higher pressure on the seat, they are engaged	$\mu = \frac{1}{R \times C} \sum_{i=1}^{R} \sum_{j=1}^{c} P_{ij}$ $C = \frac{1}{R \times C} \sum_{i=1}^{R} \sum_{j=1}^{c} X_{ij}$ $a_{coverage} = \frac{1}{N} \sum_{i=1}^{N} |C_t - C_{t+1}|$ Where R refers to the number of rows in the pressure mat, C the number of columns, and P_{ij} is the pressure of a sensing element in row i and column j $X_{ij} = 1, \, if \, Pt_{ij} > 1$ $X_{ij} = 0, \, if \, Pt_{ij} = 0$ Taken from D'Mello and Graesser [107]
Respiratory Rate	1) Rest Respiratory Rate (Baseline) 2) Task development period	Breaths per unit time		Compare the Respiratory user rate (RR) during the task development against their baseline rate	During mental stress, difficulty, or excitement, a random breathing variability (RR) tends to increase and tends to decrease with the presence of task-related attention	$RR = \frac{60}{T_{otalbreathtime}}$ $MV = \frac{RR}{VT}$ Respiratory volume (Vi), Respiration Rate (RR), Minute ventilation (MV) Taken from Vlemincx et al. [116]
Facial Expression	1) Task development period	Depth map in output from the 3D sensor		Using the trained Convolutional neural network (CNN) with a facial dataset is possible to classify the engagement level according to a defined scale	*Classes* / *Engagement level* / *Expressions* **Deactivation** / Low / Relaxed, Calm **Average** / Medium / Happy, Contented, Serene, elated **Activation** / High / Alert, Excitement Engagement scale, according to Nonis et al. [121], based on Russell's circumplex model of affect	

Table 8 Conceptual framework content

Framework reference	Method
Self-report UE measurements (Table 4)	Questionnaire
	Survey
	Interview
UE Dimensionality sub-constructs (Table 5)	Uni-dimensionality
	Multi-dimensionality
Physiological methods Guideline (Table 7)	Skin conductance
	Heart rate
	Electroencephalography
	Pupillometry
	Posture analysis
	Respiratory rate
	Facial expression

definition of the UE through its dimensionalities, and the last one focused on the physiological methods.

The findings verify that the UE is defined and measured differently depending on the application context; yet, getting to know better the engagement level of a user with the product under analysis can give valuable information that could be used as feedback for the design process, enhancing the chances of success of a product. In many aspects, the focus and extent of the analysis performed will create a product that users wish to interact with [68]. For that reason, we have included examples of good practices in this framework.

This study identified several authors who have conducted significant studies on self-report methods and their characteristics. For future studies on this topic, we see opportunities for research related to identifying and standardizing a methodology based on UED sub-constructs to complement the development of self-report instruments.

Referring to physiological methods, in the literature, we found different physiological approaches for conducting the UE measurement; in the framework, we have included the most commonly used. From the revision, we concluded that there are no standardized scales since the measurements are highly dependent on the conditions under which the studies are carried out. Consequently, the standardized scales for some physiological methods such as heart rate, pupillometry, posture, and respiratory rate are still a research gap. A common scale could be beneficial, facilitating the measurement and guaranteeing a clear understanding of the results. In this sense, there is an opportunity for future studies interested in this topic that wish to focus on less-used physiological methods for measuring the UE.

Most existing studies on UE focus their goals on determining the engagement value using a parameter explicitly created for their research. In this way, finding lessons learned and guidance from other studies could become difficult due to the lack of standardized methods. The available literature review documents analyzed a particular UE domain and referred to a specific application context. For example, in the study conducted by Pontes et al. [142], the authors made a systematic literature review of the instruments used to measure political engagement. The Handbook of Research on Customer Engagement [143] is a good source of information regarding exclusively marketing practices and organizational performance. Kulikowski [144] explored the literature on work engagement as a predictor of health. Furthermore, Bazzani et al. [145] presented an overview of EEG application in consumer neuroscience, briefly discussing the drivers of engagement.

Some evaluated studies concluded the need to adopt physiological approaches to assess engagement. Current studies on UE do not necessarily focus exclusively on one field; thus, considering this framework as a starting point for further research can give a global vision of the most commonly used methods in different fields and help the reader delimit their research area.

The best method to use will depend on many factors; so far, the previous studies are making considerable progress on physiological technics, analyzing the physical user responses to a stimulus, and obtaining a reliable engagement value. These findings demonstrate that the inclusion of physiological characteristics provides information that allows the researcher to determine whether the findings from self-reports are trustworthy and representative; however, the sample size is much must lower compared to traditional self-report methods, limiting the generalization of the results.

On the other hand, self-reported methods have been in the UE context for more time, and it is usual to find validated instruments for specific fields. It is also essential to consider resource availability, budget, time, and experience. In the case of physiological methods, we realize that the equipment needed is much more expensive than self-reports, and the researchers must follow a considerable training period to collect and evaluate reliable data. Fortunately, recording equipment is getting relatively cheaper and more user-friendly; that is the case, for example, of the new EEG headsets available on the market.

The findings from this framework suggest the need to apply a mixed UE measurement technic using self-report and, whenever possible, a physiological method to validate the results.

In addition, there is a need for an agreed conceptualization of UE measurement; future studies from the same contexts should try to address their finding as an input to create a standardized UE measure. The information in this framework could then be used as the basis to explore already validated methods and, based on them, develop reliable and standardized UE measurements.

