



The Validity of Heart Rate Variability (HRV) in Educational Research and a Synthesis of Recommendations

Hyun Jin Kim¹ · Yuyi Park¹ · Jihyun Lee¹ 

Accepted: 18 March 2024
© The Author(s) 2024

Abstract

This study provides a comprehensive overview of methodological aspects when using heart rate variability (HRV) measures in educational research. Following PRISMA 2020 guidelines, we searched four databases for relevant studies published until March 5, 2024. From the 48 studies reviewed, we extracted data across three analytical categories: (1) area of study interest and participant populations, (2) data collection and analysis methodologies, and (3) the concurrent and predictive validity of HRV measurement for educational research. Study quality was evaluated using QualSyst assessment criteria. Most studies measured stress and enlisted undergraduate students as participants. Data were predominately collected using wearable devices, measuring HRV for durations of less than 30 min, and in varied contexts, including during exams, while learning, and in experiments. The parameters analyzed varied within both time and frequency domains. HRV data had a moderate level of concurrent validity as a measure of stress in an educational context. The concurrent validity of HRV data for measuring attention remains uncertain with insufficient evidence. Limited correlations appeared between stress and performance. The findings, potentials, and limitations of HRV measures are discussed, and synthesized recommendations for educational research using HRV data are provided.

Keywords Heart rate variability (HRV) · Psychophysiological data · Concurrent validity · Predictive validity · Learner stress · Learner attention

✉ Jihyun Lee
leeji1@snu.ac.kr

Hyun Jin Kim
childkim21@snu.ac.kr

Yuyi Park
premisshu@snu.ac.kr


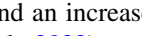
¹ Department of Dental Education, School of Dentistry & Dental Research Institute, Seoul National University, 103 Daehak-Ro, Jongno-Gu, Seoul 03080, Republic of Korea

A large body of research has documented a variety of learning outcomes affected by a wide spectrum of psychological factors, such as stress, anxiety, attention, and other diverse responses to lectures, exams, presentations, and normal school life. Stress, one of the most researched phenomena, has been found to have significant cognitive, psychological, social, and behavioral consequences for many learners, ultimately impacting academic success (Ahmady et al., 2021; Kwon et al., 2017; Lisnyj et al., 2020; Moreira et al., 2016; Ribeiro et al., 2018). Extensive research has revealed intricate relationships among emotional and cognitive variables and learning. These encompass test anxiety effects on educational outcomes (von der Embse et al., 2012; Zeidner, 2007, 2014); the influence of stress on attention, memory, and higher cognitive functions (Kofman et al., 2006; Redondo et al., 2019); and its impact on task performance in high-stakes environments (Martin & Naziruddin, 2020; Wetzel et al., 2010).

In traditional educational research, factors related to learning have typically been measured using self-report instruments (Weenk et al., 2018), which, while statistically confirmed for validity and reliability, have consistently faced criticism due to their inherent limitations (Williamson, 2007). Among the concerns is the subjective nature of self-reports, making them vulnerable to distortion. For example, respondents can misunderstand or misinterpret the literal or practical meaning of the survey questions (Hunt & Bhopal, 2004; Sudman et al., 1996). They can also result in mis-recalled information given that people often forget their experiences, even very significant ones, over surprisingly short periods (Schmier & Halpern, 2004; Wagenaar & Keren, 1986). Respondents must rely on their estimations, judgments, and perspectives (Brann et al., 2021; Chafouleas, 2011); may overestimate a certain state or trait (Williamson, 2007); or may answer in a way they perceive to be socially desirable (Krumpal, 2011). Further, the subjectivity inherent in self-report measures inevitably leads to reliability problems, since responses are notoriously changeable. Fatigue, anger, or broad situational changes also can cause response variation in the same person (Salinsky et al., 2001). Finally, self-report instruments typically inquire about learning experiences post-occurrence, presenting limitations in capturing learners' states in real time during the learning process and in monitoring these states over prolonged learning periods (Pekrun, 2006; Prokofieva et al., 2019; Weenk et al., 2018).

Researchers have recognized and addressed the limitations of self-report instruments by pioneering the measurement of psychophysiological variables as a direct and objective reflection of learners' cognitive and emotional states (e.g., Paas and Van Merriënboer (1994)). Psychophysiological measures assess the interplay between psychological and physiological conditions using a variety of tools and data, including electroencephalography (EEG) and event-related potentials (ERP), heart rate (HR) and heart rate variability (HRV), blood pressure (BP), electromyography (EMG), thermal imaging, and pupillometry (Gaffey & Wirth, 2014; Lohani et al., 2019). Physiological responses avoid the self-bias risks and are generally not distorted by respondents' recall, understanding, interpretations, judgments, or intentions. Given their increased objectivity, psychophysiological measures capture stable individual features, increasing the reliability of the collected data (Scrimin et al., 2019, 2021). With the development of newer wearable technologies, the use of

psychophysiological measures in educational research has become even more attractive. Such devices allow for real-time continuous monitoring of changes in individual psychological states, data that self-report pre- and post-surveys have only been able to collect to a limited degree (Castaldo et al., 2019; Huhn et al., 2022).

In an extension of clinical research, HRV is a psychophysiological measure increasingly being used in educational research. HRV is a relatively simple, affordable, noninvasive, and reliable measure that can capture changes in learner states over time (Castaldo et al., 2019; Laborde et al., 2017; Weenk et al., 2018). HRV represents the variability of the time interval between two consecutive heartbeats. It is a quantitative expression of balance in the functioning of the autonomic nervous system (ANS), also known as the sympathetic (signaling of a body in a state of alertness) and parasympathetic (signaling a body in a state of relaxation) modulation (Huhn et al., 2022; Prichard et al., 2012; Tracy et al., 2016; Wimmer et al., 2019). In the presence of high demands, like stressors, the sympathetic nervous system predominates over the parasympathetic, leading to an increased heart rate (HR) and a decreased HRV (). Conversely, dominant parasympathetic nervous activity reflecting a decreased HR and an increased HRV () is associated with a state of relaxation (Huhn et al., 2022).

HRV is an index of the complex mix of physiological, emotional, and cognitive processes involved in self-regulation and adaptability due to brain–heart connections (Laborde et al., 2017; Malik, 1996). These connections work via two pathways: (1) inhibitory pathways via GABA (gamma-aminobutyric acid, a neurotransmitter with a calming effect on the nervous system) from the prefrontal cortex to the amygdala and (2) additional regulatory pathways by the amygdala on ANS through the production of several different hormones and neurotransmitters including dopamine, serotonin, cortisol, adrenaline, and noradrenaline. These two pathways modulate heart rate and thus HRV. Hence, HRV is influenced by a range of factors involved in physiological, emotional, and cognitive regulation, including physiological factors such as cardiac function, metabolic processes, hormonal regulation, respiration rate, and age; emotional factors such as stress, depression, anxiety, and anger; and cognitive factors such as working memory, attention, and executive function. All these factors are closely interrelated with inhibitory GABAergic pathways and the hormones and neurotransmitters produced from amygdala. High HRV implies a balanced ANS in a state of robust physiological, emotional, and cognitive regulation, allowing learners to respond to varied external stimuli with flexibility and resilience (Yee Chung et al., 2021), while lower HRV can be a sign of unbalanced ANS implying physiological, emotional, and/or cognitive issues.

HRV is typically measured using an electrocardiogram (ECG), which records the electrical activity of the heart. HRV analysis includes the time domain, frequency domain, and non-linear domain parameters. Time domain parameters are the simplest form of HRV analysis, involving direct measurements of the time intervals between heartbeats. Common metrics include the standard deviation of normal-to-normal (NN) intervals (SDNN) and the root mean square of successive differences (RMSSD). Frequency domain analysis examines the spectral components of HRV, decomposing the HR signal into frequency bands such as high-frequency (HF) bands (0.15–0.40 Hz), low-frequency (LF) bands (0.04–0.15 Hz),

and very low-frequency (VLF) bands (0.003–0.04 Hz). The duration of HRV recordings can vary; short-term recordings lasting around 5 min are typical for quick assessments, while long-term recordings (over 24 h) offer a more comprehensive overview. Ensuring data quality in HRV measurement is crucial. In the medical context, HRV measurement standards were provided by the Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology in 1996, with recommendations for meticulous pre-processing, artifact-free ECG recordings, and accounting for age, gender, health status, current medications, and the environment during the recording (Malik, 1996).

Educational studies have been conducted using HRV data in diverse learning contexts to provide an objective understanding of learners' regulation mechanisms regarding emotional and cognitive phenomena (Forte et al., 2019; Lehrer et al., 2020). Studies on learner emotions have investigated the association between HRV and stress, anxiety, panic disorder, depression, emotion, empathy, burnout, and social competence (Dimitriev et al., 2008; Hammoud et al., 2018; Huhn et al., 2018; Melillo et al., 2011; Scrimin et al., 2018; Tharion et al., 2009). Similarly, studies of learner cognition have revealed a link between HRV and learner attention, working memory, executive functioning, problem-solving, general intelligence, and performance (Hansen et al., 2004; Melis & van Boxtel, 2001; Scrimin et al., 2018).

