Computers & Education 200 (2023) 104797

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Measuring emotions in education using wearable devices: A systematic review

Shen Ba, Xiao Hu *

Faculty of Education, The University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

Keywords:
Teaching/learning strategies
Data science applications in education
Evaluation methodologies
Distributed learning environments

ABSTRACT

Wearable devices that detect real-time and fine-grained physiological signals offer potentials for understanding the intricate mechanisms of emotions in education. However, due to the diversities of wearable devices, physiological signals, educational emotions, and educational contexts, there is lack of consensus on the affordance and constraints of wearable devices for measuring emotions in education. The present study conducted a systematic literature review and examined 50 peer-reviewed journal articles and influential proceedings published over the last 15 years (January 2008 to December 2022). Five research questions were addressed concerning research backgrounds, theoretical frameworks, methodologies, remaining challenges, and ethical considerations. Findings demonstrated that while most studies focused on university students in controlled environments, recent advances in wearable devices have enabled emotion measurements of younger learners in natural settings. Research interests have developed towards understanding the theoretical connections between emotion and cognition leveraging wearable devices. Electrodermal activity and heart rate were the most frequently measured signals whereas "engagement", "positive", and "anxiety" were the most studied emotions. Machine learning and inferential statistics were often adopted to examine associations between physiological signals and educational emotions. Moreover, we identified a need for updated ethical guidelines in advanced data collection using wearable devices. This review can not only inform wearable device usages in educational practices but also shed light on future research.

1. Introduction

Emotions are something almost everyone feels and expresses every day. Yet, it is challenging to precisely define emotions. Even before 1981, researchers have proposed nearly 100 different definitions of emotions (Kleinginna & Kleinginna, 1981). While it can be difficult to determine a one-size-fits-all definition, certain behavioral, psychological, and physiological processes that accompany the manifestations of emotions have been established. Physiological activations (e.g., heartbeat) were recognized as closely associated with emotions, which enabled the quantifications and computations of emotions (Picard, 1997).

This study coined the term "educational emotion" to refer to any emotions occurred in educational settings, including emotions related to learning mathematics (Strohmaier, Schiepe-Tiska, & Reiss, 2020), emotions expressed during teaching (Al-Fudail & Mellar, 2008), and emotions experienced while participating in serious games (DeFalco, Rowe, Paquette, Georgoulas-Sherry, Brawner, Mott et al., 2018). Picard, Papert, Bender, Blumberg, Breazeal, and Cavallo et al. (2004) advocated that emotions were intertwined with

* Corresponding author. Room 209, Runme Shaw Building, The University of Hong Kong, Pokfulam, SAR, Hong Kong, China.
E-mail addresses: woloshen@hku.hk (S. Ba), xiaoxhu@hku.hk (X. Hu).

https://doi.org/10.1016/j.compedu.2023.104797
Received 14 October 2022; Received in revised form 31 March 2023; Accepted 3 April 2023
Available online 8 April 2023

cognition in a complex manner by affecting what humans think, how they behave, and other cognitive incidents. To build a comprehensive and nuanced understanding of learning process, it is critical to consider the functions and effects of emotions. Since then, there have been increasingly more theoretical and empirical studies exploring the roles and functions of emotions in diverse educational settings (Buissink-Smith, Mann, & Shephard, 2011; Eliot & Hirumi, 2019; Linnenbrink, 2006; Pekrun, 2006). For improving self-learning, emotion recognition was added to intelligent tutoring systems to detect students' emotions such as confusion through gestures, conversations, and facial movements (D'mello & Graesser, 2013; D'Mello, Picard, & Graesser, 2007; Nye, Graesser, & Hu, 2014). In school settings, based on the control-value theory (CVT) of achievement emotions, a meta-analysis consisting of 68 studies revealed that students' learning enjoyment was associated positively with their academic performance (Camacho-Morles et al., 2021). Similarly, in multimedia learning settings, the visually appealing design of learning content was proposed to enhance positive emotions (Wong & Adesope, 2021), highlighting the impact of emotions on human information processing. Furthermore, in MOOC discussions, Liu, Liu, Liu, Peng, and Yang (2022) detected students' emotional and cognitive engagement through text classification models and uncovered their associations with learning achievement.

Given the significance of emotions in education, a critical question is how to effectively recognize and monitor emotions in educational settings. In the last two decades, extensive efforts have been devoted to addressing this question, which were summarized by several literature reviews. Feidakis (2016), Harley (2016), and Imani and Montazer (2019) examined systems and tools for detecting emotions in e-learning environments. The three reviews considered the pros and cons of several measuring modalities and methods but were limited to only digital learning environments. Moreover, Hasan, Noor, Rahman, and Rahman (2020) looked into transitions from intelligent tutoring system (ITS) to affective tutoring system (ATS) to investigate the implementations and influence of emotional components in ATS. Thus, only studies involving ITS or ATS were included. Furthermore, from a more comprehensive perspective, Wu, Huang, and Hwang (2016) reviewed studies on affective computing in education published from 1997 to 2013. Specifically, research questions related to research trends, sample types, learning domains, application issues, and remaining challenges were explored. Following Wu et al. (2016), Yadegaridehkordi, Noor, Ayub, Affal, and Hussin (2019) further investigated the research purposes, emotion models, and measuring methods of more recent (i.e., 2010–2017) studies on educational affective computing. According to results from Wu et al. (2016) and Yadegaridehkordi et al. (2019), textual and self-report data have been the primary modalities for educational emotion detection. For instance, Yadegaridehkordi et al. (2019) synthesized 49 studies using textual modality (e.g., online discussion transcripts) while only nine adopted physiological data. Consequently, their reported trends and patterns could favor the dominant modalities while the affordance and suitability of wearable devices for affective computing are largely underrepresented. Similarly, a recent review examined affordance of wearable devices in supporting learning analytics (Liu, Ren, Kong, & Liu, 2022). While their review mentioned usages of wearable devices for measuring emotions, the content analysis focused more on cognitive and behavioral domains. To our best knowledge, there has not been a study that systematically and exclusively reviews the progress of measuring emotions in education with wearable devices.

Wearable devices refer to a branch of mobile devices which are attachable to human bodies and use built-in biosensors to monitor physical and physiological signals such as electrocardiogram (ECG), electroencephalogram (EEG), and electrodermal activity (EDA) (Billinghurst & Starner, 1999; Seneviratne, Hu, Nguyen, Lan, Khalifa, Thilakarathna et al., 2017). There has been an observable increase in educational studies using physiological sensors to detect emotions (Wu et al., 2016). In addition, a significant number of studies in fields such as health care has adopted wearable devices and advocated their advantages for producing real-time and fine-grained emotional information (Shu, Xie, Yang, Li, Li, Liao et al., 2018).

Considering the importance of emotions in education as well as the enormous potential of wearable technologies in providing non-invasive, objective, and real-time measurements, we deem a retrospective work that summarizes previous experiences and identifies future directions necessary. Moreover, as indicated by Graesser (2020) and Mayer (2020), since there has been no gold standard for emotion detection in education yet, a better understanding of each modality will be instructive for improving emotion measurement by incorporating multiple modalities. Therefore, the present study intends to synthesize existing empirical studies and uncover the research status of measuring emotions in education with wearable devices. A systematic literature review (SLR) approach was employed. Particularly, we adopted the research framework from Wu et al. (2016) and Yadegaridehkordi et al. (2019) concerning five important aspects of measuring emotions in education, including research backgrounds, theoretical frameworks, methodologies, challenges and limitations, and ethical considerations. The following research questions (RQs) guided this SLR:

RQ1. What are the research purposes, learning environments, and subjects of the reviewed studies?

RQ2. What theories, models, or frameworks do the reviewed studies refer to?

RQ3. What wearable devices, emotions, and modeling methods are adopted in the reviewed studies?

RQ4. What challenges and limitations are stated in the reviewed studies?

RQ5. What are the ethical considerations of measuring emotions with wearable devices?

By answering these RQs, this SLR aims to build upon the current understanding of affective computing in education and provides a comprehensive overview regarding the usage of wearable devices for measuring emotions in education. Findings from this SLR can not only inform researchers and practitioners in terms of when, what, and how wearable devices may be employed to detect educational emotions, challenges, limitations, and ethical issues encountered in empirical studies will also be examined with the intention of acknowledging research boundaries and enlightening future directions.

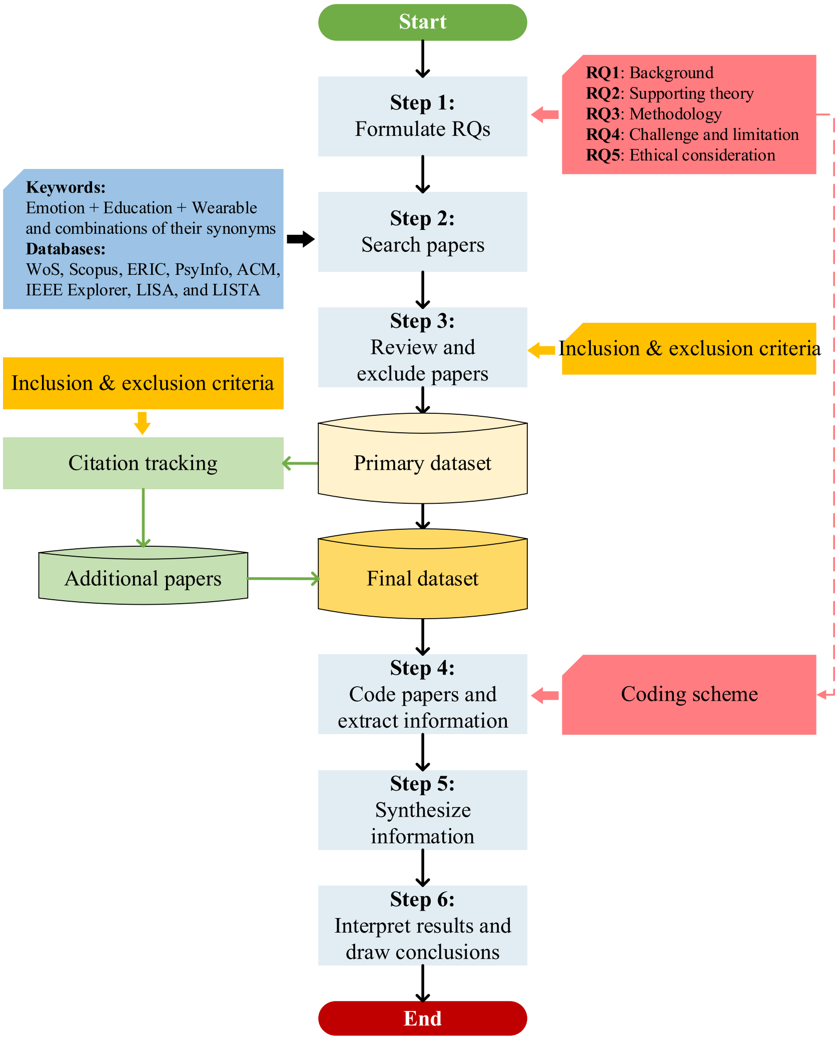

Fig. 1. The systematic literature review procedures.

2. Methods

We followed procedures in the *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) statement (Page, McKenzie, Bossuyt, Boutron, Hoffmann, Mulrow et al., 2021). PRISMA provides reporting guidelines for various reviews and has been applied extensively in education research (e.g., Crompton & Burke, 2018; Yadegaridehkordi et al., 2019). As presented in Fig. 1, this SLR consisted of six steps. First, the purposes and research questions (RQs) that motivated this research were identified. Second, we searched multiple digital databases with predefined search terms for relevant studies. Third, an initial selection of papers was performed to build the primary dataset according to the inclusion and exclusion criteria. Based on the primary dataset, citation tracking was conducted to ensure more complete coverage. Fourth, we coded the final dataset and extracted information relevant to the RQs. In the final two steps, the extracted information was compared, synthesized, and discussed in response to the RQs.