Through a review of the different methodologies, the framework developed in this paper provides a comprehensive summary of current evidence obtained through an objective and transparent approach to minimize bias without preference and favoritism. This study is a valuable help for other researchers and practitioners involved in UE and can be used as a guide to quickly choose an alternative from all those present in the literature.

5 Conclusions and discussion

Our research question in this study was whether in the literature exits a common scale/metric for measuring the user engagement. To address this question this conceptual framework aimed to identify a range of studies that met the criteria and reported engagement measurements to extract conclusions about the methods used. A comprehensive search was conducted to determine as many studies as possible, which included Scopus DB and Google Academics but without excluding proceeding documents, to identify further prominent studies that may not have been wholly published yet. We only included articles that reported a precise engagement measure.

The number of articles and citations regarding UE studies has been increasing in recent years. As stated before, many disciplinary fields are interested in engagement measurement. Indicatively, in our study, we found that the sector that has carried out the most research on this topic is the Education field which represents 27% of the total documents analyzed; lately, interest has grown regarding student engagement during online lessons. The following field by the number of studies (20%) is health services and their interest in patient engagement during different treatments. Followed by engineering and manufacturing context (19%), this field includes studies related to new technology developments, product design processes, and workers' engagement while developing a specific task. Marketing-related studies represent 15%. Virtual reality applications the 8%. Social services the 4%. Videogames 3%, while other services interested in user engagement and their perceptions like music, sports or food represent the remaining 4%.

For carrying out the literature review, an initial search was done, then the representative data was selected and filtered according to the scope of the research to create clusters based on data classification. After the review, a total of 1479 documents were analyzed, screened, and selected to understand the methods, tools/instruments, and metrics to measure the UE. We found that many studies described the use of self-report methods, while others used physiological methods.

Based on the results, we classified the methods to measure the UE with a product into clusters. The previous revision is gathered within a framework that organizes the current methodologies as Self-report and Physiological methods. The self-report methods refer to measurement instruments that evaluate engagement from different perspectives. This category included the UE dimensionality (UED) as a way to compare the instruments [16], allowing a comparison between sub-constructs that defines the interaction level between user and product. It was possible to identify a wide range of dimensionalities considered as the basis of a self-report instrument. However, according to the findings, the multi-dimensionality that better supports the UE analysis includes the emotional, behavioral, and cognitive sub-constructs.

There is a higher prevalence of self-report methods, opening new possibilities for the incursion of physiological methods to identify in real-time physical metrics and establish if a user is more or less engaged. Many of the studies used only self-report methods to measure UE, (despite the reported limitations of self-report [139]), and used engagement assessments specific to one context; therefore, generalizability is limited. However, in the last years, there has been an increase in the use of physiological methods.

We explored seven physiological approach methods, finding similarities in the process's measurement phases between the different approaches. These phases allow comparing a standard data against its variation, usually defined as baseline period, task development period, and in some cases, a resting task period. On the other hand, for some methods (Heart rate, Pupillometry, Posture, Respiratory rate), there is not a numerical scale that classifies whether or not the user is engaged; however, there are insights to facilitate the understanding of the information regarding the UE according to the method used. While for other methods (Skin conductance, Electroencephalography, Facial expressions), indices and classification levels define the UE more clearly and concisely. Furthermore, considering these metrics as an additional source of information about the user requirements and desires, it is possible to conclude that the physiological methods still have a wide field to be explored concerning UE.

In this context, we consider user engagement a handy indicator in any research field, including interactive design and manufacturing. The interdisciplinarity aimed by UCD methods and the lessons learned from other areas will provide visions that will enhance the current design processes. For that reason, this study described different scenarios, including marketing, health, and education, that can be supportive case studies for future UE applications.

This conceptual framework has analyzed the most significant number of publications available and shows examples of good practices from different fields and domains for both self-report and physiological methods. Additionally provides, whenever possible, the measurement scale and equation used to calculate the UE.

Finally, we recognize the challenge of selecting a method for measuring UE since there is a wide range of possibilities between the various existing methods. In this regard, the results of this study can allow future researchers, developers, or designers to consider the UE as one of the most prominent indicators of the product or service success and to use this general guideline as support to find the more appropriate method, technique, and metric for measuring the UE based on the requirements of their studies.

Declarations

Conflict of interest The authors report no conflict of interest.