Since the use of HRV in educational research is still incipient, numerous challenges related to HRV measurement and interpretation should be addressed to provide sound guidelines for educational contexts. While HRV measurement presents an alternative and complementary method for assessing learners' psycho-cognitive states, its validity has yet to be established, particularly regarding sensitivity and specificity, crucial attributes of the validity of a measure. HRV measures must be sensitive enough to accurately identify the areas of interest in a study (e.g., stress and attention) and specific enough to accurately identify a state of disinterest (e.g., relaxation and distraction). Both sensitivity and specificity pertain to the validity and accuracy of a measure, along with the concurrent validity (the extent to which HRV correlates with similar measures assessed approximately at the same time, such as subjective self-reported data) and the predictive validity (the extent to which HRV predicts certain learning outcomes) of the HRV measure.

Concerns regarding HRV utilization in educational research are as follows. First, several factors can limit its validity, such as different physiological variables like health status, particularly respiratory, and cardiovascular diseases (Campbell & Ehlert, 2012; Malik, 1996; Redondo et al., 2019), which should be controlled through robust methodological approaches (Laborde et al., 2017). Second, despite the ease of data collection, the wide range of information collected by an HRV measure can be easily misconstrued, and difficulties with interpreting findings can limit real-world implications (Laborde et al., 2017; Scrimin et al., 2021). Although Malik's (1996) pivotal guidelines provided early standards for HRV data's clinical use, they are dated and primarily focused on clinical medicine, necessitating more relevant guides for current educational research contexts. Third, few studies have systematically examined the concurrent or predictive validity of HRV measures in educational research.

This study addresses these caveats and provides a comprehensive overview of the methodological aspects of using HRV measures in educational research. Accordingly, we posed the following research questions: (1) To what extent has HRV been used to measure stress (as a representative emotional area) and attention (as a representative cognitive area) in educational research, and for which learners have HRV data been used? (2) How has HRV data been collected and analyzed in educational research (device, measuring time, parameters, precautions, and analysis software)? (3) To what extent does HRV data demonstrate concurrent and predictive validity in relation to other data types, such as self-reported or achievement data, in educational research? We intend to provide educational researchers with practical recommendations and methodological considerations regarding HRV use.

Method

This review was designed, conducted, and reported in accordance with the updated PRISMA (Preferred Reporting Items for Systematic review and Meta-Analysis) 2020 statement and guideline (Page et al., 2020).

Search Strategy

We searched four databases judged to be most relevant to our topic: PubMed (National Library of Medicine; 1781~), PsycINFO (1887~), Web of Science (1990~), and ERIC (1966~). We initially performed database searches for all records up to March 5, 2024, using search terms in three categories: heart rate variability (heart rate variability* OR HRV), educational context (education OR educat* OR learn* OR student* OR teach*), and stress or attention (attention* OR stress*). Detailed search queries for each database are presented in Table S1 (*Search Queries* in Online Supplementary File 1). In addition, we performed a hand search for gray literature in Google Scholar. An experienced research librarian supervised the development and implementation of the search strategies.

Inclusion and Exclusion Criteria

Inclusion criteria initially comprised original English research articles that investigated stress or attention using HRV measurements in educational contexts. We excluded research with extremely young (below 5 years) or older age groups (above 65 years) based on existing research evidence regarding HRV age dependency (Antelmi et al., 2004; Bonnemeier et al., 2003; Michels et al., 2013; Voss et al., 2015; Zhang, 2007). We only reviewed studies involving healthy individuals with no physical or psychological problems. Our inclusion criteria are presented in Table 1. After removing duplicates, two authors independently reviewed the titles and abstracts of all articles. Disagreement was resolved by consulting a third author to reach a consensus.

Table 1 Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Time period	Published up to March, 5 th , 2024	
Language	English	Not English
Type of article	Peer-reviewed journal publications	Non-peer-reviewed publications
Type of study	Full-text research papers	Commentaries, editorials, book chapters, letters
Context	Educational context (e.g., exam, learning, clinical practice, experiments)	Not related to educational context (e.g., military, police, navy, machine learning, meditation, etc.)
Study focus	Studies using HRV data to measure learners' stress or attention in educational context	<ul style="list-style-type: none"> - Not related to stress and attention - Not provided details of HRV measurement, only for technical description
Participants	Healthy individuals aged from 5 to 65 years	<ul style="list-style-type: none"> - Not human - Unhealthy individuals with physical or psychological symptoms - Individuals younger than 4 or older than 66

Data were extracted from all eligible studies based on our research questions, resulting in the following data extraction categories and subcategories. More detailed data extracted by the classification scheme (as in Lee et al. (2020)) are provided in Table S2 (*Classification Scheme* in Online Supplementary File 1).

- Areas of interest (stress/attention), participants (sample size, age, and school level group), and participant control (excluded participants and participant instructions)
- HRV data collection and analysis (educational setting, HRV device, software, pre-processing, measurement standards compliance, baseline measure, measuring time, parameters, and data analysis and interpretation)
- Concurrent and predictive validity (concurrently measured psychophysiological, self-report, and performance data) and correlation level (L1/L2/L3/L4)

As per the areas of interest, data were extracted into either *stress* or *attention* or both. Due to the blurred definitions and mixed use of the terms, *stress* in this study is defined as *a complex psychological and physiological phenomenon that arises in response to perceived challenges or demands related to learning*. Stress in this study is associated with *specific situations or events rather than being chronic* in nature. It manifests as *anticipatory anxiety, increased mental*

and physiological arousal, and emotional tension. This phenomenon affects various aspects of an individual's *cognitive, emotional, behavioral, and physiological processes* (Blackburn & Epel, 2017; Fink, 2017). We coded both studies of stress and those of anxiety uniformly as studies of stress.

We defined *attention* as a *multifaceted cognitive function that involves the allocation of cognitive resources to stimuli or tasks. It encompasses processes such as focusing, sustaining concentration over time, managing cognitive load, and the ability to selectively prioritize certain stimuli over others* (Guo et al., 2020; Redondo et al., 2019). We coded studies with experiments or interventions that required attention, such as the Stroop test or psychomotor vigilance task (PVT), as studies of attention.

We extracted information on HRV data collection and analysis, including participants (sample size, age, school level) and participant control (excluded participants participant instructions); the educational setting and the presence of professors/teachers; the HRV devices and software used; the presence of pre-processing; compliance with measurement standards; the presence of baseline measurements; the duration of measuring time; the specific HRV parameters collected; and the methods of data analysis and interpretation.

Regarding the correlation level for examining concurrent validity of HRV measurement, we coded the extent of the correlation between HRV and other data into four levels according to the magnitude of Cohen's (1992) measure of effect size: level 1 (large effect size, $r > 0.5$), level 2 (medium and small effect size, $r < 0.3$ and $r < 0.1$, respectively), level 3 (r value was not reported, or no information on statistical correlation was provided), and level 4 (not correlated).

Using a data extraction form, we developed for this study through iterative testing and revision, the three authors worked independently. Disagreements were resolved through joint discussion.

Assessment of Study Quality

To assess the quality of the selected studies, we used the QualSyst Standard Quality Assessment Criteria by Kmet et al. (2004). This assessment tool was selected due to its applicability to evaluating the quality of quantitative studies, including non-randomized controlled trials. The criteria for evaluating quantitative studies are (1) clear question/objective, (2) appropriate study design, (3) proper subject/comparison group selection, (4) detailed subject characteristics, (5) random allocation in interventional studies, (6) investigator blinding, (7) subject blinding in interventional studies, (8) robust outcome and exposure measures, (9) appropriate sample size, (10) adequate analytic methods, (11) variance estimates for results, (12) control of confounding, (13) detailed results reporting, and (14) evidence-supported conclusions. These 14 criteria for quantitative studies are scored from 0 (lowest quality) to 2 (highest quality). The study quality was defined using a final mean score, with lower than 0.50 indicating inadequate quality, 0.50–0.70 indicating adequate, 0.71–0.80 indicating good, and higher than 0.80 indicating strong (Lee et al., 2008). We chose a conservative cutoff point of 0.71 for inclusion. The final quality scores ranged from 0.71 to 0.95, demonstrating that all included studies were of good quality.

Results

Trial Flow

A total of 1726 articles were identified from the four academic databases. None was identified in the gray literature. The first screening process eliminated 1025 duplicate articles, leaving 701 potentially relevant articles for title review. After 294 articles had been eliminated based on study titles, 407 potentially relevant articles remained for abstract review. A thorough abstract review further excluded 190 articles. The remaining 217 articles were retrieved for full-text review, resulting in an additional 169 being excluded. A final set of 48 articles were included in this systematic review. The exclusion process details and reasons for exclusion at each step are presented in Fig. 1.

Study Features

The 48 studies that met the inclusion criteria were published from 2006 to 2024 (March 5) and involved 2706 people in total. Most of the studies were quantitative ($n=43$, 89.6%), whereas a few studies used mixed methods ($n=5$, 10.4%), adopting both qualitative (interview) and quantitative approaches.

Areas of Interest and Participants

Of the studies, 44 out of 48 focused on stress (91.7%), while a few focused on attention ($n=12$, 25%). Eight out of 48 studies focused on both (Fig. 2). Participants were aged between 6 and 58 years, and the sample size of each study ranged from 5 to 160. While there were relatively few studies with fewer than 10 participants ($n=2$, 5.6%) or more than 100 participants ($n=6$, 12.5%), most studies had sample sizes between 10 and 50 ($n=22$, 45.8%) or between 50 and 100 ($n=18$, 37.5%). By school level, the largest group of studies used HRV with undergraduate students ($n=33$, 68.8%), half of whom were health sciences students, namely, medical ($n=15$), pharmacy/biotechnology ($n=1$), nursing students ($n=1$), and physiotherapy ($n=1$). Studies with participants in other school levels were equally represented: primary school students ($n=3$, 8.3%), secondary school students ($n=3$, 8.3%), and postgraduate students or professionals ($n=4$, 8.3%). All the postgraduates or professionals were interns, residents, surgeons, or medical staff from medical fields.