2.1. Inclusion and exclusion criteria

This SLR developed several criteria for selecting and screening research papers. First, we considered studies published in the past 15 years (i.e., January 2008 to December 2022) which witnessed the release of first-generation commercialized wearable devices (e.g., Fitbit, Jawbone) and the rapid development of wearable technologies (Ometov, Shubina, Klus, Skibińska, Saafi, Pascacio et al., 2021). Second, to ensure the representativeness of research articles, only those published in reputable and peer-reviewed journals (e.g., Computers & Education) and conferences (e.g., International Conference of Learning Analytics & Knowledge) were considered. Third, we focused on papers written in English. Fourth, we included studies that have measured at least one emotion with automatic means. Last, while wearable devices have been employed for emotion recognition in various fields, this SLR concentrated on empirical studies

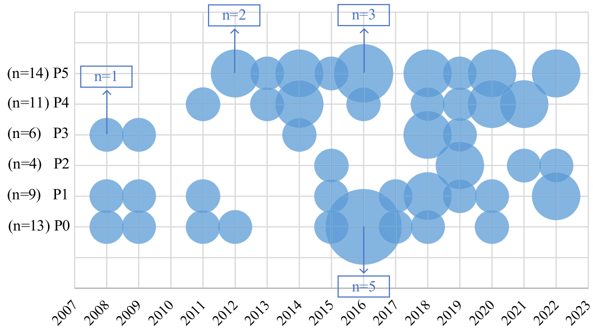

Fig. 2. Number of studies carrying each purpose per year.

within educational contexts. Consequently, studies that did not satisfy the above criteria were excluded, including those that were 1) outside the coverage period, 2) non-English, 3) without full text, 4) non-empirical, or 5) did not measure emotions with wearable devices.

2.2. *Searching strategies*

The search process included two phases. In phase one, a series of keywords under three categories namely “emotion”, “education”, and “wearable” were identified and organized as queries using ‘AND’ and ‘OR’ operators (e.g., ["emotion recognition” OR “affective computing” OR “mood detection” OR “sentiment analysis”] AND [education OR learner OR student OR teacher OR instructor] AND [wearable OR physiological OR sensor]). Given that this SLR covered multiple related research fields including education, information science, and engineering, we searched eight popular databases (Scopus, Web of Science, ACM Digital Libraries, IEEE Explorer, LISA, LISTA, ERIC, and PsycInfo). Metadata of the retrieved literature, such as titles, author names, and abstracts, were obtained and stored. The inclusion and exclusion criteria were then applied to identify studies that met all the requirements through three rounds of screenings based on titles, abstracts, and full texts respectively.

In phase two, based on the primary dataset established in phase one, we further performed citation tracking on each selected paper with a network tool called Connected Papers (https://www.connectedpapers.com/). This tool can detect papers with high similarity based on citations, co-citations, and bibliographic coupling. The same screening criteria were then applied to the newly found papers.

2.3. *Coding and data extraction*

Following the searching and screening procedures, two experienced researchers reviewed and coded the final dataset. Both researchers have bachelor’s or above degrees in Educational Technology and more than three years of full-time or part-time experience in education-related research. A coding scheme was designed based on the five RQs (Appendix A). The coders first established a consistent understanding of the codes through working together on two sample papers and then coded independently 25% of the studies in the final database. Inter-rater reliability was measured by Cohen’s kappa (Cohen, 1960) to report consistency between coders. Afterwards, the coders discussed extensively to resolve disagreements. If the inter-rater reliability was acceptable ($0.61 <$ kappa <0.80) or higher (McHugh, 2012), one coder continued coding the rest 75% of the papers.

3. Results

From the initial literature search, this SLR gathered a total of 665 studies. After screening and selection according to the inclusion and exclusion criteria, we found 44 studies that met all requirements, which formed the primary dataset. Through citation tracking, six additional studies were found to fit the scope and thus the final dataset consisted of 50 studies. The majority of the studies were reported in journal articles while six were published in proceedings of the Association for Computing Machinery (ACM), International Conference on Learning Analytics & Knowledge (LAK), and International Conference on Educational Data Mining (EDM). The six conference papers have been cited from 12 to 105 times according to the Google Scholar metrics (data was collected on March 20, 2023). As for the journal papers, four of them (8%) were found in Computers & Education, four (8%) from the British Journal of Educational Technology, and the remaining 82% were scattered across 32 different publications, such as Computers in Human Behavior and Educational Technology & Society. Furthermore, nine studies (20%) were published in 2016, followed by six (13%) in 2018, 2019, and 2020 respectively. In terms of inter-rater reliabilities, Cohens’ kappa for each coding field ranged from 0.79 for “research purpose” to 0.96 for “level of education”, indicating substantial to near-perfect agreement (McHugh, 2012).

3.1. RQ1*: research purposes, learning environments, and subjects*

The purposes of reviewed studies were summarized into six categories: (P0) Developing systems that recognize and respond to learner emotions; (P1) Designing methods for classifying/predicting emotions; (P2) Comparing different emotional measurement

Table 1
Number of studies conducted in different learning environments.

Learning environments		Number of studies
Lab (36)	E-learning system	14
	Multimedia learning	10
	Game-based learning	4
	Math test	4
	Intelligent tutoring system	3
	Cognitive and social tasks	1
Classroom (11)	Face-to-face lecture	7
	Face-to-face collaboration	3
	Virtual lecture	1
Out-of-classroom (3)	Learners' own choices	3

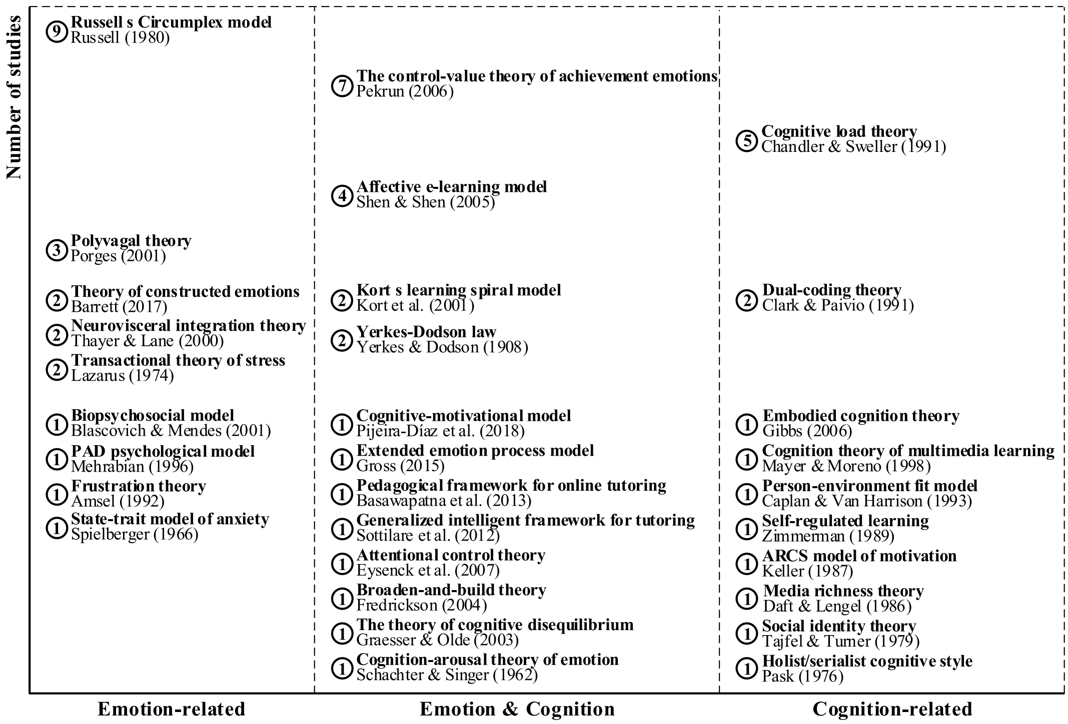

Fig. 3. Number of studies referring to each theoretical framework.

methods; (P3) Examining emotion changes during learning activities; (P4) Investigating relationships between emotions and other learning-related constructs (e.g., performance); (P5) Assessing effects of instructional designs/interventions on learners' emotions. Fig. 2 shows distributions of the reviewed studies across the six research purposes and publication years. The diameter of a circle indicates the number of studies. A study may have multiple purposes.

From 2008 to 2012, most studies developed systems or methods that utilized wearable devices to detect emotions (P0 and P1). This research direction continued after 2012 and witnessed an outbreak around 2016. Meanwhile, another group of studies (P3, P4, and P5) started to accumulate after 2012 which focused on the roles of emotions in education and their associations with other learning constructs. Comparatively speaking, studies with purposes P3, P4, or P5 started to outnumber those on P0, P1, or P2 since 2018.

Table 1 presents three general settings where the reviewed studies were conducted. First, the research lab refers to designated rooms where researchers set up necessary equipment and minimize possible distractions. Among the reviewed studies, 36 studies were conducted in research labs and many of them invited participants to interact with e-learning systems (14 studies) or multimedia materials (10). Second, 11 studies were conducted in classroom environments where a group of learners participated in lectures or collaborations. These studies were carried out in typical educational contexts and measured students' emotions during learning process. Results showed that face-to-face lectures took up the largest proportion (7) in this category. Last, three studies were conducted in out-of-classroom settings. In this case, participants were instructed to follow ordinary schedules and upload necessary information (e.g., momentary emotions) through self-report or wearable devices.

For research participants, most reviewed studies (72%) involved university students while 16% and 12% involved secondary school and primary school students respectively.

Table 2
Wearable devices and measured signals.

Type	Device	Signal											
		Heart			Brain/Face			Skin			Others		
		HR	BVP	ECG	Pulse	EEG	fEMG	EOG	EDA	ST	Respiration	Context	Movement
Finger	Biosemi								1				
	Biopac	1	1	2					3	1			
	Shimmer GSR+			1					1				1
	Smartphone	1										2	
	HBE Ubi-nanoLoc				1								
Wrist	Empatica E3, E4	1	1						9	2			1
	Q sensor								4	2		1	1
Arm	HEM-7201		1										
	BodyMedia								1	1			3
	Tyco	1	1										
Chest	Polar H7, H10	2		1									1
	VU-AMS5fs	1											
	X-Vest	1	1						1				
	FirstBeat	1		1									
Head	Microsoft HoloLens											1	
	Nexus-10	1				1			1				
	NeuroSky					5							
	emWave	4											
	EMOTIV EPOC+					1							
	AB-Medica					1							
Multiple	AICARP v2			1	1				1	1	1		
	ProComp5	1	3			2			3				
	Wireless arduino system				1				2	1			
	Physiologger	1							1				
	Unspecified	1		1	1	1	1	1	2		1		
	Sum	**17**	**8**	**7**	**4**	**11**	**1**	**1**	**30**	**8**	**2**	**4**	**7**

Note: ECG-electrocardiogram, fEMG-facial electromyography, EOG-electrooculogram, EDA (i.e., galvanic skin response/skin conductance/skin resistance), Movement (i.e., acceleration, physical activity, energy expenditure, intensity of body movement).

Regarding the learning environments of studies with post-secondary students, 31 studies (86%) were conducted in laboratories while distributions between labs and classrooms are more even in studies involving secondary (3 in labs, 3 in classrooms) and elementary (3 in labs and classrooms respectively) school learners. It is noteworthy that two out-of-classroom studies were conducted at post-secondary level, and the other one was at secondary level. In terms of sample sizes, half of the reviewed studies recruited 30 to 100 participants, while 13 studies had fewer than 30 participants, and 12 studies more than 100.

3.2. RQ 2: theories, models, or frameworks

The reviewed studies were based on various theories, models, or frameworks that can be grouped into three main categories as shown in Fig. 3: 1) cognitive consideration of learning, 2) structure and mechanism of emotions, and 3) shared importance and mutual influence between cognitive and emotional constructs in education. Each study could adopt multiple theories. Citations of reviewed studies adopting different frameworks were listed in Appendix B.

For cognition-related theories, we found that cognitive load theory (Chandler & Sweller, 1991) was quoted most frequently, followed by dual-coding theory (Clark & Paivio, 1991) (twice) and other theoretical frameworks (once).

Regarding theories on emotions, Russell's circumplex model (Russell, 1980) outnumbered others by a large margin, followed by polyvagal theory (Porges, 2001) (three times). The reviewed studies also referred to the theory of constructed emotions (Barrett, 2017), the neurovisceral integration theory (Thayer & Lane, 2000), and the transactional theory of stress (Lazarus, 1974) twice respectively. The remaining theories appeared only once.

In terms of theoretical frameworks that connect emotion with cognition, the control value theory (CVT) of achievement emotions (Pekrun, 2006) was leading the chart occurring in seven studies, followed by the affective e-learning model (Shen & Shen, 2005) (four studies), Yerles-Dodson law (Yerkes & Dodson, 1908) (twice) and Kort's learning spiral model (Kort, Reilly, & Picard, 2001) (twice). The others were only referred to once.