References

1. O'Brien, H., Cairns, P. (eds.): Why Engagement Matters. Springer International Publishing, Cham (2016)
2. Di Lascio, E., Gashi, S., Santini, S.: Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **2**, 1–21 (2018). https://doi.org/10.1145/3264913
3. Scholz, J., Smith, A.N.: Augmented reality: designing immersive experiences that maximize consumer engagement. Bus. Horiz. **59**, 149–161 (2016). https://doi.org/10.1016/j.bushor.2015.10.003
4. Pagulayan, R.J., Keeker, K., Fuller, T., Wixon, D., Romero, R.L., Gunn, D.V.: User-Centered Design in Games, pp. 795–821. CRC Press, London (2012)
5. Savazzi, F., Isernia, S., Jonsdottir, J., Di Tella, S., Pazzi, S., Baglio, F.: Engaged in learning neurorehabilitation: development and validation of a serious game with user-centered design. Comput. Educ. **125**, 53–61 (2018). https://doi.org/10.1016/j.compedu.2018.06.001
6. Sturm, U., Tscholl, M.: The role of digital user feedback in a user-centred development process in citizen science. J. Sci. Commun. JCOM (2019). https://doi.org/10.22323/2.18010203
7. Calder, B.J., Malthouse, E.C., Maslowska, E.: Brand marketing, big data and social innovation as future research directions for engagement. J. Mark. Manag. **32**, 579–585 (2016). https://doi.org/10.1080/0267257X.2016.1144326
8. Lalmas, M., O'Brien, H., Yom-Tov, E.: Measuring user engagement. Synth. Lect. Inf. Concepts Retrieval Services. **6**, 1–132 (2014). https://doi.org/10.2200/s00605ed1v01y201410icr038
9. McNeal, K.S., Spry, J.M., Mitra, R., Tipton, J.L.: Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course. J. Geosci. Educ. **62**, 655–667 (2014). https://doi.org/10.5408/13-111.1
10. Darnell, D.K., Krieg, P.A.: Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures. PLoS ONE **14**, e0225709 (2019). https://doi.org/10.1371/journal.pone.0225709
11. McMahan, T., Parberry, I., Parsons, T.D.: Evaluating player task engagement and arousal using electroencephalography. Proc. Manuf. **3**, 2303–2310 (2015). https://doi.org/10.1016/j.promfg.2015.07.376
12. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biol. Psychol. **40**, 187–195 (1995). https://doi.org/10.1016/0301-0511(95)05116-3
13. Schimanski, C.P., Pasetti, G., Matt, C.: Pushing digital automation of configure-to-order services in small and medium enterprises of the construction equipment industry: a design science research approach. Appl. Sci. **9**, 3780 (2019). https://doi.org/10.3390/app9183780
14. Dresch, A., Lacerda, D.P., Antunes, J.A.V.: Design Science Research: A Method for Science and Technology Advancement. Springer International Publishing, Cham (2015)
15. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information. Syst. Res. **28**, 75–105 (2004)
16. Fredricks, J.A., McColskey, W.: The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In: Christenson, S.L., Reschly, A.L., Wylie, C. (eds.) Handbook of Research on Student Engagement, pp. 763–782. Springer, US, Boston, MA (2012)
17. Henrie, C.R., Halverson, L.R., Graham, C.R.: Measuring student engagement in technology-mediated learning: a review. Comput. Educ. **90**, 36–53 (2015). https://doi.org/10.1016/j.compedu.2015.09.005
18. O'Brien, H.L., Toms, E.G.: The development and evaluation of a survey to measure user engagement. J. Am. Soc. Inf. Sci. **61**, 50–69 (2010). https://doi.org/10.1002/asi.21229
19. Leung, D.Y.P., Kember, D.: Comparability of data gathered from evaluation questionnaires on paper and through the internet. Res High Educ. **46**, 571–591 (2005). https://doi.org/10.1007/s11162-005-3365-3
20. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The development of the game engagement questionnaire: a measure of engagement in video game-playing. J. Exp. Soc. Psychol. **45**, 624–634 (2009). https://doi.org/10.1016/j.jesp.2009.02.016
21. O'Sullivan, T.L., Corneil, W., Kuziemsky, C.E., Toal-Sullivan, D.: Use of the structured interview matrix to enhance community resilience through collaboration and inclusive engagement: the structured interview matrix tool for collaboration and engagement. Syst. Res. **32**, 616–628 (2015). https://doi.org/10.1002/sres.2250
22. Yeung, A., Trinh, N.-H.T., Chang, T.E., Fava, M.: The engagement interview protocol (EIP): improving the acceptance of mental health treatment among Chinese immigrants. Int. J. Cult. Ment. Health **4**, 91–105 (2011). https://doi.org/10.1080/17542863.2010.507933

23. Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., Huggins, J.: Validation of a comprehensive stress state questionnaire: towards a state "Big Three." 16 (1999)

24. Helton, W.S.: Validation of a short stress state questionnaire. In: 5th annual meeting. (2004)

25. Herrin, J., Harris, K.G., Kenward, K., Hines, S., Joshi, M.S., Frosch, D.L.: Patient and family engagement: a survey of US hospital practices. BMJ Qual Saf. **25**, 182–189 (2016). https://doi.org/10.1136/bmjqs-2015-004006

26. Shirom, A.: Feeling vigorous at work? The construct of vigor and the study of positive affect in organizations. In: Research in occupational stress and well-being. pp. 135–164. Emerald (MCB UP), Bingley (2003)

27. Schaufeli, W.B., Bakker, A.B., Salanova, M.: The measurement of work engagement with a short questionnaire: a cross-national study. Educ. Psychol. Measur. **66**, 701–716 (2006). https://doi.org/10.1177/0013164405282471

28. Yarosh, S., Markopoulos, P., Abowd, G.D.: Towards a questionnaire for measuring affective benefits and costs of communication technologies. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14. pp. 84–96. ACM Press, Baltimore, Maryland, USA (2014)

29. Hart, S.R., Stewart, K., Jimerson, S.R.: The student engagement in schools questionnaire (SESQ) and the teacher engagement report form-new (TERF-N): examining the preliminary evidence. Contemp. Sch. Psychol. **15**, 13 (2011)

30. Hannum, M.E., Simons, C.T.: Development of the engagement questionnaire (EQ): a tool to measure panelist engagement during sensory and consumer evaluations. Food Qual. Prefer. **81**, 103840 (2020). https://doi.org/10.1016/j.foodqual.2019.103840

31. Chin, T., Rickard, N.S.: The music Use (MUSE) questionnaire: an instrument to measure engagement in music. Music Percept. **29**, 429–446 (2012). https://doi.org/10.1525/mp.2012.29.4.429