Of the 48 studies, 27 (56.3%) controlled for participant factors, either excluding participants (marked as *Excl.* in Table S3 (*Details of Reviewed Studies* in Online Supplementary File 2)) ($n=30$) or providing instructions (marked as *Instr.* in Table S3) ($n=13$), or both ($n=16$). Exclusions were most common for cardiovascular diseases (e.g., hypertension, heart diseases, coronary artery disease, arrhythmias, and tachycardia; $n=12$), followed by mental health issues ($n=13$), obesity or weight/BMI-related criteria (BMI > 25 or 30 kg/m², BMI < 18 or 18.5 kg/m²; $n=7$), and endocrine diseases including metabolic disorders (e.g., diabetes; $n=8$). Criteria limiting smoking varied: Seven studies included only

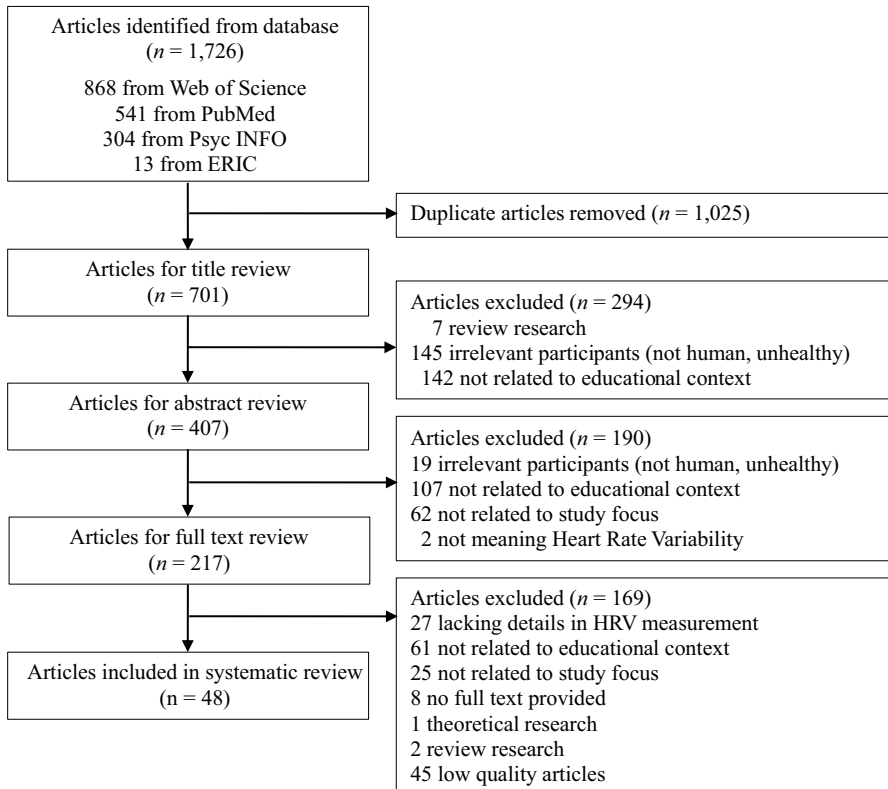


Fig. 1 Trial flow for this systematic review

nonsmokers, five studies restricted participants from smoking before measuring HRV, and one study did not restrict participants from smoking due to the risk of withdrawal symptoms. Because hormonal changes can affect the cardiac ANS, four studies limited participants based on women's menstrual cycles and only included females in the proliferative menstrual phase. Two studies recruited only males. Studies varied in participant exclusion based on medication ($n = 14$); some barred any regular or long-term medication users, and others excluded those on cardiovascular drugs, or required a medication-free period prior to HRV data collection. Instructions given before HRV measurement ($n = 17$, 35.4%) included avoiding caffeine ($n = 15$, 88.2% of instruction studies), alcohol ($n = 11$), and intense exercise ($n = 4$) and ensuring adequate sleep ($n = 3$). Eight controlled for noise, seven for temperature (e.g., 23 °C), four for illumination, and three for humidity (e.g., 45–55%). Seven studies specified the time frame for measuring HRV (e.g., between 9:00 and 12:30 pm after breakfast), and 11 studies asked participants to maintain their posture during HRV measurements (e.g., prone and supine).

HRV Data Collection and Analysis

Data Collection Context

Studies collected HRV data from varied contexts, which were classified into three categories: exam ($n = 11$, 22.9%), learning ($n = 22$, 45.8%), and experiments ($n = 23$, 47.9%; Fig. 2). The exam category included written exams, oral exams, and skill assessments; learning included face-to-face lectures, online lectures, clinical practice, and peer tutoring; experiments included cognitive tasks (e.g., Stroop test, mathematics, or memory), psychomotor tasks, and educational intervention (e.g., stress management program and video stressors). Stress measurements were predominantly conducted in examination ($n = 11$) and learning contexts ($n = 22$), whereas attention was primarily assessed in experimental settings ($n = 12$), with two of these studies also measuring attention in a learning context.

Device and Measuring Time

Among the 48 studies analyzed, 26 (54.2%) explicitly reported adherence to Malik's (1996) guidelines. One study reported compliance with both Malik (1996) and Laborde et al. (2017). Although those remaining 22 studies are presumed to generally follow the recommended steps (e.g., baseline measurement and pre-processing) as suggested in the standard, they did not explicitly mention this compliance. Measuring devices were either wearable or static. Of the studies, 21 (43.8%) were conducted using wristband ($n = 11$, 22.9%) or chest belt ($n = 13$, 27.1%) wearable devices, and 3 studies involved the use of both. Twenty-eight studies (58.3%) used static devices, among which one study additionally employed a chest belt. The Polar series ($n = 15$, 31.3%) was the most used HRV device, and Kubios ($n = 19$, 39.6%) was the predominant HRV analysis software. Software like Kubios, MATLAB, and some from Polar offered pre-processing features, although their usage was variably

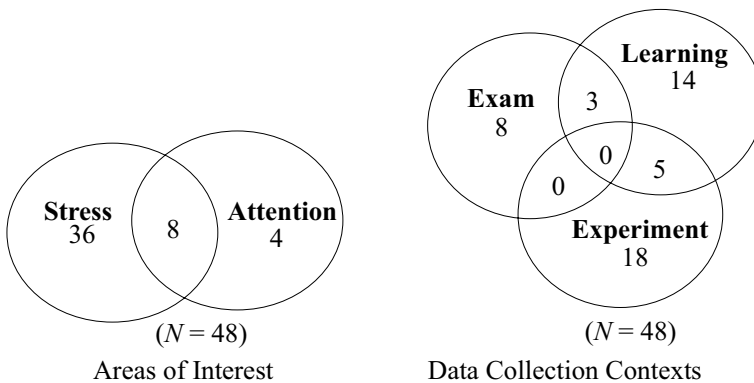


Fig. 2 Area of interest and educational settings of the included studies

reported (see Table S2 (*Classification Scheme* in Online Supplementary File 1) for all devices and software used in the studies).

For pre-processing, nearly one third of the studies explicitly mentioned their engagement in this process ($n=19$, 39.6%), primarily through software functions. Although the majority did not specifically report on pre-processing, it can be inferred that they implemented it, as indicated by the software used ($n=17$, 48.6%). Altogether, approximately 80% of the studies are estimated to have involved pre-processing. Most studies collected HRV data for less than 30 min ($n=37$, 77.1%), and seven studies collected HRV data for more than 30 min, even up to several days (14.6%). The measurement duration in nine studies depended on the completion of a designed task (18.9%). More details regarding the research designs associated with the different HRV measurement durations are provided in Table S3 (*Details of Reviewed Studies* in Online Supplementary File 2).

Parameters

HRV data can be analyzed using different methods: time domain analysis, frequency domain analysis, and non-linear domain analysis. Of the studies, 38, 38, and 11 studies adopted each of the analysis methods, respectively. The most frequently analyzed time domain parameters were RMSSD ($n=29$, 60.4%), indicating parasympathetic activation; SDNN ($n=15$, 31.3%), the standard deviation of normal-to-normal R-R intervals; and pNN50 ($n=12$, 25%), percentage of differences between normal-to-normal R-R intervals greater than 50 ms. In the frequency domain, the most frequently analyzed parameters were HF ($n=32$, 66.7%), which reflects parasympathetic activity, LF/HF ($n=24$, 50.0%), which reflects sympathovagal balance, and LF ($n=26$, 54.2%), which reflects sympathetic activity. The non-linear domain was analyzed relatively rarely. Eight papers investigated SD1 and SD2 non-linear domain parameters, which mirror short-term and long-term HRV variability. LF/HF in the frequency domain and SDNN and pNN50 in the time domain were more frequently used in the context of measuring stress than measuring attention (see Table S2 (*Classification Scheme* in Online Supplementary File 1) for the complete parameter frequencies used in the studies).

Analysis and Interpretation

When interpreting HRV parameters, most of the studies evaluated the relative value—i.e., increases or decreases—in the collected data rather than considering absolute standards as used in a clinical context. Forty out of 48 examined changes from baseline values (83.3%), and ten studies (20.8%) compared HRV data between experimental and control groups. Studies considered participants' age, gender, health status, and length of recording holistically (Xhyheri et al., 2012). None of the reviewed studies considered the normal range HRV parameter standards suggested and widely used in clinical settings.