3.3. RQ 3: wearable devices, emotions, and modeling methods

Table 2 presents the number of reviewed studies using each wearable device to examine each related variable. We grouped devices by how they were worn and various types of signals from related organs (Seneviratne et al., 2017). Empatica wristbands have not only been used extensively (12) but also employed by the most recent studies. Other devices used include NeuroSky (5), Q sensor (4), eWave (4), and Polar chest bands (4). As for variables and signals, electrodermal activity (EDA, 30) and heart rate (HR, 17) were the most widely adopted for measuring educational emotions, followed by electroencephalogram (EEG, 11), skin temperature (ST, 8), and blood

Fig. 4. Emotions and signals measured in reviewed studies
Note: Resp-respiration, Move-movement; Numbers indicate the number of reviewed studies.

volume pressure (BVP, 8). Notably, while most studies paid attention to signals obtained from human bodies, four studies also included context variables measured by wearables such as location and light.

Furthermore, looking at the intersections between devices and signals, it was learned that nine studies employed Empatica wristbands to measure EDA. Five utilized NeuroSky to obtain EEG brain activities of students, while emWave was the most popular for measuring HR (4).

There are a total of 30 types of emotions measured in the reviewed studies, plus three sentiment polarities, “positive”, “neutral” and “negative”. Fig. 4 presents the top 20 emotions and sentiment polarities based on the number of studies examining them, while the citations of the actual studies are listed in Appendix C. It was found that “engagement” (12), “anxiety” (9), “positive” (8) and “boredom” (8) appeared the most. The emotions are ordered in Fig. 4 roughly based on their natures of being positive (i.e., “engagement”, “happy/joy”, “peacefulness”, “enjoyment”) and negative (i.e., “anxiety”, “boredom”, “confusion”, “frustration”, “stress”, “anger”, “fear/scared”, “disgust”, “sadness”, and “nervousness”)

Moreover, Fig. 4 demonstrates the relationships between measured signals and emotions. Among the top 20 emotions, 16 were measured using EDA, marking the most frequently adopted signal. Besides, HR was employed to measure “positive” (5), “negative” (4), “engagement” (3), “stress” (3), and “peacefulness” (3). Specifically, for measuring sentiment polarities “positive” and “negative”, HR was used most often among all physiological signals. Furthermore, ST was recorded five times for “engagement” and four for “arousal”. The “nervousness” measurement included pulse three times. The measurement of “engagement”, the most studied emotion, involved various signals (i.e., EDA, Movement, ST, BVP, and HR) at least three times.

In the reviewed literature, researchers have adopted five types of approaches to study emotions based on data collected by wearable devices. First, 15 studies used machine learning to model the predictive relationships between collected data and emotions, particularly traditional machine learning classifiers (e.g., support vector machine) rather than deep learning approaches. Second, 14 studies implemented inferential statistics (e.g., correlation and analysis of variance) to test if a statistically significant relationship existed between physiological signals and other emotion measurements such as self-reported arousal and valence. Third, eight studies derived emotions from physiological values based on conclusions in previous literature. For example, Cowley, Ravaja, and Heikura (2013, p. 300) stated that “recording at the Orbicularis Oculi (periocular) can index high arousal positive valence (Ekman, Davidson, & Friesen, 1990)”. Moreover, seven studies applied descriptive statistics to compare educational emotions measured by wearable devices with results produced by other more established approaches. Lastly, this review also identified a smaller group of studies (6) that used

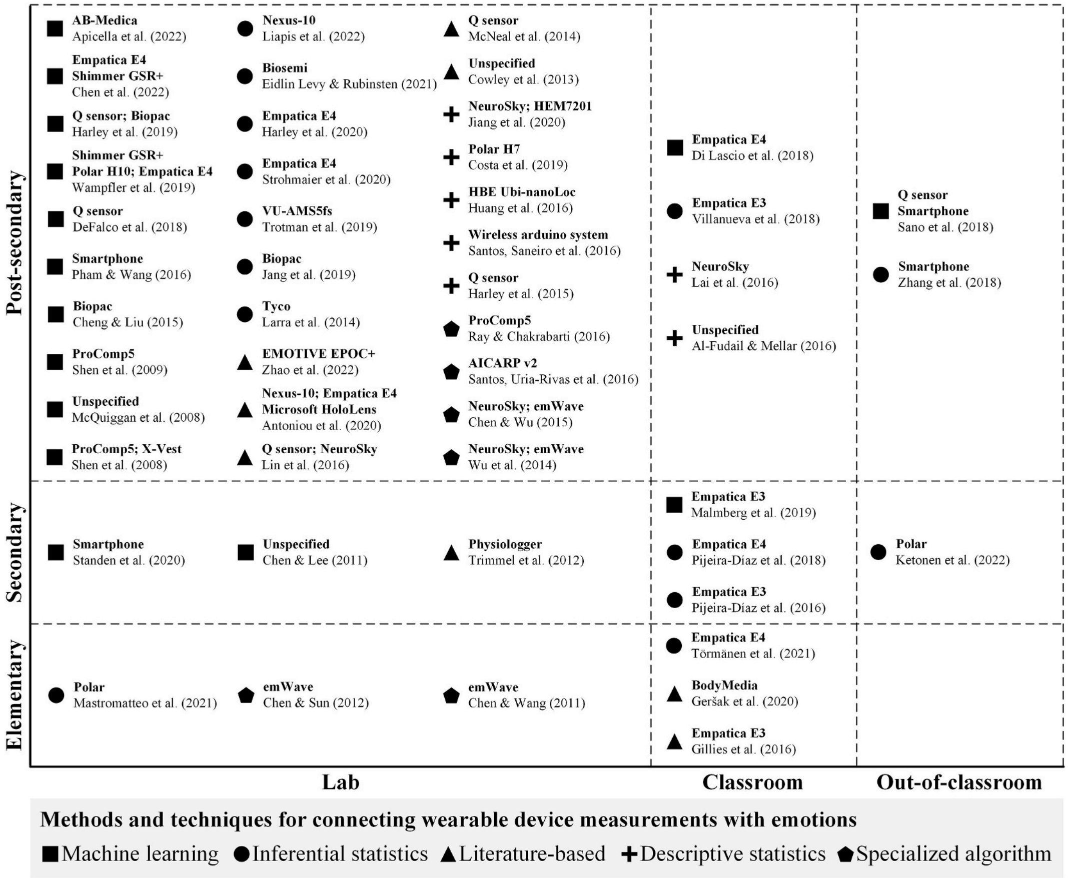

Fig. 5. Emotion measuring mechanisms in diverse contexts.

Table 3
Challenges and limitations proposed in the reviewed studies.

Categories	Challenges and limitations	Number of studies
Generalizability	Focused only on learners of a particular gender, age, personality, or health status.	14
	Study should be repeated in other contexts.	14
	The sample size is too small.	14
Method	Multimodal data might improve predictions of emotions.	10
	The accuracy of certain wearable devices should be considered.	6
	Advanced machine/deep learning algorithms should be considered.	3
	Contextual factors should also be considered.	2
	Multiple emotions could appear at one time.	2
	Aligning learning and emotional events with physiological measures.	2
	Data should be collected at a finer granularity.	2
	Some devices are not convenient to wear.	1
	Physiological data need to be better represented.	1
	The data collection time is too short, and some changes may not be observed.	1
	Lack of self-report measures to compare with.	1
Theory	Relationships between emotional and educational constructs are unclear.	4
	Emotional reactions are subject to personal factors and differences.	2
	The meanings of physiological features need to be better explained.	1
	No causal relationship between physiological signals and emotions are identified.	1
Pedagogy	Learning materials and interventions need to be better designed.	6
	Effects of specific pedagogical actions on emotions are less examined.	2

specialized algorithms to deduce emotions with physiological data. For instance, by calculating three indices (i.e., low-, medium-, and high-frequency zones) in power spectral density analysis of heart rate, Chen and Wang (2011) estimated positive and negative emotions of learners. To further depict different mechanisms for connecting educational emotions with wearable devices in diverse teaching and learning contexts, an overview was offered in Fig. 5.

3.4. RQ 4: challenges and limitations

In Table 3, we summarized challenges and limitations discussed in the reviewed studies into four categories. First, for theory-related limitations, four studies suggested that the associations between emotional and educational constructs were under-studied. Two studies expressed concerns about the precision of predicting educational emotions, as emotional responses could vary even for the same student. Second, pedagogy-related limitations were raised, particularly on the difficulty in inducing learners' emotions through pedagogical interventions (6). Third, challenges in data collection and analysis methods were mentioned, including the need for multimodal data to improve the accuracy of emotion predictions (10), accuracies of wearable data that might be affected by personal and contextual factors (6), and the need for more advanced algorithms in emotion predictions (3). Finally, the generalizability of findings was discussed at length in nearly all reviewed papers, ranging from the necessity to repeat the study with different participants (14) and in different learning contexts (14) to limited sample size (14).

3.5. RQ 5: ethical considerations

Around half (24) of the reviewed studies considered ethical issues related to their research. We coded the ethical considerations into three categories according to Tzimas and Demetriadis (2021) and Yan, Zhao, Gasevic, and Martinez-Maldonado (2022), namely 1) privacy (4) referring to protecting learners' identities and secure data access, 2) transparency (17) meaning to ensure learners' understanding of the research procedure such as informed consent, and 3) fairness/bias (3) referring to potential biases caused by sample selection or measuring tool. Among the four studies that included privacy considerations, Di Lascio, Gashi, and Santini (2018) raised concerns that anonymous research data could be intentionally or unintentionally deanonymized and used for inappropriate interventions. To reduce this type of risk, another reviewed study (Wampfler, Klingler, Solenthaler, Schinazi, & Gross, 2019) argued that log data and physiological data were relatively safer than video and auditory data in protecting learners' identities. Moreover, when designing a learning application, Zhang, Li, Chen, and Lu (2018) specifically discarded privacy-sensitive information (e.g., phone numbers and usernames) and provided learners with options to opt out of being recorded. For research transparency, while 17 studies claimed to have obtained informed consent from participants and thus showing some degree of transparency, none of them mentioned any details of the consent. In terms of fairness and bias, Strohmaier et al. (2020) and Zhang et al. (2018) suggested that wearable-based measurements were more objective than self-report which suffered from social desirability and stereotypes. Meanwhile, McNeal, Spry, Mitra, and Tipton (2014) and Strohmaier et al. (2020) both considered data representativeness by pointing out that mathematics students may express different emotions towards math learning than students from other majors. Nevertheless, the reviewed studies did not mention algorithmic bias during data modeling or pattern interpretation. Citations of reviewed studies addressing each of these ethical issues can be found in Appendix D.

4. Discussion

In this section, we discuss our findings for each research question.

4.1. Research purposes, learning environments, and subjects

The research purpose reflects the core objectives of a study. Synthesizing the purposes of existing works can facilitate understanding of the research field and illuminate future directions (Cohen, Manion, & Morrison, 2018). From the reviewed studies, we observed the parallel development of two major research themes: 1) building systems or methods that monitor emotions with wearable technologies (e.g., McQuiggan, Mott, & Lester, 2008; Pham & Wang, 2015; Shen, Callaghan, & Shen, 2008); and 2) examining the roles of emotions in learning and regulating emotions through pedagogical designs (e.g., Chen & Sun, 2012; DeFalco et al., 2018; Strohmaier et al., 2020). This interplay between methodological explorations and empirical investigations demonstrates positive dynamics in the field. Nonetheless, compared to textual, visual, and auditory modalities, the implementation of wearable devices in affective learning is still in its infancy (Imani & Montazer, 2019; Wu et al., 2016; Yadegaridehkordi et al., 2019). Future studies are encouraged to continue the current directions both by developing new computational methods for emotion recognition and by implementing empirical interventions to further clarify the functions of emotions in education.

It is often discussed that individuals at different life stages and education levels differ in their emotional responses (Bailen, Green, & Thompson, 2019; Mora, Urdaneta, & Chaya, 2018). This leads to the necessity to examine diverse participants. The findings that over 70% of the reviewed studies recruited university/college students is not surprising since many researchers work at universities, and sophisticated experiment setups with wearable devices often require high-level cooperation (Antoniou, Arfaras, Pandria, Athanasiou, Ntakakis, Babatsikos et al., 2020; Shen et al., 2008). Nevertheless, adolescence signifies a phase when children develop rapidly in all aspects including emotion, cognition, and physicality (Guerra-Bustamante, León-del-Barco, Yuste-Tosina, López-Ramos, & Mendo-Lázaro, 2019). During this phase, their emotions are more sensitive and may vary due to daily events (Ba, Hu, Kong, & Law, 2022; Lennarz, Hollenstein, Lichtwarck-Aschoff, Kuntsche, & Granic, 2019). Therefore, more research is needed to uncover the emotional variations of younger learners. Moreover, compared to other more adopted emotion measurements (e.g., text analysis and facial expression analysis), lightweight and robust wearable devices can be more ubiquitous and less obtrusive in monitoring the emotions and well-being of adolescents (under appropriate ethical terms), which is also promising for helping prevent potential mental problems and risks (Hu, Chen, & Wang, 2021).