32. Vanstone, A.D., Wolf, M., Poon, T., Cuddy, L.L.: Measuring engagement with music: development of an informant-report questionnaire. Aging Ment. Health **20**, 474–484 (2016). https://doi.org/10.1080/13607863.2015.1021750

33. Handelsman, M.M., Briggs, W.L., Sullivan, N., Towler, A.: A measure of college student course engagement. J. Educ. Res. **98**, 184–192 (2005). https://doi.org/10.3200/JOER.98.3.184-192

34. Phan, M.H., Keebler, J.R., Chaparro, B.S.: The development and validation of the game user experience satisfaction scale (GUESS). Hum Factors. **58**, 1217–1247 (2016). https://doi.org/10.1177/0018720816669646

35. Brodie, R.J., Hollebeek, L.D., Jurić, B., Ilić, A.: Customer engagement: conceptual domain, fundamental propositions, and implications for research. J. Serv. Res. **14**, 252–271 (2011). https://doi.org/10.1177/1094670511411703

36. Dessart, L., Veloutsou, C., Morgan-Thomas, A.: Capturing consumer engagement: duality, dimensionality and measurement. J. Mark. Manag. **32**, 399–426 (2016). https://doi.org/10.1080/0267257X.2015.1130738

37. Kuvykaitė, R., Tarutė, A.: A critical analysis of consumer engagement dimensionality. Procedia. Soc. Behav. Sci. **213**, 654–658 (2015). https://doi.org/10.1016/j.sbspro.2015.11.468

38. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School engagement: potential of the concept, state of the evidence. Rev. Educ. Res. **74**, 59–109 (2004). https://doi.org/10.3102/00346543074001059

39. Hollebeek, L.: Exploring customer brand engagement: definition and themes. J. Strateg. Mark. **19**, 555–573 (2011). https://doi.org/10.1080/0965254X.2011.599493

40. Achterberg, W., Pot, A.M., Kerkstra, A., Ooms, M., Muller, M., Ribbe, M.: The effect of depression on social engagement in newly admitted Dutch nursing home residents. Gerontologist **43**, 213–218 (2003). https://doi.org/10.1093/geront/43.2.213

41. Jennings, M.K., Stoker, L.: Social trust and civic engagement across time and generations. Acta Polit. **39**, 342–379 (2004). https://doi.org/10.1057/palgrave.ap.5500077

42. Algesheimer, R., Dholakia, U.M., Herrmann, A.: The social influence of brand community: evidence from European car clubs. J. Mark. **69**, 19–34 (2005). https://doi.org/10.1509/jmkg.69.3.19.66363

43. Whelan, S., Wohlfeil, M.: Communicating brands through engagement with 'lived' experiences. J Brand Manag. **13**, 313–329 (2006). https://doi.org/10.1057/palgrave.bm.2540274

44. London, B., Downey, G., Mace, S.: Psychological theories of educational engagement: a multi-method approach to studying individual engagement and institutional change. Vanderbilt Law Rev. **60**, 28 (2007)

45. Bejerholm, U., Eklund, M.: Occupational engagement in persons with Schizophrenia: relationships to self-related variables, psychopathology, and quality of life. Am. J. Occup. Ther. **61**, 21–32 (2007). https://doi.org/10.5014/ajot.61.1.21

46. Bowden, J.L.-H.: The process of customer engagement: a conceptual framework. J. Market. Theory Pract. **17**, 63–74 (2009). https://doi.org/10.2753/MTP1069-6679170105

47. Higgins, E.T., Scholer, A.A.: Engaging the consumer: the science and art of the value creation process. J. Consum. Psychol. **19**, 100–114 (2009). https://doi.org/10.1016/j.jcps.2009.02.002

48. Calder, B.J., Malthouse, E.C., Schaedel, U.: An experimental study of the relationship between online engagement and advertising effectiveness. J. Interact. Mark. **23**, 321–331 (2009). https://doi.org/10.1016/j.intmar.2009.07.002

49. Sprott, D., Czellar, S., Spangenberg, E.: The importance of a general measure of brand engagement on market behavior: development and validation of a scale. J. Mark. Res. **46**, 92–104 (2009). https://doi.org/10.1509/jmkr.46.1.92

50. Gambetti, R.C., Graffigna, G.: The concept of engagement: a systematic analysis of the ongoing marketing debate. Int. J. Mark. Res. **52**, 801–826 (2010). https://doi.org/10.2501/S147078531020166

51. Mollen, A., Wilson, H.: Engagement, telepresence and interactivity in online consumer experience: reconciling scholastic and managerial perspectives. J. Bus. Res. **63**, 919–925 (2010). https://doi.org/10.1016/j.jbusres.2009.05.014

52. van Doorn, J., Lemon, K.N., Mittal, V., Nass, S., Pick, D., Pirner, P., Verhoef, P.C.: Customer engagement behavior: theoretical foundations and research directions. J. Serv. Res. **13**, 253–266 (2010). https://doi.org/10.1177/1094670510375599

53. Phillips, B.J., McQuarrie, E.F.: Narrative and persuasion in fashion advertising. J Consum Res. **37**, 368–392 (2010). https://doi.org/10.1086/653087

54. Verhoef, P.C., Reinartz, W.J., Krafft, M.: Customer engagement as a new perspective in customer management. J. Serv. Res. **13**, 247–252 (2010). https://doi.org/10.1177/1094670510375461