Concurrent and Predictive Validity of HRV Measurement

Concurrently Measured Data

In all reviewed studies, HRV was not the only data collected; other psychophysiological, self-reported, and performance data were also collected. Of the 48 studies, 35 (72.9%) collected other psychophysiological data such as heart rate, cortisol levels, or blood pressure, and 35 studies ($n = 72.9\%$) collected self-report data, such as stress scale scores, affect scale scores, or cognitive load scores, which we used to evaluate the concurrent validity of the HRV data. Eighteen (37.5%) studies collected performance data, such as test scores or skill performance scores, from which we evaluated the predictive validity of the HRV data.

Regarding psychophysiological measures, many stress studies ($n = 44$) concurrently collected heart rate ($n = 25$, 56.8%) and cortisol ($n = 12$, 27.3%) data, as did attention studies; four collected heart rate data (33.3%) and two collected cortisol data (16.7%) along with HRV data.

Validated self-reported stress scales such as the STAI ($n = 10$) or Subjective Perception of Distress Scale (SUDS; $n = 3$) were used in many stress studies, and some studies used invalidated self-report scales ($n = 6$, 13.6%) to complement the HRV data ($n = 27$, 56.3%). HRV data used to measure attention was more often paired with cognitive performance data ($n = 5$, 41.7%) than with self-reported scale data ($n = 1$, 8.3%).

Among the performance data used by studies, grades ($n = 4$, 9.0%), skill assessment ($n = 3$, 6.8%), and surgical training ($n = 3$, 6.8%) were primarily used to measure stress, whereas only cognitive performance scores were primarily used to measure attention ($n = 5$, 41.7%). Studies reported other diverse psychological data, including affect, depression, personality traits, empathy, resilience, mindfulness, or burnout, as detailed in Table S2 (*Classification Scheme* in Online Supplementary File 1).

Correlation Between HRV and Other Data: Concurrent Validity and Predictive Validity

To assess the concurrent validity of HRV data used to measure stress or attention in an educational context, one of four correlation levels (L1, L2, L3, and L4) was coded for each HRV-variable pair (studies had more than one pair). Level 1 (large effect size, $r > 0.5$) and level 2 (medium and small effect sizes, $r < 0.3$ and $r < 0.1$, respectively) correlations were considered as one group to contrast with level 4 correlations (not correlated). For stress measurement, there were 14 L1/L2 pairs and 8 L4 pairs, implying that the concurrent validity of HRV data used to measure stress has moderate support. For attention measurement, there was one L1/L2 pair and two L4 pairs, the sample size is notably small, suggesting that there is insufficient evidence to support the concurrent validity of HRV data for measuring attention.

The top three data types assessed were heart rate, cortisol, and validated stress scales. The L1/L2 (high correlation) group included two studies with heart rate data, three studies with cortisol data, and eight studies with stress scale data. Heart rate is known to relate to HRV (Kazmi et al., 2016; Sacha, 2014); 25 studies out of 48

co-measured heart rates of participants, among which only two studies (8.0%) pertained to the L1/L2 group. Among the 12 studies that collected cortisol data, three studies (one that used blood samples and one that used saliva samples) reported a high correlation with HRV data (RMSSD, LF, HF), but one study showed that the parameter LF/HF had no correlation with cortisol data. The remaining studies that used psychophysiological measurements corresponded to L3, implying that most was contextually consistent with HRV data.

Among the 22 studies that employed validated stress scales, 8 (36.4%) reported a high correlation between HRV data and stress scale scores. The instruments used in L1 studies were Escala de Valoración del Estado de Ánimo (Mood Assessment Scale; EVEA) and Test Anxiety Inventory (TAI). The parameters used were balanced between time domain (RMSSD, NN50, SDNN) and frequency domain (LF, HF, LF/HF) data. EVEA scores highly correlated with HRV RMSSD, LF, and HF data but did not correlate with LF/HF data. Visual Analogue Scale (VAS) scores highly correlated with HRV data in students' passive conditions, but no correlation with HRV data was found for students' active or interactive conditions.

In attention studies, the concurrent validity of HRV data was supported due to a high correlation between HRV data and cortisol data used to measure attention, but validity was weakened as there was no correlation between HRV data and respiration rate. The overall correlations relevant to concurrent validity are provided in Table 2.

To determine whether HRV data can predict academic or other performance, we considered the correlation between performance data and HRV data. Among the 21 studies that measured HRV and performance data, only 4 studies fell into the L1/L2 group, with 11 studies in the L4 group. Among stress studies, this trend was more pronounced. All three studies that co-measured test scores reported no correlation between test scores and HRV data, and two out of three studies that used skill assessment data provided evidence against the predictive validity of HRV. Two studies that used PVT test scores indicated low predictive validity of HRV data for measuring attention. The overall correlation information relevant to predictive validity is provided in Table 3. The study features by areas of interest, namely, stress and attention, and their proportional composition within the set of the reviewed studies are provided in Table 4.

Discussion

Given the scarcity of HRV guiding research for educational use, this review explored the potential, measurement guides and validity of HRV measures for educational research. By systematically reviewing 48 educational studies that utilized HRV data, we synthesized the literature into (Q1) areas of interest and learner groups, (Q2) methodologies for data collection and analysis, and (Q3) concurrent and predictive validity of HRV measurement for educational research. Addressing challenges related to HRV measurement in educational research as a method to complement self-report instruments, findings and implications by research questions are further discussed below.

Table 2 Concurrent validity

		Correlation Between		Context	Device (SW)	Time	Participants (n)	Study
STRESS	L1/L2 (n = 14)	Heart rate	SDNN	Learning	Static (N/R)	Task-dependent	Professionals (28)	[3]
		Heart rate	LF, HF, LF/HF	Exam	Static (N/R)	Short	Undergraduates (65)	[12]
		Cortisol (blood)	SDNN	Learning	Static (N/R)	Task-dependent	Professionals (28)	[3]
		Cortisol (salivary)	RMSSD, LF, HF	Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
		Cortisol (salivary)	RMSSD, pNNS0, LF, HF	Learning, Experiment	Wearable (Kubios)	Long	Undergraduates (82)	[42]
		Blood pressure	RMSSD, LF, HF	Experiment	Wearable (BioSign)	Short	Undergraduates (60)	[41]
		Stress Scale (EVEA)	RMSSD, LF, HF	Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
		Stress Scale (STAI-SF)	RR Interval, LF, HF, LF/HF	Learning	Static (Kubios)	Short	Undergraduates (21)	[25]
		Stress Scale (VAS) (Passive Condition)	RMSSD, LF/HF	Learning, Experiment	Static (Qhrv Assessment)	Short	K-12 Students (104)	[14]
		Stress Scale (KUSTA)	LF, HF, LF/HF	Exam	Static (N/R)	Short	Undergraduates (65)	[12]
		Stress Scale (AMOS 8-15)	NN50	Experiment	Static (Kubios)	Short	K-12 Students (91)	[26]
		Stress Scale (TAI)	LnRMSSD, LnLF, LnHF	Exam	Wearable (Biopac)	Short	Undergraduates (97)	[37]
		Stress Scale (SSQ)	VLF	Experiment	Static (MATLAB)	Short	Undergraduates (28)	[45]
		Stress Scale (Not validated)	SDNN	Learning	Static (N/R)	Task-dependent	Professionals (28)	[3]
		STRESS	L4 (n = 8)	Heart rate	LF, HF, LF/HF	Learning	Static (N/R)	Short
Cortisol (salivary)	LF/HF			Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
Respiration rate	HF			Experiment	Static (LabChart Pro)	Task-dependent	Undergraduates (40)	[32]
Blood volume	LF, LF/HF			Experiment	Static (Biotrace+)	Short	Undergraduates (73)	[47]
Skin conductance	LF, LF/HF			Experiment	Static (Biotrace+)	Short	Undergraduates (73)	[47]
Stress Scale (EVEA)	LF/HF			Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
Stress Scale (VAS) (Active/Interactive Condition)	RMSSD, LF/HF			Learning, Experiment	Static (Qhrv Assessment)	Short	K-12 Students (104)	[14]
Stress Scale (PSRS)	RMSSD, LF, HF			Experiment	Wearable (BioSign)	Short	Undergraduates (60)	[41]
ATTENTION	L1/L2 (n= 1)	Cortisol (salivary)	LF, HF	Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
		Respiration rate	HF	Experiment	Static (LabChart Pro)	Task-dependent	Undergraduates (40)	[32]
	L4 (n= 2)	Respiration rate	RMSSD	Experiment	Static (GSR+Unit)	Short	Undergraduates (30)	[39]

Measuring Stress and Attention with HRV in Educational Contexts

Most studies measured stress, which is consistent with the general clinical context (Glass, 2009; Sassi et al., 2015). In the clinical context, HRV data have been used for assessing stress at both the trait or tonic level and the state or phasic level (Malik,