This review uncovered three major environments where studies were conducted, namely laboratory, classroom, and out-of-classroom settings. Laboratories are ideal for controlled experiments that focus on evaluating certain e-learning systems, pedagogical designs, or learning interventions. Moreover, some wearable devices are too delicate or expensive to be used outside of laboratories (Antoniou et al., 2020; Shen et al., 2008). However, researchers have argued that emotions in controlled environments might be different from naturalistic settings (Goldstein & Strube, 1994; Reichenberger, Schnepper, Arend, & Blechert, 2020). In education, students may be accustomed to learning alone without being observed or in the presence of classmates and friends. Familiar environments help students express authentic emotions during learning (Borup, West, & Graham, 2012). In this regard, future studies should make more efforts to examine student emotions in authentic learning environments with wearable devices.

4.2. Theories, models, or frameworks

Among the reviewed studies, we categorized theoretical frameworks into three groups, namely cognition-related, emotion-related, and those on the interrelationship between emotion and cognition (Fig. 3). While distinguishing emotion models is undoubtedly essential, we propose that contextualizing emotions in educational settings and making sense of their associations with cognitive constructs is instructive for improving teaching and learning. For instance, while constructs such as test performance and higher-order thinking are both related to emotions, they describe different aspects of learning and may connect to emotions differently (Schillinger, Mosbacher, Brunner, Vogel, & Grabner, 2021). The control-value theory (CVT) of achievement emotions defined a set of emotions in educational settings and linked them to students' appraisals of their controls and values of learning process and environments (Pekrun, Frenzel, Goetz, & Perry, 2007). Therefore, CVT helps stakeholders understand how emotions and cognition interact in educational settings. Identifying the relationships between emotional and cognitive constructs will help enhance our understanding of the roles that emotions play in learning and interpret the causes and changes of educational emotions.

For theories associated with cognition, cognitive load theory (CLT) was referred to most among the reviewed studies. CLT indicates that working memory can only process a limited amount of information at one time (Chandler & Sweller, 1991). Thus, instructional designers should be cautious and selective about materials that are presented to learners. Plass and Kalyuga (2019) summarized four possible ways to consider emotion under the scope of CLT, and each can be observed in empirical studies. First, emotional designs (e.g., appealing visual elements) may be seen as external information that competes with learning content for cognitive resources. Second, some emotions may help expand the capacity of cognition and in turn promote learning. Third, emotions may be associated with the intrinsic cognitive load in the process of self-regulation. Lastly, emotions may also influence the mental efforts of students. These four roles of emotions may happen at the same time and one role could dominate others under certain circumstances. For instance, when designing background music for learning, music can sometimes promote learning performance by enhancing learners' positive emotions, while risking learning by distracting learners' attention at other times (Que, Zheng, Hsiao, & Hu, 2023). These different effects are suggested to depend on the properties of music (e.g., fast or slow tempo) (Hu, Li, & Kong, 2019), preferences of learners (Li, Wang, Ng, & Hu, 2021), and types of learning tasks (Li & Hu, 2023). For emotion-related theoretical frameworks, Russell's circumplex model was the most popular. It is also known as the dimensional model that projects emotional states onto two continuous dimensions (Russell, 1980). One dimension indicates the valence level (i.e., from negative to positive) of an emotional state; the other describes the arousal level (i.e., from calm to energetic). While the dimensional model supports quantifying emotions in fine granularity, it also faces challenges such as the interpretations of emotional states (Larsen & Diener, 1992).

In contrast, the CVT of achievement emotions adopts the discrete emotion model which uses human natural language to label emotions such as "happy", "sad", and "angry", and thus are more explainable (Pekrun, 2006). Nevertheless, since a discrete model contains a finite set of emotion labels, its completeness is always debatable.

To compensate for the inadequacies of dimensional and discrete emotion models, some studies suggested integrating the two. For instance, based on the circumplex model, Fernández-Caballero, Martínez-Rodrigo, Pastor, Castillo, Lozano-Monasor, López et al. (2016) classified discrete emotion labels across two dimensions (i.e., high-low activation and pleasant-unpleasant). Similarly, Pekrun et al. (2007) grouped achievement emotions into three dimensions (i.e., activity focus vs outcome focus, positive vs negative, and activating vs deactivating). However, it remains largely an open question regarding the exact positions of emotion labels in the dimensional space. Wearable devices carry the advantage of monitoring emotion-induced physiological responses continuously and can detect minor variations. This provides researchers with important references for quantifying emotions. More studies are thus called for to reveal and establish the numerical associations between emotions and physiological responses.

4.3. Wearable devices, emotions, and modeling methods

In Table 2, this SLR identified wearable devices employed in the reviewed studies and signals measured. Regarding wearable devices, Empatica E3/E4, Q sensor, and NeuroSky have been employed most widely. Empatica E3/E4 and Q sensor are wireless wristbands with a series of embedded sensors, and NeuroSky is a portable headset intended for measuring brain activities. Unlike traditional physiological devices that have many electrodes and wires, these three devices are wireless and thus convenient implementation is a major advantage. For education studies, it is desirable to measure authentic learning processes with minimum disruptions, and cumbersome devices may induce negative emotions from learners such as fatigue or boredom. In this regard, wearable technologies that mimic daily accessories such as watches and headsets would be preferable. Moreover, portable devices enable researchers to conduct experiments outside the laboratories and collect data in naturalistic learning settings. Nevertheless, due to the cost of research-grade devices, their applications in large cross-sectional or longitudinal studies are very limited. It is still an ongoing task to explore wearables that suit the needs of educational applications.

Since emotions are associated with a collective of organ activities, it may not be accurate to identify emotions with only one channel of signal (Di Mitri, Schneider, Specht, & Drachsler, 2018). By examining the connections between wearables and signals (Table 2), we found that Empatica E3/E4, Biopac, Q sensor, ProComp, and BodyMedia Core armband could measure multi-channel signals. Multi-channel devices can synchronize signals from multiple sources with the same clock, which is advantageous over using multiple single-channel devices (Ayata, Yaslan, & Kamasak, 2020). Nonetheless, it remains a challenging task to process signals, select features, and build computational models that combine multiple channels of signals to accurately recognize educational emotions.

Fig. 4 presents 20 emotions that have been measured most frequently using wearables where “engagement” was deemed as an emotion and examined in 12 studies. To explain this, Di Lascio et al. (2018) suggested that “engagement” was a compound variable that has an emotional aspect similar to “enthusiasm” and “enjoyment”. Also, Shen et al. (2008) and Shen, Wang, and Shen (2009) stated that “engagement” was an emotional state that carried positive valence and arousal. In other words, “engagement” can be regarded as a process when students enjoy the learning content and are highly active in learning.

Moreover, among the top 20 emotions, “positive”, “negative”, and “neutral” describe the polarities of emotional states rather than carrying more specific emotional meanings. The frequent occurrences of these polarities in the reviewed studies reflected their popularity in learning contexts and could partially attribute to the measurement affordance of wearable devices. For example, the eWave system was relatively easy to operate and thus was adopted by multiple studies under review (e.g., Chen & Sun, 2012; Chen & Wu, 2015). However, by measuring heart rate variability (HRV) only, eWave could only distinguish different affective polarities rather than specific emotional states which puts a higher requirement on the sensors and corresponding algorithms. Based on a scientific finding that emotions occur with the collective changes of multiple physiological signals (Schmidt, Reiss, Dürichen, & Laerhoven, 2019), future studies may employ more than one wearable device and/or combine physiological measurements from various sensors to identify specific emotional states. Meanwhile, another possible research direction is the exploration of robust classifiers and machine learning models that are capable of distinguishing emotional states at a finer granularity. Furthermore, we found 11 emotions in negative wordings such as “anxiety”, “boredom”, and “fear” while only five on the positive side (e.g., “happy/joy”). This observation is similar to that of Yadegaridehkordi et al. (2019), revealing that more studies on affective computing in education tended to support learning by identifying and addressing negative emotions.

For relationships between emotions and signals, EDA was employed for detecting nearly all emotions, which could be attributable to the convenience of measuring EDA with wristband sensors (e.g., Empatica E3/E4, Q sensor). Besides, an important factor associated with human emotional responses is sweating, which alters the properties of skin in conducting current (Stern, Ray, & Quigley, 2001) and is closely related to EDA. Therefore, the measurement of EDA can be a reliable approach for representing changes in emotions (Caruelle, Gustafsson, Shams, & Lervik-Olsen, 2019). In addition, HR has also been adopted in many studies as an indicator of emotions. HR reflects activities of the autonomic nervous system based on time intervals between heartbeats (Shi, Yang, Zhao, Su, Mao, & Zhang et al., 2017). Our results indicate that HR was used more often for distinguishing polarities of emotions (i.e., positive, neutral, and negative). This is consistent with Chung, So, Choi, Yan, and Wong (2021) who showed that HR could be used for classifying happiness and sadness but not for other emotions. Therefore, while HR contributes to emotion measurements, it is not adequate if the goal is to identify more than just sentiment polarities. Future studies may include HR as a part of multimodal approaches for measuring emotions in education. Other physiological signals were less examined in the reviewed studies, which echoes the general trend of wearable-based emotion recognition in engineering (Saganowski, Dutkowiak, Dziadek, Dzieżyc, Komoszyńska, & Michalska et al., 2020). Possible reasons could be that some devices are not as convenient (e.g., EEG caps) and methods for inferring emotions from some physiological signals have not been fully developed.

The reviewed studies have followed five different approaches to infer emotions based on measurements obtained from wearables (Fig. 5). Unlike affective computing studies in engineering or computer science, which primarily used machine learning to model the relationships between physiological data and emotions, studies in education were more diverse (Rim, Sung, Min, & Hong, 2020). Only 30% of reviewed studies utilized machine learning while other studies employed inferential statistics (28%), descriptive statistics (14%), and specialized algorithms (12%). A few reasons might explain this difference. On the one hand, many of the reviewed studies aimed to examine the effects of pedagogical designs or learning interventions and it was sufficient to refer to previous studies to identify wearables that could measure certain emotions. Meanwhile, some studies included physiological signals to triangulate other emotion measurements such as self-report or facial expression, for which descriptive or inferential statistics were sufficient for comparisons. An advantage of inferential statistics is being explainable while many machine learning models are challenging to interpret.

To ensure performance and generalizability, many machine learning algorithms require large data sizes and varieties, yet it is difficult for education studies to scale up due to limited participants or wearable devices. Moreover, prediction or classification belongs to supervised learning for which ground truth (i.e., label or answer) is necessary. Usually, the ground truth labels are obtained through learners’ self-report which could disrupt their learning process (Csikszentmihalyi & Larson, 1987). Among the reviewed studies that used machine learning to predict educational emotions, the accuracies of emotion recognition ranged from around 60% (Pham & Wang, 2015; Ray & Chakrabarti, 2016) to over 90% (Malmberg et al., 2019; Shen et al., 2009), indicating room for improvement. Moreover, few studies tested the generalizability of their prediction models, suggesting that it was not clear whether and how robust their models could be used in different contexts. To improve the accuracy and versatility of wearable-based emotion recognition methods, future studies are encouraged to construct and share datasets of emotions and physiological signals collected in education contexts and to experiment with advanced machine learning models.