55. Abdul-Ghani, E., Hyde, K.F., Marshall, R.: Emic and etic interpretations of engagement with a consumer-to-consumer online auction site. J. Bus. Res. **64**, 1060–1066 (2011). https://doi.org/10.1016/j.jbusres.2010.10.009

56. Wang, M.-T., Willett, J.B., Eccles, J.S.: The assessment of school engagement: examining dimensionality and measurement invariance by gender and race/ethnicity. J. Sch. Psychol. **49**, 465–480 (2011). https://doi.org/10.1016/j.jsp.2011.04.001

57. Vivek, S.D., Beatty, S.E., Morgan, R.M.: Customer engagement: exploring customer relationships beyond purchase. J. Market. Theory Pract. **20**, 122–146 (2012). https://doi.org/10.2753/MTP1069-6679200201

58. Gambetti, R.C., Graffigna, G., Biraghi, S.: The grounded theory approach to consumer-brand engagement: the practitioner's standpoint. Int. J. Mark. Res. **54**, 659–687 (2012). https://doi.org/10.2501/IJMR-54-5-659-687

59. Gummerus, J., Liljander, V., Weman, E., Pihlström, M.: Customer engagement in a Facebook brand community. Manag. Res. Rev. **35**, 857–877 (2012). https://doi.org/10.1108/01409171211256578

60. Cambra, J., Melero, I., Sese, F.J.: Aproximación al concepto de engagement: un estudio exploratorio en el sector de la telefonía móvil. Univ. Bus. Rev. (2012)

61. Vivek, S.D., Beatty, S.E., Dalela, V., Morgan, R.M.: A generalized multidimensional scale for measuring customer engagement. J. Market. Theory Pract. **22**, 401–420 (2014). https://doi.org/10.2753/MTP1069-6679220404

62. Hollebeek, D.L., Chen, T.: Exploring positively- versus negatively-valenced brand engagement: a conceptual model. J. Product Brand Mgt. **23**, 62–74 (2014). https://doi.org/10.1108/JPBM-06-2013-0332

63. Hollebeek, L.D., Glynn, M.S., Brodie, R.J.: Consumer brand engagement in social media: conceptualization, scale development and validation. J. Interact. Mark. **28**, 149–165 (2014). https://doi.org/10.1016/j.intmar.2013.12.002

64. Baldus, B.J., Voorhees, C., Calantone, R.: Online brand community engagement: scale development and validation. J. Bus. Res. **68**, 978–985 (2015). https://doi.org/10.1016/j.jbusres.2014.09.035

65. Hollebeek, L.D., Malthouse, E.C., Block, M.P.: Sounds of music: exploring consumers' musical engagement. J. Consum. Mark. **33**, 417–427 (2016). https://doi.org/10.1108/JCM-02-2016-1730

66. Violante, M.G., Vezzetti, E., Piazzolla, P.: How to design a virtual reality experience that impacts the consumer engagement: the case of the virtual supermarket. Int. J. Interact. Des. Manuf. **13**, 243–262 (2019). https://doi.org/10.1007/s12008-018-00528-5

67. Furbee, M.R.: Actions speak louder than words: a study in body language. 10 (2017)

68. Veryzer, R.W., de Borja, M.B.: The impact of user-oriented design on new product development: an examination of fundamental relationships. J. Product. Innov. Manag. **22**, 128–143 (2005). https://doi.org/10.1111/j.0737-6782.2005.00110.x

69. Morrison, A.L., Rozak, S., Gold, A.U., Kay, J.E.: Quantifying student engagement in learning about climate change using galvanic hand sensors in a controlled educational setting. Clim. Change **159**, 17–36 (2019). https://doi.org/10.1007/s10584-019-02576-6

70. Poh, M.-Z., Swenson, N.C., Picard, R.W.: A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE Trans. Biomed. Eng. **57**, 1243–1252 (2010). https://doi.org/10.1109/TBME.2009.2038487

71. Pecchinenda, A., Smith, C.A.: The affective significance of skin conductance activity during a difficult problem-solving task. Cogn. Emot. **10**, 481–504 (1996). https://doi.org/10.1080/026999396380123

72. Wang, C., Cesar, P.: Measuring audience responses of video advertisements using physiological sensors. In: Proceedings of the 3rd international workshop on immersive media experiences - ImmersiveME '15. pp. 37–40. ACM Press, Brisbane, Australia (2015)

73. Gregersen, A., Langkjær, B., Heiselberg, L., Wieland, J.L.: Following the viewers: investigating television drama engagement through skin conductance measurements. Poetics **64**, 1–13 (2017). https://doi.org/10.1016/j.poetic.2017.06.002

74. Mandryk, R.L., Atkins, M.S.: A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int. J. Hum Comput Stud. **65**, 329–347 (2007). https://doi.org/10.1016/j.ijhcs.2006.11.011

75. Villanueva, I., Campbell, B.D., Raikes, A.C., Jones, S.H., Putney, L.G.: A multimodal exploration of engineering students' emotions and electrodermal activity in design activities. J. Eng. Educ. **107**, 414–441 (2018). https://doi.org/10.1002/jee.20225

76. Nourbakhsh, N., Chen, F., Wang, Y., Calvo, R.A.: Detecting users' cognitive load by galvanic skin response with affective interference. ACM Trans. Interact. Intell. Syst. **7**, 1–20 (2017). https://doi.org/10.1145/2960413

77. Pendleton, D.M., Sakalik, M.L., Moore, M.L., Tomporowski, P.D.: Mental engagement during cognitive and psychomotor tasks: effects of task type, processing demands, and practice. Int. J. Psychophysiol. **109**, 124–131 (2016). https://doi.org/10.1016/j.ijpsycho.2016.08.012