Table 3 Predictive validity

		Correlation Between		Context	Device (SW)	Time	Participants (n)	Study
STRESS	L1/L2 (n = 2)	Psychomotor Vigilance Task (PVT)	RMSSD, LF, HF	Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
		Test Score (interactive learning)	RMSSD, pNN50, LF, HF, LF/HF	Learning, Experiment	Wearable (Kubios)	Long	Undergraduates (104)	[44]
	L4 (n = 8)	Test Score	RMSSD, pNN50, LF, HF, LF/HF	Exam	Wearable (Kubios)	Long	Undergraduates (110)	[1]
		Test Score	LF, HF, LFn, HFn, LF/HF, TotPow	Learning	Static (SA-6000)	Short	Undergraduates (109)	[5]
		Test Score	NN50	Experiment	Static (Kubios)	Short	K-12 Students (91)	[26]
		Test Score (passive learning)	RMSSD, pNN50, LF, HF, LF/HF	Learning, Experiment	Wearable (Kubios)	Long	Undergraduates (104)	[44]
		Skill Assessment	SDNN, RMSSD, LF, HF	Exam, Learning	Static (Biopac MP150W)	Task-dependent	Undergraduates (24)	[6]
		Skill Assessment (EPA, OSATS, OTAS)	RR interval, C_HRV, LF/HF	Learning	Wearable (Polar Precision)	Task-dependent	Professionals (30)	[21]
		Tutoring Performance	RMSSD	Learning	Wearable (Kubios)	Short	Undergraduates (22)	[22]
		Psychomotor Vigilance Task (PVT)	LF/HF	Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
ATTENTION	L1/L2 (n = 2)	Stroop Test Score	RMSSD	Experiment	Wearable (Kubios)	Short	K-12 Students (62)	[28]
		Stroop Test Score (errors)	nLF, nHF	Experiment	Static	Short	Undergraduates (160)	[40]
	L4 (n = 3)	Psychomotor Performance (PVT)	RMSSD	Experiment	Wearable (Kubios)	Short	Undergraduate (25)	[29]
		Psychomotor Performance (PVT)	LF/HF	Experiment	Wearable (Kubios)	Short	Professionals (58)	[36]
		Stroop Test Score (time)	nLF, nHF	Experiment	Static	Short	Undergraduates (160)	[40]

1996), although clinicians tend to focus more on chronic stress from more continuous stressors rather than event-based stressors. However, in the studies analyzed, HRV data were predominantly utilized to represent stress at the state or phasic level, where stressors were academic events such as lectures, practice, or exam situations. Castaldo et al. (2019) reviewed 12 studies of acute mental stress measured with HRV in a clinical context, all of which involved stressors from cognitive tasks, such as arithmetic, Stroop tasks, and academic exams. Hair cortisol data, contrary to saliva and blood cortisol, is indicative of accumulated or chronic stress and may not be deemed suitable as a co-measured variable (Huhn et al., 2018). Self-report data, which has typically been used in existing educational research, measures stress at the end of a learning event, and HRV in psychiatric clinical settings measures chronic stress. HRV in the reviewed studies captured changes in learner stress throughout

Table 4 Study features by areas of interest

		Area of interest			
		Stress (n=44)	Attention (n=12)		
Participants	Primary school students	1 (2.0%)	2 (16.7%)		
	Secondary school students	2 (4.5%)	1 (8.3%)		
	Undergraduates	General	15 (34.0%)	6 (50.0%)	
		Health sciences	18 (40.9%)	1 (8.3%)	
	Professionals	Health professionals	7 (15.9%)	1 (8.3%)	
		Others	2 (4.5%)	1 (8.3%)	
Data Collection	Context	Exam	11 (25.0%)	-	
		Learning	22 (50.0%)	2 (16.7%)	
		Experiment	20 (45.4%)	12 (100.0%)	
	Measuring Time	Long-term	6 (13.6%)	2 (16.7%)	
		Task Dependent	Short-term	7 (15.9%)	2 (16.7%)
			RMSSD	30 (68.1%)	8 (66.7%)
HRV Parameters	Time Domain	SDNN	26 (59.1%)	7 (58.3%)	
		pNN50	15 (34.0%)	1 (8.3%)	
		LF	11 (25.0%)	2 (16.7%)	
	Frequency Domain	HF	25 (56.8%)	5 (41.7%)	
		LF/HF	30 (68.1%)	7 (58.3%)	
		SD1	25 (56.8%)	2 (16.7%)	
HRV Tools	Data Collection Device	SD2	7 (15.9%)	1 (8.3%)	
		SD2	8 (18.2%)	2 (16.7%)	
	Analysis Software	Polar series	13 (29.5%)	4 (33.3%)	
		3-lead ECG	6 (13.6%)	2 (16.7%)	
	Co-measures	Psychophysiological Data	Kubios	17 (38.6%)	6 (50.0%)
			Heart rate	25 (56.8%)	4 (33.3%)
Cortisol			12 (27.3%)	2 (16.7%)	
Self-reported Scale Data		Blood pressure	8 (18.2%)	2 (16.7%)	
		SUDS	3 (6.8%)	-	
		STAI	10 (22.7%)	-	
Performance Data	MAAS	-	1 (8.3%)		
	Not validated scales	6 (13.6%)	-		
	Grades	4 (9.0%)	-		
	Skill assessment	3 (6.8%)	-		
Correlation Level/Validity	Concurrent Validity	Surgical training	3 (6.8%)	-	
		Cognitive performance	-	5 (41.7%)	
	Predictive Validity	L1 or L2	15 (34.0%)	1 (8.3%)	
		L4	9 (20.5%)	1 (8.3%)	
Validity	Predictive Validity	L1 or L2	1 (2.0%)	2 (16.7%)	
		L4	8 (18.2%)	4 (33.3%)	

Note. Three most frequently used items included.

Color display according to frequency 0–25%, 25–50%, 50–75%, 75–100%

educational scenes in real time, thereby providing valuable insight into instructional design; this highlights the potential of using HRV measures in the field of education.

Attention was researched relatively less than stress. Measuring attention via HRV data in an educational context is a newer research area. Given that HRV can be used as a marker of attentional processes (Mathewson et al., 2007), HRV measures were used to assess different attentional demands (the extent to which a task is

challenging and requires a high level of attention) in real time. Higher HF-HRV may be indicative of enhanced attention (e.g., Causse et al. (2011), Quintana et al. (2012), and Ramírez et al. (2015)). Among diverse psychophysiological data, EEG data are more commonly utilized to measure attention than HRV data (Schoenberg & David, 2014). However, the mobility of students during learning activities can hinder EEG measurement reliability, leading to a high degree of measurement error and making large-scale, long-term studies challenging (Xu & Zhong, 2018). The scarcity of attention studies using HRV highlights the need for further research in this area.

The studies included in this review showed that adult participants—undergraduates and professionals—were studied more frequently than children or adolescents. Specifically, stress was studied more widely, while attention was studied primarily in undergraduates. Although the therapeutic value of HRV biofeedback has been established in children, adolescents, and adults (Dormal et al., 2021), research in educational contexts with children and adolescents has yet to be adequately explored in comparison to research with adults (Goessl et al., 2017; Lehrer et al., 2020). In addition, participants in the healthcare field (medical students, residents, surgeons, or medical staff) accounted for 54.2% of the study populations. This is likely due to the nature of HRV and the fact that healthcare researchers have the necessary expertise to collect and analyze HRV data considering associated medical factors. This highlights the need for proper guidelines on the collection, analysis, and interpretation of HRV data for educational research.

Collecting and Analyzing HRV Data from Educational Contexts

The present study synthesized the methods that 48 studies reported to provide educational researchers with guidelines for utilizing HRV data to measure stress and attention in educational contexts. A classic work on HRV measurement standards by the Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology (Malik, 1996) originally targeted heart disease patients based on the equipment available in 1996, providing standards for collecting and processing HRV data, interpreting results, and reporting findings. This guideline has been pivotal in HRV research. In the reviewed studies, about 54.2% explicitly reported following Malik's, 1996 guidelines. One reported following the recommendations by Laborde et al. (2017). This study provided methodological considerations crucial for experimental designs, data analysis, and reporting using HRV data in the context of self-regulation research. While it signifies a step forward from clinical diagnostics to the context of psychological research, it acknowledges its own limitations in serving as an imperative guideline. This highlights the nuanced transition HRV research is undergoing, from its roots in clinical applications to broader psychological contexts, but the ongoing need for comprehensive and directly applicable guidelines for its integration into educational research still persists. The other studies did not mention any alternative standards but are presumed to generally follow the same recommended steps, implying the importance of reporting compliance with these standards.