4.4. Challenges and limitations reported in reviewed studies

In Table 3, we summarized the challenges and limitations indicated by the reviewed studies in four aspects. From the theoretical aspect, there is a need to further understand the relationships between physiological signals, emotional states, and educational constructs. Notably, physiological signals are not only associated with emotions but also with learning variables such as mental effort and self-efficacy (Nourbakhsh, Chen, Wang, & Calvo, 2017; Setz et al., 2009). Therefore, future studies need to identify and account for learning-related variables while measuring emotions with wearable devices. For the pedagogical aspect, several studies have stressed the importance of designing proper learning materials and interventions for isolating complex emotions into individual ones that are easier to analyze (Antoniou et al., 2020; Cheng & Liu, 2015; Shen et al., 2008, 2009; Standen, Brown, Taheri, Trigo, Boulton, Burton et al., 2020; Strohmaier et al., 2020). This is particularly important because educational emotions may change along with pedagogical content and context. To improve the accuracy of measuring emotions with wearable devices, researchers may need to segment learning content into smaller units and explore what emotions are associated with specific learning designs. This would also help build more ground truth datasets of emotions in different learning contexts. From the methodological aspect, multiple studies have argued that the accuracy of wearables should be ensured before using them in education settings (Huang, Hwang, & Chen et al., 2016; Malmberg et al., 2019; Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2016; Santos, Saneiro, Boticario, & Rodriguez-Sanchez, 2016; Törmänen, Järvenoja, & Mänty, 2021; Zhang et al., 2018) since decisions are made based on those physiological signals. Moreover, while some devices may be accurate in controlled environments, further evidence is needed to evaluate if they can maintain similar levels of accuracy while used in naturalistic settings where temperature, humidity, and personal factors can vary, sometimes to a large extent. Lastly, in terms of generalizability, it is commonly suggested that more studies are needed in the future to test hypotheses and results from previous research with different participants and in different contexts (e.g., Chen & Sun, 2012; McQuiggan et al., 2008; Pijeira-Díaz et al., 2016; Wampfler et al., 2019).

4.5. Ethical considerations in reviewed studies

It is observed that more studies have become aware of the ethical issues of measuring emotions with wearables in the recent five years. However, the lack of widely recognized guidelines has prevented researchers from adequately considering and reporting the potential ethical issues (Yan et al., 2022). For example, while most studies claimed that approvals and consents were granted by participants, parents, or related institutions (e.g., ethical committee), those approvals might not always cover all the potential aspects of ethical problems given the complex nature of data collection, signal processing, data analysis, and data management for wearable devices (Resnick & Chircu, 2018). Moreover, for students and parents who do not have a background in information technology or data science, they may not have fully understood the implications of data collection, processing, storage or sharing. First, regarding data collection, wearable devices possess the advantage of measure objective learning data in real-time and in a continuous manner, which may easily cover some aspects that participants did not want to disclose (e.g., locations they have visited). Second, for data analysis, researchers often implement machine learning techniques to model the physiological and behavioral data and make predictions/classifications (Fig. 5). However, people argue that machine learning can have bias issues (e.g., sampling and algorithmic bias) which may lead to discrimination and unfairness (Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021). Third, while wearable devices normally do not record visual or auditory data, they do produce a large amount of information, which may be used to infer emotional responses and even learning performance of participants. Therefore, it is necessary to ensure the security of data and devise appropriate plans for data usage and long-term storage. Any usage beyond the original plan needs to be reviewed. Given that most data are stored in digital format, it is also important to consider how the data should be dealt with after the research. Furthermore, as advocated by Open Science and Open Data movements (Gewin, 2016; Vicente-Saez & Martinez-Fuentes, 2018), data sharing among researchers can greatly advance research in a field. While strict anonymization has become a standardized procedure for human-contributed research data, whether and how to ensure participants' right of permitting or declining data sharing is still under debate. Overall, more efforts are needed to establish standard procedures for managing data with wearable devices.

5. Conclusion

This study presents a systematic review of the research progress on measuring emotions with wearable devices in education. Following the PRISMA guidelines, 50 studies published in the last 15 years have been systematically examined. Information associated with research backgrounds, theoretical frameworks, methodologies, remaining challenges, and ethical considerations was extracted to address the five research questions. Overall, nine new findings have been identified: 1) there has been an interplay between methodological and empirical research on detecting emotions in education with wearable devices, indicating healthy dynamics in the field; 2) while many existing studies were conducted with university participants in controlled environments, lightweight and robust wearable devices make it possible to monitor emotions of younger learners ubiquitously and unobtrusively; 3) Besides the comparison between dimensional and discrete emotion models, educational studies have been particularly interested in examining the role of emotion in learning and its theoretical association with cognitive constructs; 4) wearable devices such as Empatica E3/E4, Q sensor, and NeuroSky that are portable, convenient and can detect multiple physiological signals have been more preferable in recent studies; 5) "engagement", "positive affect", and "anxiety" have been studied the most as educational emotions using wearable devices; 6) EDA and HR have been included most frequently as predictors or indicators of emotions in education; 7) machine learning and inferential statistics have been used equally for inferring the associations between physiological signals and educational emotions; 8) generalizability has been the most significant limitation mentioned by the reviewed studies; 9) although studies have become aware of ethical

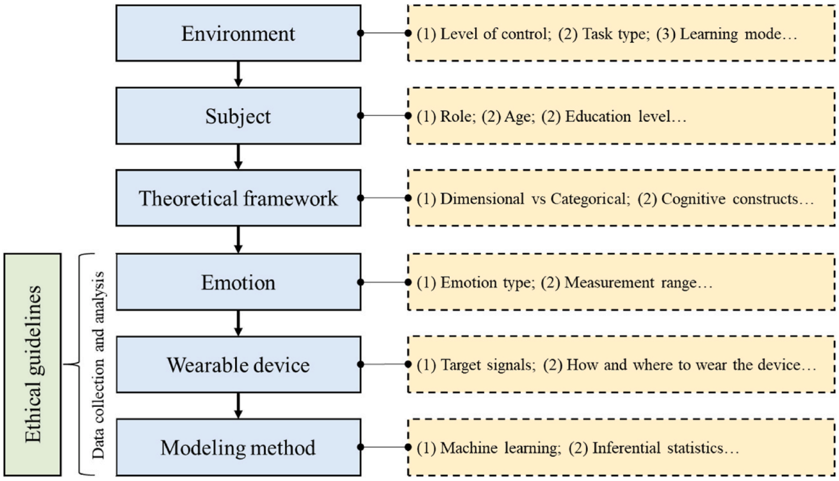

Fig. 6. Educational emotion measurement based on wearable devices.

issues in collecting physiological data, there is a need for updated ethical guidelines.

To summarize the findings, we provide a procedure diagram for guiding future applications of wearable devices in educational emotion measurements (Fig. 6). Findings of this review can inform practitioners of wearable device usage in educational settings as well as shed light on future research opportunities for researchers from education, computer science, physiology and more.

Despite research and educational implications brought by this review, certain limitations need to be acknowledged. To ensure the research quality, we have restricted our scope to peer-reviewed journal articles and influential conference proceedings (based on conference scale and Google Scholar citation number). This restriction could inevitably exclude studies that might be highly relevant and insightful, especially when measuring emotions in education with wearable devices is still an under-studied and fast-developing research topic. Besides, retrospective work was often inevitably subject to publication bias where statistically significant results were favored more. Furthermore, some physiological signals might stem from the same measurement but appear in different forms (e.g., skin conductance level and skin resistance level), we have combined these sub-features to keep the results consistent and reveal the general trend. Future studies should consider these variations when a particular physiological signal is of interest.

Credit author statement

Shen Ba: Formal analysis, Methodology, Validation, Writing - Original Draft, Visualization, Project administration, Xiao Hu: Conceptualization, Methodology, Data Curation, Writing - Review & Editing, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgments

Funding: This work was supported by the Research Grants Council of the Hong Kong S. A. R., China [grant numbers HKU17607018 & T44-707/16-N] and National Natural Science Foundation of China [grant number 61703357]. We thank Dr. Jingyan Lu and Dr. Ling Li for their helpful discussions at early stages of this study.

Appendix A

Scheme for coding and data extraction

Extracted data		Description/Example codes
RQ 1	Research purpose	Research aims stated.
	Learning environment	e.g., lab, classroom, or online learning.
	Education level	e.g., post-secondary, secondary, or elementary.
	Sample size	Number of participants.
RQ 2	Theoretical framework	Theories that support the studies. e.g., circumplex model, control-value theory of achievement emotions.
RQ 3	Studied emotion	e.g., anxiety, happy, and boredom.

(continued on next page)

(*continued*)

Extracted data		Description/Example codes
	Wearable device	e.g., Empatica E4 wristband, Polar H10 belt.
	Variable measured	e.g., heart rate, EDA.
	Analytical method	e.g., machine learning models, inferential statistics.
RQ 4	Limitation	Challenges or limitations stated.
RQ 5	Ethics	Ethical considerations stated, e.g., privacy.

Appendix B

Theories, models, or frameworks used in reviewed studies

Cognition-related theories	Studies reviewed
Cognitive load theory (Chandler and Sweller, 1991)	Zhao, Zhang, Chu, Zhu, Hu, He et al. (2022); Chen and Wu (2015); Wu, Tzeng, and Huang (2014); Chen and Sun (2012); Chen and Wang (2011);
Dual-coding theory (Clark and Paivio, 1991)	Chen and Wang (2011); Chen and Sun (2012)
Embodied cognition theory (Gibbs, 2006)	Geršak et al. (2020)
Cognitive theory of multimedia learning (Mayer and Moreno, 1998)	Chen and Wu (2015)
Person-environment fit model (Caplan and Van Harrison, 1993)	Al-Fudail and Mellar (2008)
Self-regulated learning (Zimmerman, 1989)	Malmberg et al. (2019)
ARCS model of motivation (Keller, 1987)	Wu et al. (2014)
Media richness theory (Daft and Lengel, 1986)	Chen and Wu (2015)
Social identity theory (Tajfel and Turner, 1979)	DeFalco et al. (2018)
Holist/serialist cognitive style (Pask, 1976)	Huang et al. (2016)
Emotion-related theories	**Studies reviewed**
Russell's circumplex model (Russell, 1980)	Chen, Xie, Li, and Wang (2022); Ketonen, Salonen, Lonka, and Salmela-Aro (2022); Törmänen et al. (2021); Wampfler et al. (2019); Pijeira-Díaz, Kirschner, Järvelä, and Chsler (2018); Villanueva, Campbell, Raikes, Jones, and Putney (2018); Harley, Bouchet, Hussain, Azevedo, and Calvo (2015); Shen et al. (2009); Shen et al. (2008);
Polyvagal theory (Porges, 2001)	Mastromatteo, Zaccoletti, Mason, and Scrimin (2021); Chen and Wu (2015); Chen and Sun (2012);
Theory of constructed emotions (Barrett, 2017)	Mastromatteo et al. (2021); Törmänen et al. (2021)
Neurovisceral integration theory (Thayer and Lane, 2000)	Chen and Wu (2015); Chen and Sun (2012);
Transactional theory of stress (Lazarus, 1974)	Mastromatteo et al. (2021); Al-Fudail and Mellar (2008);
Biopsychosocial model (Blascovich and Mendes, 2001)	Mastromatteo et al. (2021)
PAD psychological model (Mehrabian, 1996)	Antoniou et al. (2020)
Frustration theory(Amsel, 1992)	DeFalco et al. (2018)
State-trait model of anxiety (Spielberger, 1966)	Eidlin Levy and Rubinsten (2021)
Emotion & Cognition theories	**Studies reviewed**
The control-value theory of achievement emotions (Pekrun, 2006)	Törmänen et al. (2021); Strohmaier et al. (2020); Harley, Liu, Ahn, Lajoie, and Grace (2020); Harley, Jarrell, and Lajoie (2019); Villanueva et al. (2018); DeFalco et al. (2018); Harley et al. (2015)
Affective e-learning model (Shen and Shen, 2005)	Ray and Chakrabarti (2016); Cheng and Liu (2015); Shen et al. (2009); Shen et al. (2008)
Kort's learning spiral model (Kort et al., 2001)	Shen et al. (2009); Shen et al. (2008)
Yerkes-Dodson law (Yerkes and Dodson, 1908)	Costa, Guimbretière, Jung, and Choudhury (2019); Pijeira-Díaz et al. (2018)
Cognitive-motivational model (Pijeira-Díaz et al., 2018)	Pijeira-Díaz et al. (2018)
Extended emotion process model (Gross, 2015)	Harley et al. (2019)
Pedagogical framework for online tutoring (Basawapatna et al., 2013)	Standen et al. (2020)
Generalized intelligent framework for tutoring (Sottilare et al., 2012)	DeFalco et al. (2018)
Attentional control theory (Eysenck et al., 2007)	Costa et al. (2019)
Broaden-and-build theory (Fredrickson, 2004)	Lai, Liu, Liu, and Huang (2016)
The theory of cognitive disequilibrium (Graesser and Olde, 2003)	Cheng and Liu (2015)
Cognition-arousal theory of emotion (Schachter and Singer, 1962)	Pijeira-Díaz et al. (2018)