78. Richardson, D.C., Griffin, N.K., Zaki, L., Stephenson, A., Yan, J., Hogan, J., Skipper, J.I., Devlin, J.T.: Measuring narrative engagement: the heart tells the story. Neuroscience **29**(2), 278 (2018)

79. Scholey, A.B., Moss, M.C., Neave, N., Wesnes, K.: Cognitive performance, hyperoxia, and heart rate following oxygen administration in healthy young adults. Physiol. Behav. **67**, 783–789 (1999). https://doi.org/10.1016/S0031-9384(99)00183-3

80. Monkaresi, H., Bosch, N., Calvo, R.A., D'Mello, S.K.: Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans. Affective Comput. **8**, 15–28 (2017). https://doi.org/10.1109/TAFFC.2016.2515084

81. Stikic, M., Berka, C., Levendowski, D.J., Rubio, R.F., Tan, V., Korszen, S., Barba, D., Wurzer, D.: Modeling temporal sequences of cognitive state changes based on a combination of EEG-engagement, EEG-workload, and heart rate metrics. Front. Neurosci. (2014). https://doi.org/10.3389/fnins.2014.00342

82. Nearing, B.D., Libbus, I., Amurthur, B., Kenknight, B.H., Verrier, R.L.: Acute autonomic engagement assessed by heart rate dynamics during vagus nerve stimulation in patients with heart failure in the ANTHEM-HF trial: acute heart rate dynamics during VNS. J. Cardiovasc. Electrophysiol. **27**, 1072–1077 (2016). https://doi.org/10.1111/jce.13017

83. Poh, M.-Z., McDuff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt. Express. **18**, 10762 (2010). https://doi.org/10.1364/OE.18.010762

84. Chaouachi, M., Frasson, C.: Exploring the Relationship between Learner EEG Mental Engagement and Affect. In: Aleven, V., Kay, J., Mostow, J. (eds.) Intelligent Tutoring Systems, pp. 291–293. Springer, Berlin (2010)

85. Freeman, F.G., Mikulka, P.J., Prinzel, L.J., Scerbo, M.W.: Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. Biol. Psychol. **50**, 61–76 (1999). https://doi.org/10.1016/S0301-0511(99)00002-2

86. Eldenfria, A., Al-Samarraie, H.: Towards an online continuous adaptation mechanism (OCAM) for enhanced engagement: an EEG study. Int. J. Hum. Comput. Interact. **35**, 1960–1974 (2019). https://doi.org/10.1080/10447318.2019.1595303

87. Nuamah, J.K., Seong, Y.: Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. Brain-Comput. Interfaces. **5**, 1–12 (2017). https://doi.org/10.1080/2326263X.2017.1338012

88. Kang, D., Kim, J., Jang, D.-P., Cho, Y.S., Kim, S.-P.: Investigation of engagement of viewers in movie trailers using electroencephalography. Brain-Comput. Interfaces. **2**, 193–201 (2015). https://doi.org/10.1080/2326263X.2015.1103591

89. Yamada, F.: Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics. Ergonomics **41**, 678–688 (1998). https://doi.org/10.1080/001401398186847

90. Dehais, F., Roy, R.N., Durantin, G., Gateau, T., Callan, D.: EEG-Engagement Index and Auditory Alarm Misperception: An Inattentional Deafness Study in Actual Flight Condition. In:

Baldwin, C. (ed.) Advances in Neuroergonomics and Cognitive Engineering, pp. 227–234. Springer International Publishing, Cham (2018)

91. Rabbi, A.F., Zony, A., de Leon, P., Fazel-Rezai, R.: Mental workload and task engagement evaluation based on changes in electroencephalogram. Biomed. Eng. Lett. **2**, 139–146 (2012). https://doi.org/10.1007/s13534-012-0065-8

92. Ales, F., Giromini, L., Zennaro, A.: Complexity and cognitive engagement in the rorschach task: an eye-tracking study. J. Personal. Assess. (2019). https://doi.org/10.1080/00223891.2019.1575227

93. Hopstaken, J.F., van der Linden, D., Bakker, A.B., Kompier, M.A.J.: A multifaceted investigation of the link between mental fatigue and task disengagement: mental fatigue and task disengagement. Psychophysiol. **52**, 305–315 (2015). https://doi.org/10.1111/psyp.12339

94. Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J.: The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology **45**, 602–607 (2008). https://doi.org/10.1111/j.1469-8986.2008.00654.x

95. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. 17 (1982)

96. Binda, P., Pereverzeva, M., Murray, S.O.: Pupil size reflects the focus of feature-based attention. J. Neurophysiol. **112**, 3046–3052 (2014). https://doi.org/10.1152/jn.00502.2014

97. van den Brink, R.L., Murphy, P.R., Nieuwenhuis, S.: Pupil diameter tracks lapses of attention. PLoS ONE **11**, e0165274 (2016). https://doi.org/10.1371/journal.pone.0165274

98. Gilzenrat, M.S., Nieuwenhuis, S., Jepma, M., Cohen, J.D.: Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn. Affect. Behav. Neurosci. **10**, 252–269 (2010). https://doi.org/10.3758/CABN.10.2.252

99. Jepma, M., Nieuwenhuis, S.: Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. J. Cogn. Neurosci. **23**, 1587–1596 (2011). https://doi.org/10.1162/jocn.2010.21548