Table 5 Synthesized recommendations for educational research using HRV data

(1) Areas of Measurement	<u>Stress</u>	<ul style="list-style-type: none"> - a complex psychological and physiological phenomenon that arises in response to perceived challenges or demands related to learning - associated with specific situations rather than being chronic in nature - encompassing anticipatory anxiety, increased mental and physiological arousal, and emotional tension. 																				
	<u>Attention</u>	<ul style="list-style-type: none"> - a multifaceted cognitive function involving the allocation of cognitive resources to stimuli or tasks - encompassing focusing, sustaining concentration, managing cognitive load, and the ability to selectively prioritize certain stimuli over others 																				
(2) Participant Control	<u>Participant exclusion</u>	<ul style="list-style-type: none"> - cardiovascular diseases (e.g., hypertension, heart disease, coronary artery disease, arrhythmias, tachycardia) - mental health issues - obesity or weight/BMI-related criteria - endocrine diseases including metabolic disorders (e.g. diabetes) - women not in the proliferative menstrual phase - long-term or regular medication/smoking 																				
	<u>Instructions given before HRV measurement</u>	<ul style="list-style-type: none"> - avoiding caffeine, alcohol, intense exercise, impending smoking - ensuring adequate sleep - stopping medication 																				
<ul style="list-style-type: none"> • The movement intensity of learning activities and desired accuracy of HRV data should be weighed in determining the appropriate device. When feasible, the combined use of different devices is also a viable option. 																						
(3) HRV Measurement Device	<u>Wearable HRV devices</u>	<ul style="list-style-type: none"> - <i>wristband</i>: favored for everyday health tracking; user-friendly - <i>chest belt</i>: preferred by athletes and professionals, ranking accuracy-wise between wristbands and static devices; provides a balance of mobility and precision. 																				
	<u>Static HRV devices</u>	<ul style="list-style-type: none"> - participants remain seated or lie down during measurement - typically measure HRV for a short duration (e.g., 5 minutes) - relatively accurate and precise 																				
<table border="0" style="width: 100%; text-align: center;"> <tr> <td colspan="2">Wearable Device</td> <td></td> <td>Static Device</td> </tr> <tr> <td><i>Wristband</i></td> <td><i>Chest Belt</i></td> <td></td> <td></td> </tr> <tr> <td colspan="2">← low</td> <td>Data Accuracy</td> <td>high →</td> </tr> <tr> <td colspan="2">← high</td> <td>Mobility Ease</td> <td>low →</td> </tr> <tr> <td colspan="2">← long-term (e.g. 24 h)</td> <td>Measuring Duration</td> <td>short-term (e.g. 5 min) →</td> </tr> </table>			Wearable Device			Static Device	<i>Wristband</i>	<i>Chest Belt</i>			← low		Data Accuracy	high →	← high		Mobility Ease	low →	← long-term (e.g. 24 h)		Measuring Duration	short-term (e.g. 5 min) →
Wearable Device			Static Device																			
<i>Wristband</i>	<i>Chest Belt</i>																					
← low		Data Accuracy	high →																			
← high		Mobility Ease	low →																			
← long-term (e.g. 24 h)		Measuring Duration	short-term (e.g. 5 min) →																			

The data collection contexts of the studies suggest that most HRV data measuring stress in an educational setting were collected during learning, with a relatively high proportion collected during exams and in experimental settings, whereas HRV data measuring attention were predominantly collected in an experimental context.

Table 5 (continued)

(4) Data Collection Condition	<ul style="list-style-type: none"> • <i>Noise, temperature, illumination, humidity, the time frame, and participants' posture during HRV data collection need to be controlled.</i> • Participants need to be advised to <i>rest</i> and <i>relax</i> and to practice <i>guided breathing</i> before measuring baseline. • <i>The presence of teachers/professors during data collection must be carefully deliberated in the study design, considering its potential to stress learners.</i> • <i>Baselines</i> should be measured and explicitly reported.
(5) Measurement Duration	<ul style="list-style-type: none"> • The duration of measurement can vary depending on the purpose of the study. • In educational research, HRV can be more commonly measured to represent the <i>state or phasic level of stress</i> and attention rather than the trait or tonic level, i.e., chronic stress or pathological inattentiveness. • Recordings under 30 minutes, preferably <i>around 5 minutes</i>, are typically recommended; however, in educational research, <i>task-dependent measures</i> of HRV are commonly used. • Longer measurements require consideration of <i>data stationarity issues</i>, necessitating the selection of appropriate parameters and preprocessing for non-stationary data and additional analysis, such as the <i>Binary Symbolic Dynamics</i> and <i>Heart Rate Asymmetry (HRA)</i>.
(6) Parameters	<p><u>Time-domain analysis</u></p> <ul style="list-style-type: none"> - considers the time intervals between heartbeats - RMSSD (Root Mean Square of Successive Differences); SDNN (Standard Deviation of NN Intervals); pNN50 (Percentage of Successive NN Intervals that Differ by More than 50 ms) - simple and easy to understand but may not provide much information <p><u>Frequency-domain analysis</u></p> <ul style="list-style-type: none"> - considers the different frequencies present in the HRV signal - HF (High Frequency); LF (Low Frequency); VLF (Very Low Frequency); LF/HF Ratio - provides relatively detailed information but can be difficult to understand <p><u>Non-linear analysis</u></p> <ul style="list-style-type: none"> - considers the complexity and irregularity of the HRV signal - SD1/SD2 (Poincaré Plot Analysis); Detrended Fluctuation Analysis (DFA); Entropy Measures - provides even more detailed information but is even more difficult to understand • Common combination: RMSSD, HF, and LF/HF (Stress: RMSSD and SDNN + LF, HF, and LF/HF; Attention: RMSSD + HF) (For further details, see Table S2)
(7) Co-measured Data	<ul style="list-style-type: none"> • For the list of concurrently collected psycho-physiological measures, validated self-report scales, and performance data, see Table S2.

Studies controlling for participant factors in HRV research—through excluding certain individuals, providing specific instructions, or implementing both strategies—reflect the complexity of HRV as an index of the interplay between

Table 5 (continued)

(8) HRV Software & Pre-Processing	<ul style="list-style-type: none"> • Ensure to select software that offers the most powerful pre-processing and data quality assessment capabilities, which include detrending, noise/artifact detection and removal, data filtering, ectopic beat detection and classification, and visual inspection (For currently available software, see Table S2)
(9) Interpretation	<ul style="list-style-type: none"> • HRV parameters are commonly interpreted relatively, not by absolute clinical standards (See Jarczok et al., 2019; Malik, 1996; Xhyheri et al., 2012 for normal ranges). • Interpretations commonly involve assessing <i>changes from baseline</i> and <i>comparing between experimental and control groups</i>. • Age, gender, health status, and recording duration should be <i>holistically</i> considered
(10) Concurrent & Predictive Validity	<ul style="list-style-type: none"> • Stress in educational contexts: acceptable concurrent validity, limited predictive validity • Attention in educational contexts: limited concurrent validity (suggesting a need for additional data types), limited predictive validity
(11) Standardized Reporting	<ul style="list-style-type: none"> • Published HRV research guidelines must be explicitly followed and reported. • Each aforementioned area should be documented when reporting. <ul style="list-style-type: none"> - Phenomenon or Constructs to Measure: Clearly define the phenomena and constructs being measured. - Participant Control: Detail participant exclusion criteria and any instructions provided before HRV measurement. - Measurement Device and Specifications: Describe HRV measurement devices used with specifications and rationales for their choice. - Data Collection Conditions: Report physical and temporal environmental conditions, the presence or absence of teachers, baseline measurement procedures, and any controls such as respiratory rate monitoring. - Research Design Including Measurement Duration: Outline the overall study design, including the HRV measurement duration. - Measured HRV Parameters and Concurrent Data: State which HRV parameters were measured and any other concurrent data collected. - Pre-processing Specifications: Detail the pre-processing steps, software used and its features, measures adopted to enhance data quality, and any additional analysis conducted for data stationarity. - Interpretation Criteria and Methods: Explain the criteria and methods for data interpretation

physiological, emotional, and cognitive processes. This approach, grounded in the understanding of heart-brain connections (Laborde et al., 2017; Malik, 1996), accounts for the impact of various health conditions and behaviors, like cardiovascular diseases, mental health issues, obesity, and substance intake (e.g., alcohol, caffeine, and medication), as well as hormonal influences related to female menstrual cycles. This level of detail is necessary to ensure the accuracy and reliability of HRV as a measure of these interrelated ANS balance and regulation systems.

In most studies, instructions, such as avoiding caffeine, alcohol, and intense exercise and ensuring adequate sleep, were given to participants before HRV measurement. Studies also controlled for noise, temperature, illumination, humidity, time, and participants' posture during HRV data collection. Instructions employed in the studies imply that all these factors can affect HRV and produce artifacts or noise for HRV data. Similarly, most studies recommended that participants rest, relax, and practice guided breathing before measuring the baseline. This approach, though not strictly enforced, seems to have mitigated the significant respiratory rate influence on HRV. Most studies documented baseline measurements, and we inferred that the four without explicit reports also did so, highlighting the necessity for consistent reporting of key data measurement steps.

Recent studies over the past 3 years have shown a notable surge in the utilization of HRV data in virtual reality (VR), simulation, and online learning contexts (8 out of 12 studies, or 66.7%). Consequently, participant exclusion criteria have been expanded to include technologically related issues such as vision, ear, and movement sickness, which, while not directly related to HRV, impact the study conditions. This trend signifies a shift towards more personalized learning strategies enabled by technology, wherein learning outcomes can be optimized by closely monitoring and adapting to learners' real-time needs and conditions using physiological data like HRV.

Almost 60% of the reviewed studies reported using a wearable HRV measurement device. The literature highlights several differences between wearable and static HRV devices. Static devices typically require users to remain seated or lying down during measurement and typically measure HRV for a short duration, such as 5 min (Rangaswamy et al., 2016). These devices are regarded as more accurate and precise compared to wearable HRV devices. This aligns with our findings that a significant portion of L1/L2 studies, which reported a high correlation between HRV and other measures, used static devices. Wearable devices, particularly wristband and chest belt forms, are extensively employed. Wristbands are favored for everyday health tracking due to their user-friendliness, while chest belts, offering more accurate heart rate data, are preferred by athletes and professionals. In the spectrum of data accuracy for HRV devices, chest belts are situated between wristbands and static devices, providing a balance of mobility and precision.