Appendix C

Emotions and signals measured in reviewed studies

Emotions	Signals	Studies reviewed
Engagement (12)	HR (3)	Shen et al. (2008); Cheng and Liu (2015); Antoniou et al. (2020)
	BVP (4)	Shen et al. (2008); Shen et al. (2009); Gillies et al. (2016); Di Lascio et al. (2018)
	ECG (1)	Cheng and Liu (2015)
	EEG (5)	Shen et al. (2008); Shen et al. (2009); Antoniou et al. (2020); Zhao et al. (2022); Apicella et al. (2022)
	EDA (9)	Shen et al. (2008); Shen et al. (2009); McNeal et al. (2014); Cheng and Liu (2015); Gillies et al. (2016); Di Lascio et al. (2018); DeFalco et al. (2018); Geršak et al. (2020); Antoniou et al. (2020)
	ST (5)	McNeal et al. (2014); Gillies, Carroll, Cunnington, Rafter, Palghat, & Bednark et al. (2016); Di Lascio et al. (2018); DeFalco et al. (2018); Geršak et al. (2020)
	Con (1)	Antoniou et al. (2020)
	Mov (6)	McNeal et al. (2014); Gillies et al. (2016); Di Lascio et al. (2018); DeFalco et al. (2018); Geršak et al. (2020); Antoniou et al. (2020)
Anxiety (9)	HR (3)	Costa et al. (2019); Trotman et al. (2019); Ketonen et al. (2022)
	EDA (5)	Harley et al. (2015); Villanueva et al. (2018); Harley et al. (2019); Strohmaier et al. (2020); Eidlin Levy and Rubinsten (2021)
	Mov (1)	Ketonen et al. (2022)
Positive (8)	HR (5)	Chen and Wang (2011); Chen and Sun (2012); Cowley et al. (2013); Wu et al. (2014); Chen and Wu (2015)
	BVP (1)	Jiang, Hassan, Chen, and Liu (2020)
	ECG (1)	Cowley et al. (2013)
	EEG (2)	Lin, Su, Chao, Hsieh, and Tsai (2016); Jiang et al. (2020)
	fEMG (1)	Cowley et al. (2013)
	EDA (3)	Lin et al. (2016); Malmberg et al. (2019); Törmänen et al. (2021)
Boredom (8)	HR (2)	Cheng and Liu (2015); Ketonen et al. (2022)
	BVP (1)	Shen et al. (2009)
	ECG (1)	Cheng and Liu (2015)
	Pulse (1)	Huang et al. (2016)
	EEG (1)	Shen et al. (2009)
	EDA (5)	Shen et al. (2009); Cheng and Liu (2015); Harley et al. (2015); DeFalco et al. (2018); Villanueva et al. (2018)
	ST (1)	DeFalco et al. (2018)
	Mov (3)	DeFalco et al. (2018); Standen et al. (2020); Ketonen et al. (2022)
Negative (7)	HR (4)	Chen and Wang (2011); Chen and Sun (2012); Wu et al. (2014); Chen and Wu (2015)
	EEG (1)	Lin et al. (2016)
	EDA (3)	Lin et al. (2016); Malmberg et al. (2019); Törmänen et al. (2021)
Happy/Joy (7)	HR (1)	Ray and Chakrabarti (2016)
	BVP (2)	Ray and Chakrabarti (2016); Jang, Byun, Park, and Sohn (2019)
	ECG (1)	Jang et al. (2019)
	Pulse (1)	Chen and Lee (2011)
	EEG (1)	Lai et al. (2016)
	EDA (4)	Harley et al. (2015); Ray and Chakrabarti (2016); Villanueva et al. (2018); Jang et al. (2019)
	ST (1)	Jang et al. (2019)
	Con (1)	Zhang et al. (2018)
Arousal (7)	HR (1)	Wampfler et al. (2019)
	BVP (2)	Gillies et al. (2016); Di Lascio et al. (2018)
	ECG (2)	Mastromatteo et al. (2021); Chen et al. (2022)
	EDA (7)	Gillies et al. (2016); Di Lascio et al. (2018); Pijeira-Díaz et al. (2018); Wampfler et al. (2019); Geršak et al. (2020); Eidlin Levy and Rubinsten (2021); Chen et al. (2022)
	ST (4)	Gillies et al. (2016); Di Lascio et al. (2018); Wampfler et al. (2019); Geršak et al. (2020)
	Mov (4)	Gillies et al. (2016); Di Lascio et al. (2018); Wampfler et al. (2019); Geršak et al. (2020)
Confusion (6)	HR (2)	Shen et al. (2008); Cheng and Liu (2015)
	BVP (2)	Shen et al. (2008); Shen et al. (2009)
	ECG (1)	Cheng and Liu (2015)
	EEG (2)	Shen et al. (2008); Shen et al. (2009)
	EDA (6)	Shen et al. (2008); Shen et al. (2009); Cheng and Liu (2015); Harley et al. (2015); DeFalco et al. (2018); Villanueva et al. (2018)
	ST (1)	DeFalco et al. (2018)
	Mov (1)	DeFalco et al. (2018)
Frustration (6)	HR (2)	Shen et al. (2008); Cheng and Liu (2015)
	BVP (1)	Shen et al. (2008)
	ECG (1)	Cheng and Liu (2015)
	EEG (1)	Shen et al. (2008)
	EDA (5)	Shen et al. (2008); Cheng and Liu (2015); Harley et al. (2015); DeFalco et al. (2018); Villanueva et al. (2018)
	ST (1)	DeFalco et al. (2018)

(*continued on next page*)

(*continued*)

	Mov (2)	Standen et al. (2020); DeFalco et al. (2018)
Stress (7)	HR (4)	Trimmel, Atzlsdorfer, Tupy, and Trimmel (2012); Larra, Schulz, Schilling, Ferreira de Sá, Best, and Kozik et al. (2014); Trotman et al. (2019); Liapis, Maratou, Panagiotakopoulos, Katsanos, and Kameas (2022)
	BVP (1)	Larra et al. (2014)
	Pulse (1)	Santos, Uria-Rivas, Rodriguez-Sanchez, and Boticario (2016)
	EDA (5)	Al-Fudail and Mellar (2008); Trimmel et al. (2012); Santos, Uria-Rivas, et al. (2016); Sano, Taylor, McHill, Phillips, Barger, & Klerman et al. (2018); Liapis et al. (2022)
	ST (2)	Santos, Uria-Rivas, et al. (2016); Sano et al. (2018)
	Resp (1)	Santos, Uria-Rivas, et al. (2016)
	Con (1)	Sano et al. (2018)
	Mov (1)	Sano et al. (2018)
Anger (6)	HR (1)	Ray and Chakrabarti (2016)
	BVP (2)	Ray and Chakrabarti (2016); Jang et al. (2019)
	ECG (1)	Jang et al. (2019)
	EDA (5)	Harley et al. (2015); Ray and Chakrabarti (2016); Villanueva et al. (2018); Harley et al. (2019); Jang et al. (2019)
	ST (1)	Jang et al. (2019)
	Con (1)	Zhang et al. (2018)
Fear/Scared (6)	HR (1)	Ray and Chakrabarti (2016)
	BVP (2)	Ray and Chakrabarti (2016); Jang et al. (2019)
	ECG (1)	Jang et al. (2019)
	EEG (1)	Lai et al. (2016)
	EDA (4)	Harley et al. (2015); Ray and Chakrabarti (2016); Villanueva et al. (2018); Jang et al. (2019)
	ST (1)	Jang et al. (2019)
	Con (1)	Zhang et al. (2018)
Surprise (6)	HR (1)	Ray and Chakrabarti (2016)
	BVP (2)	Ray and Chakrabarti (2016); Jang et al. (2019)
	ECG (1)	Jang et al. (2019)
	EDA (5)	Harley et al. (2015); Ray and Chakrabarti (2016); DeFalco et al. (2018); Villanueva et al. (2018); Jang et al. (2019)
	ST (2)	DeFalco et al. (2018); Jang et al. (2019)
	Con (1)	Zhang et al. (2018)
	Mov (1)	DeFalco et al. (2018)
Neutral (5)	HR (1)	Cheng and Liu (2015)
	BVP (1)	Jiang et al. (2020)
	ECG (1)	Cheng and Liu (2015)
	EEG (1)	Jiang et al. (2020)
	EDA (4)	Cheng and Liu (2015); Harley et al. (2015); Malmberg et al. (2019); Törmänen et al. (2021)
Hope (5)	HR (1)	Shen et al. (2008)
	BVP (2)	Shen et al. (2008); Shen et al. (2009)
	EEG (2)	Shen et al. (2008); Shen et al. (2009)
	EDA (5)	Shen et al. (2008); Shen et al. (2009); Harley et al. (2015); Villanueva et al. (2018); Harley et al. (2019)
Disgust (5)	HR (1)	Ray and Chakrabarti (2016)
	BVP (2)	Ray and Chakrabarti (2016); Jang et al. (2019)
	ECG (1)	Jang et al. (2019)
	EEG (1)	Lai et al. (2016)
	EDA (3)	Harley et al. (2015); Ray and Chakrabarti (2016); Jang et al. (2019)
	ST (1)	Jang et al. (2019)
	Con (1)	Zhang et al. (2018)
Sadness (5)	HR (1)	Ray and Chakrabarti (2016)
	BVP (2)	Ray and Chakrabarti (2016); Jang et al. (2019)
	ECG (1)	Jang et al. (2019)
	EEG (1)	Lai et al. (2016)
	EDA (3)	Harley et al. (2015); Ray and Chakrabarti (2016); Jang et al. (2019)
	ST (1)	Jang et al. (2019)
	Con (1)	Zhang et al. (2018)
Peacefulness (4)	HR (3)	Chen and Wang (2011); Chen and Sun (2012); Wu et al. (2014)
	Pulse (1)	Chen and Lee (2011)
Nervousness (3)	Pulse (3)	Chen and Lee (2011); Huang et al. (2016); Santos, Saneiro, Boticario, and Rodriguez-Sanchez (2016)
	EDA (1)	Santos, Saneiro, et al. (2016)
	ST (1)	Santos, Saneiro, et al. (2016)
Enjoyment (3)	EDA (3)	Harley et al. (2015); Villanueva et al. (2018); Harley et al. (2019)

Note: Resp-respiration, Con-context, Mov-movement; Numbers in brackets next to each signal indicates the number of studies that used the signals.

Appendix D

Ethic issues mentioned in the reviewed studies.

Ethical consideration	Reviewed studies
Transparency	Di Lascio et al. (2018); Geršak et al. (2020); Harley et al. (2020); Huang et al. (2016); Jang et al. (2019); Jiang et al. (2020); Larra et al. (2014); Eidlin Levy and Rubinsten (2021); McNeal et al. (2014); Pijeira-Díaz et al. (2018); Sano et al. (2018); Standen et al. (2020); Strohmaier et al. (2020); Törmänen et al. (2021); Trotman et al. (2019); Ketonen et al. (2022); Apicella et al. (2022)
Privacy	Di Lascio et al. (2018); Wampfler et al. (2019); Zhang et al. (2018); Ketonen et al. (2022)
Fairness/bias	McNeal et al. (2014); Strohmaier et al. (2020); Zhang et al. (2018)

References

Al-Fudail, M., & Mellar, H. (2008). Investigating teacher stress when using technology. *Computers & Education, 51*(3), 1103–1110. https://doi.org/10.1016/j.compedu.2007.11.004

Amsel, A. (1992). Frustration theory: Many years later. *Psychological Bulletin, 112*(3), 396. https://doi.org/10.1037/0033-2909.112.3.396

Antoniou, P. E., Arfaras, G., Pandria, N., Athanasiou, A., Ntakakis, G., Babatsikos, E., et al. (2020). Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study. *JMIR Serious Games, 8*(3), e17823. https://doi.org/10.2196/17823. e17823.

Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., & Pollastro, A. (2022). EEG-based measurement system for monitoring student engagement in learning 4.0. *Scientific Reports, 12*(1), 5857. https://doi.org/10.1038/s41598-022-09578-y, 5857.

Ayata, D., Yaslan, Y., & Kamasak, M. E. (2020). Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *Journal of Medical and Biological Engineering, 40*(2), 149–157. https://doi.org/10.1007/s40846-019-00505-7

Ba, S., Hu, X., Kong, R., & Law, N. (2022). Supporting adolescents' digital well-being in the post-pandemic era: Preliminary results from a multimodal learning analytics approach. In *International Conference on Advanced Learning Technologies (ICALT)* (pp. 177–179). IEEE. https://doi.org/10.1109/ICALT55010.2022.00059.