100. Hopstaken, J.F., van der Linden, D., Bakker, A.B., Kompier, M.A.J.: The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. Biol. Psychol. **110**, 100–106 (2015). https://doi.org/10.1016/j.biopsycho.2015.06.013

101. Murphy, P.R., O'Connell, R.G., O'Sullivan, M., Robertson, I.H., Balsters, J.H.: Pupil diameter covaries with BOLD activity in human locus coeruleus. Hum. Brain Mapp. **35**, 4140–4154 (2014). https://doi.org/10.1002/hbm.22466

102. Kassem, K., Salah, J., Abdrabou, Y., Morsy, M., El-Gendy, R., Abdelrahman, Y., Abdennadher, S.: DiVA: exploring the usage of pupil diameter to elicit valence and arousal. In: Proceedings of the 16th international conference on mobile and ubiquitous multimedia - MUM '17. pp. 273–278. ACM Press, Stuttgart, Germany (2017)

103. Skulmowski, A., Bunge, A., Kaspar, K., Pipa, G.: Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. Front. Behav. Neurosci. (2014). https://doi.org/10.3389/fnbeh.2014.00426

104. Podder, P.K., Paul, M., Debnath, T., Murshed, M.: An analysis of human engagement behaviour using descriptors from human feedback, eye tracking, and saliency modelling. In: 2015 International conference on digital image computing: techniques and applications (DICTA). pp. 1–8. IEEE, Adelaide, Australia (2015)

105. Mota, S., Picard, R.W.: Automated posture analysis for detecting learner's interest level. In: 2003 conference on computer vision and pattern recognition workshop. pp. 49–49. IEEE, Madison, Wisconsin, USA (2003)

106. Bibbo, D., Carli, M., Conforto, S., Battisti, F.: A sitting posture monitoring instrument to assess different levels of cognitive engagement. Sensors. **19**, 455 (2019). https://doi.org/10.3390/s19030455

107. D'Mello, S., Graesser, A.: Automatic detection of learner's affect from gross body language. Appl. Artif. Intell. **23**, 123–150 (2009). https://doi.org/10.1080/08839510802631745

108. Shirehjini, A.A.N., Yassine, A., Shirmohammadi, S.: Design and implementation of a system for body posture recognition. Multimed. Tools Appl. **70**, 1637–1650 (2014). https://doi.org/10.1007/s11042-012-1137-6

109. Graesser, A., McNamara, D.: Self-regulated learning in learning environments with pedagogical agents that interact in natural language. Educ. Psychol. **45**, 234–244 (2010). https://doi.org/10.1080/00461520.2010.515933

110. Boiten, F.A., Frijda, N.H., Wientjes, C.J.E.: Emotions and respiratory patterns: review and critical analysis. Int. J. Psychophysiol. **17**, 103–128 (1994). https://doi.org/10.1016/0167-8760(94)90027-2

111. Vlemincx, E., Van Diest, I., Van den Bergh, O.: A sigh following sustained attention and mental stress: effects on respiratory variability. Physiol. Behav. **107**, 1–6 (2012). https://doi.org/10.1016/j.physbeh.2012.05.013

112. Porges, S.W., Coles, M.G.H.: Individual differences in respiratory-heart period coupling and heart period responses during two attention-demanding tasks. Psychobiology **10**, 215–220 (1982). https://doi.org/10.3758/BF03332939

113. Webster, K.E., Colrain, I.M.: The respiratory-related evoked potential: effects of attention and occlusion duration. Psychophysiology **37**, 310–318 (2000). https://doi.org/10.1111/1469-8986.3730310

114. Wenger, M.J., DellaValle, D.M., Murray-Kolb, L.E., Haas, J.D.: Effect of iron deficiency on simultaneous measures of behavior, brain activity, and energy expenditure in the performance of a cognitive task. Nutr. Neurosci. **22**, 196–206 (2017). https://doi.org/10.1080/1028415X.2017.1360559

115. Gomez, P., Zimmermann, P., Guttormsen-Schär, S., Danuser, B.: Respiratory responses associated with affective processing of film stimuli. Biol. Psychol. **68**, 223–235 (2005). https://doi.org/10.1016/j.biopsycho.2004.06.003

116. Vlemincx, E., Taelman, J., De Peuter, S., Van Diest, I., Van Den Bergh, O.: Sigh rate and respiratory variability during mental load and sustained attention: sigh rate and respiratory variability. Psychophysiology **48**, 117–120 (2011). https://doi.org/10.1111/j.1469-8986.2010.01043.x

117. Vlemincx, E., Luminet, O.: Sighs can become learned behaviors via operant learning. Biol. Psychol. **151**, 107850 (2020). https://doi.org/10.1016/j.biopsycho.2020.107850

118. Jetha, M.K., Segalowitz, S.J.: Social and emotional development. In: Adolescent brain development. pp. 25–52. Elsevier (2012)

119. Nezami, O.M., Dras, M., Hamey, L., Richards, D., Wan, S., Paris, C.: Automatic recognition of student engagement using deep learning and facial expression. arXiv:1808.02324 [cs]. (2019)

120. Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., Movellan, J.R.: The faces of engagement: automatic recognition of student engagement from facial expressions. IEEE Trans. Affective Comput. **5**, 86–98 (2014). https://doi.org/10.1109/TAFFC.2014.2316163

121. Nonis, F., Olivetti, E.C., Marcolin, F., Violante, M.G., Vezzetti, E., Moos, S.: Questionnaires or inner feelings: who measures the engagement better? Appl. Sci. **10**, 609 (2020). https://doi.org/10.3390/app10020609