Although no significant differences in correlation levels of co-measured data and HRV were observed between wristband and chest belt types, Menghini et al. (2019) found that wrist-worn sensors measure HRV accurately in static conditions, like resting and guided breathing, but less so in dynamic conditions such as speaking. These findings underscore that the movement intensity of learning activities and the desired accuracy of HRV data should be weighed when determining the appropriate device. If measurement accuracy is of paramount importance, a static device may be more suitable; conversely, if ease of use or mobility is a priority, a wearable device may be more appropriate.

The HRV measurement duration varied across studies: 68.1% of studies measured HRV data for less than 30 min, with stress being measured primarily in the short term (no longer than 30 min), while attention tended to be measured in both the short and long term (Mathewson et al., 2007; Redondo et al., 2019). According to Malik (1996), the measurement duration can vary depending on the study

purpose. Short-term measurements (e.g., 5 min) are used in clinical settings, while long-term measurements (e.g., 24 h) are used to assess HRV changes during daily activities, for each of which more suitable HRV metrics were provided in the study. Task-dependent measures of HRV have commonly been employed, particularly when specific tasks need to be completed. This emphasizes the significance of considering the areas of inquiry and context in which HRV is being studied. Experimental designs employed in each study are detailed in Table S3 (*Details of Reviewed Studies* in Online Supplementary File 2).

In long-term HRV data collection and in abnormal situations like stress, the issue of data stationarity has been long addressed (Kaczmarek et al., 2019; Malik, 1996; Spellenberg et al., 2020). Over long periods, physiological responses can change significantly, making it difficult to assume a constant pattern and interpret frequency domain metrics like HRV LF and HF components. Averaging HRV data over prolonged monitoring or during stress might also mask important fluctuations and nuances in bodily responses, leading to less accurate assessments of autonomic regulation.

Addressing the issue of data stationarity and more reliable data collection, pre-processing steps were emphasized and reported in the studies reviewed. For pre-processing, studies utilized various analysis software programs. The most used software was Kubios, followed by the Polar series (including Polar Precision Performance program, Polar ProTrainer software, and Polar Flow). Other software programs used include MATLAB, Cardiovit AT-10 Plus, LabChart Pro software, Acq-Knowledge, QHRV stress assessment, and BEAT software. All these software programs feature pre-processing functions, which include detrending, noise/artifact detection and removal, data filtering, ectopic beat detection and classification, data quality assessment, and visual inspection. Under nonstationary conditions like stress, the Binary Symbolic Dynamics and Heart Rate Asymmetry (HRA) methods are useful additions to HRV analysis. The Binary Symbolic Dynamics method simplifies HRV into binary sequences to better identify stress-related changes (Speelens et al., 2020). HRA distinguishes emotional responses, with positive emotions showing more short-term deceleration (Kaczmarek et al., 2019). Although not used in the 48 studies, incorporating these methods in future research, especially in educational contexts for stress assessment, could enhance the reliability of HRV as a stress measure.

Studies employed three types of HRV analysis: time domain, frequency domain, and non-linear. Studies often used more than one type of analysis. Time domain analysis considers the time intervals between heartbeats and reports values such as the mean and standard deviation of these intervals (e.g., SDNN and RMSSD). This analysis type is simple and easy to understand, but it may not provide as much information as the other two analysis types. Frequency domain analysis considers different frequencies present in the HRV signal. This can be represented as power spectral density, which separates the signal into different frequency bands and depicts the power (or strength) of each band (e.g., HF, LF, and VLF). This analysis type can provide more detailed information regarding the HRV signal, but it can be more difficult to understand. Non-linear analysis considers the complexity and irregularity of the HRV signal. This can be implemented using techniques such as fractal analysis and entropy analysis. Non-linear analysis can provide even more detailed information about the HRV signal but can be more difficult to understand than frequency domain analysis.

Stress was primarily investigated using the RMSSD and SDNN parameters in the time domain and LF, HF, and LF/HF in the frequency domain. Attention was investigated using RMSSD in the time domain and HF in the frequency domain, but the use of LF and LF/HF was relatively limited. Within the domain of analysis, some studies used all parameters, some used a subset, and some used modified versions (C_HRV). The most common combination was RMSSD, HF, and LF/HF.

In interpreting parameters, standard normal and abnormal HRV ranges exist in a clinical context: a daytime RMSSD value lower than $25 \pm 4 \text{ ms}^2$ is associated with an increased risk of cardiovascular disease (Jarczok et al., 2019). An SDNN value of less than 50 ms^2 is interpreted as a state of poor resistance to stress, and a pNN50 value of less than 3% indicates parasympathetic dysfunction (Jarczok et al., 2019; Xhyheri et al., 2012). The LF normal range is $1170 \pm 413 \text{ ms}^2$, HF is $975 \pm 203 \text{ ms}^2$, and the LF/HF ratio is 1.5–2.0 (Nunan et al., 2010). SD1 and SD2 are standard deviations of RR interval variability calculated on a phase space plot (Poincaré plot). In healthy individuals, patterns show broad and generally symmetrical oval shapes, whereas patterns in unhealthy individuals deviate from the normal range and present clustered shapes in a narrow region (Fishman et al., 2012).

The interpretation of HRV parameters in educational research differs from the interpretation for clinical use, as the former focuses on monitoring HRV changes in response to different educational interventions or conditions. Thus, although there are established standards for clinical use, it is common in educational research to interpret HRV parameters according to personalized algorithms by comparing a baseline level to a level post-intervention to assess whether values increased or decreased (Lu et al., 2022; Xhyheri et al., 2012).

The Concurrent and Predictive Validity of HRV Measures for Educational Research

HRV data seemed to have an acceptable level of concurrent validity as a measure of stress in educational contexts due to its strong correlation with other measures of stress, such as heart rate, cortisol levels, and validated stress scale scores. Among the 39 pairs of data that were analyzed, 14 pairs were associated with high correlations, while 8 pairs indicated no correlation. The concurrent validity of HRV data as a measure of attention in an educational context is less clear. Findings showed that among the 39 pairs of data, only one pair was highly correlated, while two pairs indicated no correlation. This suggests that a correlation between HRV data and measures of attention cannot yet be confirmed.

Noteworthy, among the studies that collected heart rate data, only 4.2% were categorized within the group with high correlations with HRV data, which is inconsistent with a previous research (Kazmi et al., 2016; Sacha, 2014) where heart rate was strongly related to HRV. Additionally, among the 12 studies that collected cortisol data, only 3 (1 blood sample and 2 salivary sample) reported a high correlation with HRV data (RMSSD, LF, HF), while another study showed that the parameter LF/HF had no correlation with cortisol data. In terms of stress studies, 15.9% of studies that employed validated stress scales reported a high correlation between HRV data and stress scale scores. The instruments used in L1 studies were the TAI

and EVEA, and the parameters used were time domain (RMSSD, NN50, SDNN) and frequency domain (LF, HF, LF/HF) data. The study also found that EVEA scores highly correlated with HRV RMSSD, LF, and HF data but were not correlated with LF/HF. Overall, the correlations between HRV and other measures, such as heart rate, cortisol levels, and stress scale scores, were not always present and could have varied depending on the areas of interest, the types of scales or sub-scales used, and the parameters considered.

This study also investigated the extent to which HRV data can predict certain learning outcomes. Our findings suggest a limited correlation between performance and stress measured by HRV data. Of the studies included in the analysis that measured HRV data along with performance data, only two found a correlation between the two variables. Furthermore, most studies (80%) that measured test scores reported no correlation between test scores and HRV data, and two out of three studies using skill assessment data also reported no correlation. Additionally, two studies using PVT (psychomotor vigilance task) found low predictive validity for HRV data measuring attention. Previous research also suggests that stress can negatively impact cognitive performance. For example, studies have found that stress can impair working memory and attention, increasing the likelihood of errors or accidents (Lupien et al., 2009; McEwen & Gianaros, 2011). HRV data is useful for providing a more comprehensive understanding of learners' psychological, physiological, and cognitive states, but more research is needed to confirm its validity for measuring attention and predicting learning outcomes.

Table 5 consolidates the discussions so far, synthesizing recommendations for educational research using HRV data and covering eleven key domains. The 1996 guidelines for HRV research on cardiac patients did not include certain elements that are newly added in these guidelines, which are tailored for educational research. These new additions are (1) *areas of measurement*, (7) *co-measured data*, and (10) *concurrent and predictive validity*; due to the focus on normal learners rather than cardiac patients, (2) *participant control*, (4) *data collection condition*, and specialized (9) *interpretation* have been incorporated. Technological advancements available nearly 30 years later have led to updates in (3) *HRV measurement devices* and (8) *HRV software and pre-processing*, while (5) *measurement duration* and (6) *parameters* build upon the existing guidelines with new techniques. The guidelines under (11) *standardized reporting* delineate reporting items to ensure that all the aforementioned aspects are adhered to and duly reported.

The limitation of the findings is that it is not yet clear whether HRV data can be used to accurately measure attention in an educational context. Additionally, our review suggests that there may be a limited correlation between stress and performance as measured by HRV data, indicating that more research is needed to confirm the predictive validity of HRV data regarding learning outcomes. The review also found that correlations between HRV data and other measures such as heart rate, cortisol levels, and stress scale scores are not always present and could vary depending on the areas of interest, the kinds of scales or sub-scales employed, and the parameters used. This means that more research is needed to understand the relationship between HRV and these other measures and to determine the best ways to use HRV data in educational research. In addition, in this study, we consider stress

and attention as key psychological and cognitive phenomena in learners, focusing our review on research that includes either or both of these terms. Studies that only used terms such as anxiety, arousal, or cognitive load without mentioning stress or attention may have been excluded. We suggest that future research should conduct a more comprehensive analysis of HRV data usage across a broader spectrum of learner characteristics.