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review, 11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Basawapatna, A. R., Repenning, A., Koh, K. H., & Nickerson, H. (2013). The zones of proximal flow: Guiding students through a space of computational thinking skills and challenges. In *Proceedings of the ninth annual international ACM conference on* (pp. 67–74). International Computing Education Research. https://doi.org/10.1145/2493394.2493404.

Billinghurst, M., & Starner, T. (1999). Wearable devices: New ways to manage information. *Computer, 32*(1), 57–64. https://doi.org/10.1109/2.738305

Blascovich, J., & Mendes, W. B. (2001). *Challenge and threat appraisals* (pp. 59–82). Cambridge: *Feeling and thinking: The role of affect in social contagion*.

Borup, J., West, R. E., & Graham, C. R. (2012). Improving online social presence through asynchronous video. *The Internet and Higher Education, 15*(3), 195–203. https://doi.org/10.1016/j.iheduc.2011.11.001

Buissink-Smith, N., Mann, S., & Shephard, K. (2011). How do we measure affective learning in higher education? *Journal of Education for Sustainable Development, 5*(1), 101–114. https://doi.org/10.1177/097340821000500113

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Caplan, R. D., & Van Harrison, R. (1993). Person-environment fit theory: Some history, recent developments, and future directions. *Journal of Social Issues, 49*(4), 253–275. https://doi.org/10.1111/j.1540-4560.1993.tb01192.x

Caruelle, D., Gustafsson, A., Shams, P., & Lervik-Olsen, L. (2019). The use of electrodermal activity (EDA) measurement to understand consumer emotions-A literature review and a call for action. *Journal of Business Research, 104*, 146–160. https://doi.org/10.1016/j.jbusres.2019.06.041

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Cheng, J., & Liu, G. (2015). Affective e-learning platform based on SVM. *Journal of Information and Computational Science, 12*(11), 4225–4234. https://doi.org/10.12733/jics20106026

Chen, C. M., & Lee, T. H. (2011). Emotion recognition and communication for reducing second-language speaking anxiety in a web-based one-to-one synchronous learning environment. *British Journal of Educational Technology, 42*(3), 417–440. https://doi.org/10.1111/j.1467-8535.2009.01035.x

Chen, C. M., & Sun, Y. C. (2012). Assessing the effects of different multimedia materials on emotions and learning performance for visual and verbal style learners. *Computers & Education, 59*(4), 1273–1285. https://doi.org/10.1016/j.compedu.2012.05.006

Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research, 33*(3), 244–255. https://doi.org/10.1016/j.lisr.2010.09.010

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121. https://doi.org/10.1016/j.compedu.2014.08.015

Chen, M., Xie, L., Li, C., & Wang, Z. (2022). Research on emotion recognition for online learning in a novel computing model. *Applied Sciences, 12*(9), 4236. https://doi.org/10.3390/app12094236

Chung, J. W. Y., So, H. C. F., Choi, M. M. T., Yan, V. C. M., & Wong, T. K. S. (2021). Artificial Intelligence in education: Using heart rate variability (HRV) as a biomarker to assess emotions objectively. *Computers & Education: Artificial Intelligence, 2*, Article 100011. https://doi.org/10.1016/j.caeai.2021.100011

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210. https://doi.org/10.1007/BF01320076

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Costa, J., Guimbretière, F., Jung, M., & Choudhury, T. (2019). BoostMeUp: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch. *Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(2), 1–23. https://doi.org/10.1145/3328911

Cowley, B., Ravaja, N., & Heikura, T. (2013). Cardiovascular physiology predicts learning effects in a serious game activity. *Computers & Education, 60*(1), 299–309. https://doi.org/10.1016/j.compedu.2012.07.014

Crompton, H., & Burke, D. (2018). The use of mobile learning in higher education: A systematic review. *Computers & Education, 123*, 53–64. https://doi.org/10.1016/j.compedu.2018.04.007

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease, 175*(9), 526–536. https://doi.org/10.1097/00005053-198709000-00004

D'mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems, 2*(4), 1–39. https://doi.org/10.1145/2395123.2395128

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22*(4), 53–61. https://doi.org/10.1109/MIS.2007.79

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., et al. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education, 28*(2), 152–193. https://doi.org/10.1007/s40593-017-0152-1

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(3), 1–21. https://doi.org/10.1145/3264913

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning, 34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Eidlin Levy, H., & Rubinsten, O. (2021). Numbers (but not words) make math anxious individuals sweat: Physiological evidence. *Biological Psychology, 165*, Article 108187. https://doi.org/10.1016/j.biopsycho.2021.108187. Article.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology, 58*(2), 342. https://doi.org/10.1037/0022-3514.58.2.342

Eliot, J. A., & Hirumi, A. (2019). Emotion theory in education research practice: An interdisciplinary critical literature review. *Educational Technology Research & Development, 67*(5), 1065–1084. https://doi.org/10.1007/s11423-018-09642-3

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336. https://doi.org/10.1037/1528-3542.7.2.336

Feidakis, M. (2016). A review of emotion-aware systems for e-learning in virtual environments. *Formative Assessment, Learning Data Analytics and Gamification*, 217–242. https://doi.org/10.1016/B978-0-12-803637-2.00011-7

Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., López, M. T., et al. (2016). Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics, 64*, 55–73. https://doi.org/10.1016/j.jbi.2016.09.015

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 359*(1449), 1367–1377. https://doi.org/10.1098/rstb.2004.1512

Geršak, V., Vitulić, H. S., Prosen, S., Starc, G., Humar, I., & Geršak, G. (2020). Use of wearable devices to study activity of children in classroom; Case study-Learning geometry using movement. *Computer Communications, 150*, 581–588. https://doi.org/10.1016/j.comcom.2019.12.019

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature, 529*(7584), 117–119. https://doi.org/10.1038/nj7584-117a

Gibbs, R. W. (2006). *Embodiment and cognitive science*. Cambridge University Press.

Gillies, R. M., Carroll, A., Cunnington, R., Rafter, M., Palghat, K., Bednark, J., et al. (2016). Multimodal representations during an inquiry problem-solving activity in a year 6 science class: A case study investigating cooperation, physiological arousal and belief states. *Australian Journal of Education, 60*(2), 111–127. https://doi.org/10.1177/0004944116650701

Goldstein, M. D., & Strube, M. J. (1994). Independence revisited: The relation between positive and negative affect in a naturalistic setting. *Personality and Social Psychology Bulletin, 20*(1), 57–64. https://doi.org/10.1177/0146167294201

Graesser, A. C. (2020). Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction, 70*, Article 101212. https://doi.org/10.1016/j.learninstruc.2019.05.009

Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95*(3), 524. https://doi.org/10.1037/0022-0663.95.3.524

Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry, 26*(1), 130–137. https://doi.org/10.1080/1047840X.2015.989751

Guerra-Bustamante, J., León-del-Barco, B., Yuste-Tosina, R., López-Ramos, V. M., & Mendo-Lázaro, S. (2019). Emotional intelligence and psychological well-being in adolescents. *International Journal of Environmental Research and Public Health, 16*(10), 1720. https://doi.org/10.3390/ijerph16101720

Harley, J. (2016). Chapter 5 - measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In *Emotions, technology, design, and learning* (pp. 89–114). Elsevier.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48*, 615–625. https://doi.org/10.1016/j.chb.2015.02.013

Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science, 47*(2), 151–180. https://doi.org/10.1007/s11251-018-09480-z

Harley, J. M., Liu, Y., Ahn, B. T., Lajoie, S. P., & Grace, A. P. (2020). Examining physiological and self-report indicators of empathy during learners' interaction with a queer history app. *British Journal of Educational Technology, 51*(6), 1920–1937. https://doi.org/10.1111/bjet.13019

Hasan, M. A., Noor, N. F. M., Rahman, S. S. B. A., & Rahman, M. M. (2020). The transition from intelligent to affective tutoring system: A review and open issues. *IEEE Access, 8*, 204612–204638. https://doi.org/10.1109/ACCESS.2020.3036990

Hu, X., Chen, J., & Wang, Y. (2021). University students' use of music for learning and well-being: A qualitative study and design implications. *Information Processing & Management, 58*(1), 102409. https://doi.org/10.1016/j.ipm.2020.102409

Hu, X., Li, F., & Kong, R.. Can Background Music Facilitate Learning?: Preliminary Results on Reading Comprehension. https://doi.org/10.1145/3303772.3303839.

Huang, Y. M., Hwang, J. P., & Chen, S. Y. (2016). Matching/mismatching in web-based learning: A perspective based on cognitive styles and physiological factors. *Interactive Learning Environments, 24*(6), 1198–1214. https://doi.org/10.1080/10494820.2014.978791

Imani, M., & Montazer, G. A. (2019). A survey of emotion recognition methods with emphasis on E-Learning environments. *Journal of Network and Computer Applications, 147*, Article 102423. https://doi.org/10.1016/j.jnca.2019.102423

Jang, E. H., Byun, S., Park, M. S., & Sohn, J. H. (2019). Reliability of physiological responses induced by basic emotions: A pilot study. *Journal of Physiological Anthropology, 38*(1), 15. https://doi.org/10.1186/s40101-019-0209-y, 15.

Jiang, M., Hassan, A., Chen, Q., & Liu, Y. (2020). Effects of different landscape visual stimuli on psychophysiological responses in Chinese students. *Indoor + Built Environment, 29*(7), 1006–1016. https://doi.org/10.1177/1420326X19870578

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2–10. https://doi.org/10.1007/BF02905780

Ketonen, E. E., Salonen, V., Lonka, K., & Salmela-Aro, K. (2022). Can you feel the excitement? Physiological correlates of students' self-reported emotions. *British Journal of Educational Psychology*, e12534. https://doi.org/10.1111/bjep.12534. e12534.

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*(4), 345–379. https://doi.org/10.1007/BF00992553

Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. *Proceedings IEEE international conference on advanced learning technologies*, 43–46. https://doi.org/10.1109/ICALT.2001.943850

Lai, C. H., Liu, M. C., Liu, C. J., & Huang, Y. M. (2016). Using positive visual stimuli to lighten the online learning experience through in class questioning. *International Review of Research in Open and Distance Learning, 17*(1), 23–41. https://doi.org/10.19173/irrodl.v17i1.2114

Larra, M. F., Schulz, A., Schilling, T. M., Ferreira de Sá, D. S., Best, D., Kozik, B., et al. (2014). Heart rate response to post-learning stress predicts memory consolidation. *Neurobiology of Learning and Memory, 109*, 74–81. https://doi.org/10.1016/j.nlm.2013.12.004

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion* (pp. 25–59). Sage Publications, Inc.

Lazarus, R. S. (1974). Psychological stress and coping in adaptation and illness. *The International Journal of Psychiatry in Medicine, 5*(4), 321–333. https://doi.org/10.2190/T43T-84P3-QDUR-7RTP

Lennarz, H. K., Hollenstein, T., Lichtwarck-Aschoff, A., Kuntsche, E., & Granic, I. (2019). Emotion regulation in action: Use, selection, and success of emotion regulation in adolescents' daily lives. *International Journal of Behavioral Development, 43*(1), 1–11. https://doi.org/10.1177/01650254187555

Li, F., & Hu, X. (2023). Background music for studying: A naturalistic experiment on music characteristics and user perception. *IEEE MultiMedia*, 1–8. https://doi.org/10.1109/MMUL.2023.3243209

Li, F., Wang, Z., Ng, T. D., & Hu, X. (2021). Studying with learners' own music: Preliminary findings on concentration and task load. *LAK21: 11th. International Learning Analytics and Knowledge Conference*, 613–619. https://doi.org/10.1145/3448139.3448206

Liapis, A., Maratou, V., Panagiotakopoulos, T., Katsanos, C., & Kameas, A. (2022). UX evaluation of open MOOC platforms: A comparative study between moodle and open edX combining user interaction metrics and wearable biosensors. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2022.2048674

Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. *Educational Psychology Review, 18*(4), 307–314. https://doi.org/10.1007/s10648-006-9028-x

Lin, H. C. K., Su, S. H., Chao, C. J., Hsieh, C. Y., & Tsai, S. C. (2016). Construction of multi-mode affective learning system: Taking affective design as an example. *Educational Technology & Society, 19*(2), 132–147. http://www.jstor.org/stable/jeductechsoci.19.2.132.