122. Zhang, Z., Luo, P., Loy, C.-C., Tang, X.: Learning social relation traits from face images. In: 2015 IEEE international conference on computer vision (ICCV). pp. 3631–3639. IEEE, Santiago, Chile (2015)

123. Ramya, R., Mala, K., Sindhuja, C.: Student engagement identification based on facial expression analysis using 3D video/image of students. TAGA J. **14**, 2446 (2018)

124. Rajavenkatanarayanan, A., Babu, A.R., Tsiakas, K., Makedon, F.: Monitoring task engagement using facial expressions and body postures. In: Proceedings of the 3rd international workshop on interactive and spatial computing - IWISC '18. pp. 103–108. ACM Press, Richardson, Texas (2018)

125. Olivetti, E., Violante, M.G., Vezzetti, E., Marcolin, F., Eynard, B.: Engagement evaluation in a virtual learning environment via facial expression recognition and self-reports: a preliminary approach. Appl. Sci. **10**, 314 (2019). https://doi.org/10.3390/app10010314

126. Violante, M.G., Marcolin, F., Vezzetti, E., Ulrich, L., Billia, G., Di Grazia, L.: 3D facial expression recognition for defining users' inner requirements—an emotional design case study. Appl. Sci. **9**, 2218 (2019). https://doi.org/10.3390/app9112218

127. NASA- National Aeronautics and Space Administration: NASA TLX Task Load Index, Nasa, https://humansystems.arc.nasa.gov/groups/tlx/index.php. (2019). Accessed 18 January (2020)

128. Rogers, Kara: Electroencephalography, Encyclopædia Britannica Inc, https://www.britannica.com/science/electroencephalography. (2017). Accessed 6 January 2020

129. Gholipur, Bahar: What is a Normal Heart Rate?", Live-Science , https://www.livescience.com/42081-normal-heart-rate.html. (2018). Accessed 20 December 2019

130. Lonsdale, C., Hodge, K., Jackson, S.A.: Athlete engagement: II development and initial validation of the athlete engagement questionnaire. Int. J. Sports Psychol. **38**(4), 451–470 (2007)

131. Lykken, D.T., Venables, P.H.: Direct measurement of skin conductance: a proposal for standardization. Psychophysiology **8**, 656–672 (1971)

132. Rooney, B., Hennessy, E., Bálint, K.: Viewer versus film: exploring interaction effects of immersion and cognitive stance on the heart rate and self-reported engagement of viewers of short films. In: Proceedings at the society for cognitive studies of the moving image, Franklin & Marshall College, Lancaster, Pennsylvania, USA, 11- 14 June 2014. (2014)

133. Trans Cranial Technologies: 10–20 System Positioning MANUAL. (2012) Wanchai, Hon Kong, 13

134. Ladino Nocua, A.C., Cruz Gonzalez, J.P., Castiblanco Jimenez, I.A., Gomez Acevedo, J.S., Marcolin, F., Vezzetti, E.: Assessment of cognitive student engagement using heart rate data in distance learning during COVID-19. Edu. Sci. **11**(9), 540 (2021). https://doi.org/10.3390/educsci11090540

135. Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. Towards a science of user engagement (position paper). In: WSDM workshop on user modelling for Web applications. pp. 9–12 (2011).

136. Heather, L.O., Elaine, G.T.: What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci. Technol. **59**(6), 938–955 (2008)

137. Motyl, B., Filippi, S.: Trends in engineering education for additive manufacturing in the industry 4.0 era: a systematic literature review. Int. J. Interact. Des. Manuf. **15**, 103–106 (2021). https://doi.org/10.1007/s12008-020-00733-1

138. Maxwell, J.: Qualitative Research Design: An Interactive Approach, 3rd edn. SAGE, Thousand Oaks, CA (2013)

139. Burkholder, G., Cox, K., Crawford, L., and Hitchcock, J. Research Design and Methods: An Applied Guide for the Scholar-Practitioner (1st edn). Thousand Oaks, CA: SAGE.

140. Diamantopoulos, A., Winklhofer, H.: Index construction with formative indicators: an alternative to scale development. J. Mark. Res. **38**(2), 269–277 (2001)

141. Lavrakas, P. (2008). Encyclopedia of survey research methods. Thousand Oaks, CA: SAGE

142. Pontes, A.I., Henn, M., Griffiths, M.D.: Assessing young people's political engagement: a critical and systematic literature review of the instruments used to measure political engagement. Int. Polit. Rev. **4**, 55–72 (2016). https://doi.org/10.1057/s41312-016-0002-4

143. Hollebeek, L., Sprott, D.: Handbook of Research on Customer Engagement. Edward Elgar Publishing, Camberly, UK (2019)

144. Kulikowski, K.: Do we all agree on how to measure work engagement? Factorial validity of Utrecht work engagement scale as a standard measurement tool - a literature review. Int. J. Occup. Med. Environ. Health. **30**(2), 161–175 (2017). https://doi.org/10.13075/ijomeh.1896.00947. (**Epub 2017 Feb 17 PMID: 28366949**)

145. Bazzani, A., Ravaioli, S., Trieste, L., Faraguna, U., Turchetti, G.: Is EEG suitable for marketing research? A systematic review. Front. Neurosci. **14**, 594566 (2020). https://doi.org/10.3389/fnins.2020.594566

146. McWilliam, R.A.: Children's Engagement Questionnaire. Frank Porter Graham Child Development Center, University of North Carolina, Chapel Hill, NC (1991)