Liu, S., Liu, S., Liu, Z., Peng, X., & Yang, Z. (2022). Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement. *Computers & Education, 181*, Article 104461. https://doi.org/10.1016/j.compedu.2022.104461

Liu, Z., Ren, Y., Kong, X., & Liu, S. (2022). Learning analytics based on wearable devices: A systematic literature review from 2011 to 2021. *Journal of Educational Computing Research, 60*(6), 1514–1557. https://doi.org/10.1177/07356331211064780

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., & Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior, 96*, 235–245. https://doi.org/10.1016/j.chb.2018.06.030

Mastromatteo, L. Y., Zaccoletti, S., Mason, L., & Scrimin, S. (2021). Physiological responses to a school task: The role of student–teacher relationships and students' emotional appraisal. *British Journal of Educational Psychology, 91*(4), 1146–1165. https://doi.org/10.1111/bjep.12410

Mayer, R. E. (2020). Searching for the role of emotions in e-learning. *Learning and Instruction, 70*, Article 101213. https://doi.org/10.1016/j.learninstruc.2019.05.010

Mayer, R. E., & Moreno, R. (1998). A cognitive theory of multimedia learning: Implications for design principles. *Journal of Educational Psychology, 91*(2), 358–368. https://doi.org/10.1016/j.learninstruc.2014.02.004

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

McNeal, K. S., Spry, J. M., Mitra, R., & Tipton, J. L. (2014). Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course. *Journal of Geoscience Education, 62*(4), 655–667. https://doi.org/10.5408/13-111.1

McQuiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction, 18*(1–2), 81–123. https://doi.org/10.1007/s11257-007-9040-y

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology, 14*(4), 261–292. https://doi.org/10.1007/BF02686918

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Mora, M., Urdaneta, E., & Chaya, C. (2018). Emotional response to wine: Sensory properties, age and gender as drivers of consumers' preferences. *Food Quality and Preference, 66*, 19–28. https://doi.org/10.1016/j.foodqual.2017.12.015

Nourbakhsh, N., Chen, F., Wang, Y., & Calvo, R. (2017). Detecting users' cognitive load by galvanic skin response with affective interference. *ACM Transactions on Interactive Intelligent Systems, 7*(3), 1–20. https://doi.org/10.1145/2960413

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(4), 427–469. https://doi.org/10.1007/s40593-014-0029-5

Ometov, A., Shubina, V., Klus, L., Skibińska, J., Saafi, S., Pascacio, P., et al. (2021). A survey on wearable technology: History, state-of-the-art and current challenges. *Computer Networks, 193*, Article 108074. https://doi.org/10.1016/j.comnet.2021.108074. –.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71. n71.

Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology, 46*(2), 128–148. https://doi.org/10.1111/j.2044-8279.1976.tb02305.x

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Frenzel, A., Goetz, T., & Perry, R. (2007). Chapter 2 - the control-value theory of achievement emotions: An integrative approach to emotions in education. In *Emotion in education* (pp. 13–36). Elsevier. https://doi.org/10.1016/B978-012372545-5/50003-4.

Pham, P., & Wang, J. (2015). AttentiveLearner: Improving mobile MOOC learning via implicit heart rate tracking. *Artificial Intelligence in Education*, 367–376. https://doi.org/10.1007/978-3-319-19773-9_37

Picard, R. (1997). *Affective computing.* Cambridge, Mass: MIT Press.

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., et al. (2004). Affective learning — a manifesto. *BT Technology Journal, 22*(4), 253–269. https://doi.org/10.1023/B:BTTJ.0000047603.37042.33

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In *Proceedings of the sixth international Conference on learning analytics & knowledge* (pp. 64–73). https://doi.org/10.1145/2883851.2883897. April.

Pijeira-Díaz, H. J., Kirschner, P., Järvelä, S., & Chsler, H. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning, 34*(4), 397–408. https://doi.org/10.1111/jcal.12271

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review, 31*(2), 339–359. https://doi.org/10.1007/s10648-019-09473-5

Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology, 42*(2), 123–146. https://doi.org/10.1016/S0167-8760(01)00162-3

Que, Y., Zheng, Y., Hsiao, J. H., & Hu, X. (2023). Studying the effect of self-selected background music on reading task with eye movements. *Scientific Reports, 13*(1), 1704–1704 https://doi.org/10.1038/s41598-023-28426-1.

Ray, A., & Chakrabarti, A. (2016). Design and implementation of technology enabled affective learning using fusion of bio-physical and facial expression. *Educational Technology & Society, 19*(4), 112–125. http://www.jstor.org/stable/jeductechsoci.19.4.112.

Reichenberger, J., Schnepper, R., Arend, A. K., & Blechert, J. (2020). Emotional eating in healthy individuals and patients with an eating disorder: Evidence from psychometric, experimental and naturalistic studies. *Proceedings of the Nutrition Society, 79*(3), 290–299. https://doi.org/10.1017/S0029665120007004

Resnick, M. L., & Chircu, A. M. (2018). Wearable devices: Ethical challenges and solutions. In *Wearable technologies: Concepts, methodologies, tools, and applications* (pp. 1225–1243). IGI Global.

Rim, B., Sung, N. J., Min, S., & Hong, M. (2020). Deep learning in physiological signal data: A survey. *Sensors, 20*(4), 969. https://doi.org/10.3390/s20040969

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161. https://doi.org/10.1037/h0077714

Saganowski, S., Dutkowiak, A., Dziadek, A., Dzieżyc, M., Komoszyńska, J., Michalska, W., et al. (2020). Emotion recognition using wearables: A systematic literature review-work-in-progress. In *2020 IEEE international conference on pervasive computing and communications workshops* (pp. 1–6). IEEE. https://doi.org/10.1109/PerComWorkshops48775.2020.9156096.

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., et al. (2018). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. *Journal of Medical Internet Research, 20*(6), e210. https://doi.org/10.2196/jmir.9410. e210.

Santos, O. C., Saneiro, M., Boticario, J. G., & Rodriguez-Sanchez, M. C. (2016). Toward interactive context-aware affective educational recommendations in computer-assisted language learning. *New Review in Hypermedia and Multimedia, 22*(1–2), 27–57. https://doi.org/10.1080/13614568.2015.1058428

Santos, O. C., Uria-Rivas, R., Rodriguez-Sanchez, M. C., & Boticario, J. G. (2016). An open sensing and acting platform for context-aware affective support in ambient intelligent educational settings. *IEEE Sensors Journal, 16*(10), 3865–3874. https://doi.org/10.1109/JSEN.2016.2533266

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*(5), 379. https://doi.org/10.1037/h0046234

Schillinger, F. L., Mosbacher, J. A., Brunner, C., Vogel, S. E., & Grabner, R. H. (2021). Revisiting the role of worries in explaining the link between test anxiety and test performance. *Educational Psychology Review, 33*(4), 1887–1906. https://doi.org/10.1007/s10648-021-09601-0

Schmidt, P., Reiss, A., Dürichen, R., & Laerhoven, K. (2019). Wearable-based affect recognition-A review. *Sensors, 19*(19), 4079. https://doi.org/10.3390/s19194079

Seneviratne, S., Hu, Y., Nguyen, T., Lan, G., Khalifa, S., Thilakarathna, K., et al. (2017). A survey of wearable devices and challenges. *IEEE Communications Surveys and Tutorials, 19*(4), 2573–2620. https://doi.org/10.1109/COMST.2017.2731979

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine, 14*(2), 410–417. https://doi.org/10.1109/TITB.2009.2036164

Shen, L., Callaghan, V., & Shen, R. (2008). Affective e-Learning in residential and pervasive computing environments. *Information Systems Frontiers, 10*(4), 461–472. https://doi.org/10.1007/s10796-008-9104-5

Shen, L. P., & Shen, R. M. (2005). Ontology-based learning content recommendation. *International Journal of Continuing Engineering Education and Life Long Learning, 15* (3/6), 308.

Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. *Educational Technology & Society, 12*(2), 176–189. http://www.jstor.org/stable/jeductechsoci.12.2.176.

Shi, H., Yang, L., Zhao, L., Su, Z., Mao, X., Zhang, L., & Liu, C. (2017). Differences of heart rate variability between happiness and sadness emotion states: A pilot study. *Journal of Medical and Biological Engineering, 37*(4), 527–539. https://doi.org/10.1007/s40846-017-0238-0

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., et al. (2018). A review of emotion recognition using physiological signals. *Sensors, 18*(7), 2074. https://doi.org/10.3390/s18072074

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). In *Fundamental issues in defense training and simulation*. Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED).

Spielberger, C. D. (1966). Theory and research on anxiety. *Anxiety and Behavior, 1*(3), 3–20.

Standen, P. J., Brown, D. J., Taheri, M., Trigo, M. J. G., Boulton, H., Burton, A., et al. (2020). An evaluation of an adaptive learning system based on multimodal affect recognition for learners with intellectual disabilities. *British Journal of Educational Technology, 51*(5), 1748–1765. https://doi.org/10.1111/bjet.13010

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording*. USA: Oxford University Press.

Strohmaier, A. R., Schiepe-Tiska, A., & Reiss, K. M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. *Frontline Learning Research, 8*(1), 16–32. https://doi.org/10.14786/flr.v8i1.427

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations, 33*(47), 74.

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders, 61*(3), 201–216. https://doi.org/10.1016/S0165-0327(00)00338-4

Törmänen, T., Järvenoja, H., & Mänty, K. (2021). Exploring groups' affective states during collaborative learning: What triggers activating affect on a group level? *Educational Technology Research & Development, 69*(5), 2523–2545. https://doi.org/10.1007/s11423-021-10037-0

Trimmel, M., Atzlsdorfer, J., Tupy, N., & Trimmel, K. (2012). Effects of low intensity noise from aircraft or from neighbourhood on cognitive learning and electrophysiological stress responses. *International Journal of Hygiene and Environmental Health, 215*(6), 547–554. https://doi.org/10.1016/j.ijheh.2011.12.007

Trotman, G. P., Veldhuijzen van Zanten, J. J. C. S., Davies, J., Möller, C., Ginty, A. T., & Williams, S. E. (2019). Associations between heart rate, perceived heart rate, and anxiety during acute psychological stress. *Anxiety, Stress & Coping, 32*(6), 711–727. https://doi.org/10.1080/10615806.2019.1648794

Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research & Development, 69*(2), 1101–1133. https://doi.org/10.1007/s11423-021-09977-4

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open science now: A systematic literature review for an integrated definition. *Journal of Business Research, 88*, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

Villanueva, I., Campbell, B. D., Raikes, A. C., Jones, S. H., & Putney, L. G. (2018). A multimodal exploration of engineering students' emotions and electrodermal activity in design activities. *Journal of Engineering Education, 107*(3), 414–441. https://doi.org/10.1002/jee.20225

Wampfler, R., Klingler, S., Solenthaler, B., Schinazi, V., & Gross, M. (2019). Affective state prediction in a mobile setting using wearable biometric sensors and stylus. In *Proceedings of the 12th international conference on educational data mining* (pp. 198–207). Université du Québec; Polytechnique Montréal. https://doi.org/10.3929/ethz-b-000393912.

Wong, R. M., & Adesope, O. O. (2021). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review, 33*(2), 357–385. https://doi.org/10.1007/s10648-020-09545-x

Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology, 47*(6), 1304–1323. https://doi.org/10.1111/bjet.12324

Wu, C. H., Tzeng, Y. L., & Huang, Y. M. (2014). Understanding the relationship between physiological signals and digital game-based learning outcome. *Journal of Computers in Education, 1*(1), 81–97. https://doi.org/10.1007/s40692-014-0006-x

Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, M. N. B., Affal, H. B., & Hussin, N. B. (2019). Affective computing in education: A systematic review and future research. *Computers & Education, 142*, Article 103649. https://doi.org/10.1016/j.compedu.2019.103649

Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022). Scalability, sustainability, and ethicality of multimodal learning analytics. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 13–23). https://doi.org/10.1145/3506860.3506862

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zhang, X., Li, W., Chen, X., & Lu, S. (2018). MoodExplorer: Towards compound emotion detection via smartphone sensing. *Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(4), 1–30. https://doi.org/10.1145/3161414

Zhao, G., Zhang, L., Chu, J., Zhu, W., Hu, B., He, H., et al. (2022). An augmented reality based mobile photography application to improve learning gain, decrease cognitive load, and achieve better emotional state. *International Journal of Human-Computer Interaction, 39*(3), 643–658. https://doi.org/10.1080/10447318.2022.2041911

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*(3), 329. https://doi.org/10.1037/0022-0663.81.3.329