Although HRV data can capture changes in learner stress and/or attention in real time, providing valuable insight into instructional design, it is not yet completely clear whether HRV can be used as a valid measure of stress and/or attention in an educational context. Future research should consider the use of HRV data in educational contexts to investigate its potential and thereby develop evidence-based guidelines for the proper utilization, collection, and interpretation of HRV data for educational research, thereby ensuring its validity. In addition, the increased utilization of HRV data in digital learning research, as shown in recent studies, suggests the need for future investigations to more finely understand learners' cognitive and emotional states in technology enhanced learning (TEL) environments and to link these insights to educational interventions.

Conclusion

This study examined the potential and validity of utilizing HRV data for educational research by conducting a systematic review of 48 studies that employed HRV measurement in educational contexts. Our findings implied that HRV data can provide valuable real-time insight into instructional design by capturing changes in learner stress during educational activities. The study also highlights the need for guidelines on the proper use of HRV in educational research and calls for further detailed investigations to be conducted. HRV data had a moderate level of concurrent validity as a measure of stress in an educational context. The concurrent validity of HRV data as a measure of attention in an educational context was less clear. There were limited correlations between stress and performance as measured by HRV data. More research is needed to confirm HRV data's validity for measuring stress and attention and for predicting learning outcomes.

This study significantly extends the classic standards of utilizing HRV data, initially based on clinical contexts with heart disease patients, by synthesizing recommendations for educational research among healthy learners. It integrates the use of modern devices and software previously unavailable, providing guidelines for collecting, processing, analyzing, and interpreting HRV data. Furthermore, the study delves into the interplay of HRV with psychological and physiological aspects, such as stress and attention, thereby proposing fresh HRV research avenues and expanding its scope beyond conventional heart health metrics.

It is of great value to incorporate more objective data into educational research to make evidence-based educational decisions. However, it is important not to over-rely on a data-driven approach, but rather to supplement quantitative data

with in-depth qualitative data and provide a more comprehensive understanding of educational phenomena. We hope that this study will contribute to promote a multifaceted approach by utilizing multi-modal data for educational research.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-024-09878-x>.

Acknowledgements We acknowledge with gratitude the valuable contributions of Professor Jae Seung Kang (MD, Ph.D., Department of Anatomy and Cell Biology, College of Medicine, Seoul National University) to the medical aspects of our study, meriting authorial recognition.

Funding Open Access funding enabled and organized by Seoul National University. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Ministry of Science and ICT (MSIT) (2021R1F1A1056465) and the Ministry of Education (NRF-2020S1A5A8041947).

Data Availability The data that support the findings of this study are available on request.

Declarations

Ethics Approval N/A.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmady, S., Khajeali, N., Kalantarion, M., Sharifi, F., & Yaseri, M. (2021). Relation between stress, time management, and academic achievement in preclinical medical education: a systematic review and meta-analysis. *Journal of Education and Health Promotion, 10*(1), 10–32. https://doi.org/10.4103/jehp.jehp_600_20
- Antelmi, I., De Paula, R. S., Shinzato, A. R., Peres, C. A., Mansur, A. J., & Grupi, C. J. (2004). Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease. *The American Journal of Cardiology, 93*(3), 381–385. <https://doi.org/10.1016/j.amjcard.2003.09.065>
- Blackburn, E. H., & Epel, E. S. (2017). *The Telomere effect: a revolutionary approach to living younger, healthier, longer*. Grand Central Publishing.
- Bonnemeier, H., Wiegand, U. K., Brandes, A., Kluge, N., Katus, H. A., Richardt, G., & Potratz, J. (2003). Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. *Journal of Cardiovascular Electrophysiology, 14*(8), 791–799.
- Brann, K. L., Daniels, B., Chafouleas, S. M., & DiOrio, C. A. (2021). Usability of social, emotional, and behavioral assessments in schools: a systematic review from 2009 to 2019. *School Psychology Review, 51*(1), 6–24. <https://doi.org/10.1080/2372966x.2020.1836518>

- Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, *37*(8), 1111–1134. <https://doi.org/10.1016/j.psyneuen.2011.12.010>
- *Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making*, *19*(1), Article 12. <https://doi.org/10.1186/s12911-019-0742-y>
- Causse, M., Baracat, B., Pastor, J., & Dehais, F. (2011). Reward and uncertainty favor risky decision-making in pilots: evidence from cardiovascular and oculometric measurements. *Applied Psychophysiology and Biofeedback*, *36*(4), 231–242. <https://doi.org/10.1007/s10484-011-9163-0>
- Chafouleas, S. M. (2011). Direct behavior rating: A review of the issues and research in its development. *Education and Treatment of Children*, *34*(4), 575–591. <https://doi.org/10.1353/etc.2011.0034>
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. <https://doi.org/10.1111/1467-8721.ep10768783>
- Dimitriev, D. A., Dimitriev, A. D., Karpenko, Y. D., & Saperova, E. V. (2008). Influence of examination stress and psychoemotional characteristics on the blood pressure and heart rate regulation in female students. *Human Physiology*, *34*(5), 617–624. <https://doi.org/10.1134/s0362119708050101>
- Dormal, V., Vermeulen, N., & Mejias, S. (2021). Is heart rate variability biofeedback useful in children and adolescents? A systematic review. *Journal of Child Psychology and Psychiatry*, *62*(12), 1379–1390. <https://doi.org/10.1111/jcpp.13463>
- Fink, G. (2017). Stress: Concepts, definition and history. *Neuroscience and Biobehavioral Psychology*, *549*–555. <https://doi.org/10.1016/B978-0-12-809324-5.02208-2>
- Fishman, M., Jacono, F. J., Park, S., Jamasebi, R., Thungtong, A., Loparo, K. A., & Dick, T. E. (2012). A method for analyzing temporal patterns of variability of a time series from Poincare plots. *Journal of Applied Physiology*, *113*(2), 297–306. <https://doi.org/10.1152/jappphysiol.01377.2010>
- Forte, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. *Frontiers in Neuroscience*, *13*. <https://doi.org/10.3389/fnins.2019.00710>
- Gaffey, A. E., & Wirth, M. M. (2014). Psychophysiological measures. *Encyclopedia of Quality of Life and Well-Being Research*, *5181*–5184. https://doi.org/10.1007/978-94-007-0753-5_2315
- Glass, L. (2009). Introduction to controversial topics in nonlinear science: Is the normal heart rate chaotic? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *19*(2). <https://doi.org/10.1063/1.3156832>
- Goessl, V. C., Curtiss, J. E., & Hofmann, S. G. (2017). The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. *Psychological Medicine*, *47*(15), 2578–2586. <https://doi.org/10.1017/s0033291717001003>
- Guo, J., Dai, Y., Wang, C., Wu, H., Xu, T., & Lin, K. (2020). A physiological data-driven model for learners' cognitive load detection using HRV-PRV feature fusion and optimized XGBoost classification. *Software: Practice and Experience*, *50*(11), 2046–2064. <https://doi.org/10.1002/spe.2730>
- Hammoud, S., Karam, R., Mourad, R., Saad, I., & Kurdi, M. (2018). Stress and heart rate variability during university final examination among Lebanese students. *Behavioral Sciences*, *9*(1), 3. <https://doi.org/10.3390/bs9010003>
- Hansen, A. L., Johnsen, B. H., Sollers, J. J., Stenvik, K., & Thayer, J. F. (2004). Heart rate variability and its relation to prefrontal cognitive function: the effects of training and detraining. *European Journal of Applied Physiology*, *93*(3), 263–272. <https://doi.org/10.1007/s00421-004-1208-0>
- Huhn, D., Schmid, C., Erschens, R., Junne, F., Herrmann-Werner, A., Moltner, A., Herzog, W., & Nikendei, C. (2018). A comparison of stress perception in international and local first semester medical students using psychometric, psychophysiological, and humoral methods. *International Journal of Environmental Research and Public Health*, *15*(12), 2820. <https://doi.org/10.3390/ijerph15122820>
- Huhn, S., Axt, M., Gunga, H.-C., Maggioni, M. A., Munga, S., Obor, D., Sié, A., Boudo, V., Bunker, A., Sauerborn, R., Bärnighausen, T., & Barteit, S. (2022). The impact of wearable technologies in health research: Scoping review. *JMIR MHealth and UHealth*, *10*(1). <https://doi.org/10.2196/34384>
- Hunt, S. M., & Bhopal, R. (2004). Self-report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture. *Journal of Epidemiology & Community Health*, *58*(7), 618–622. <https://doi.org/10.1136/jech.2003.010074>
- Jarczok, M., Koenig, J., Wittling, A., Fischer, J., & Thayer, J. (2019). First evaluation of an index of low vagally-mediated heart rate variability as a marker of health risks in human adults: Proof of concept. *Journal of Clinical Medicine*, *8*(11), 1940. <https://doi.org/10.3390/jcm8111940>

- Kaczmarek, L. D., Behnke, M., Enko, J., Kosakowski, M., Hughes, B. M., Piskorski, J., & Guzik, P. (2019). Effects of emotions on heart rate asymmetry. *Psychophysiology*, *56*(4). <https://doi.org/10.1111/psyp.13318>
- Kazmi, S. Z. H., Zhang, H., Aziz, W., Monfredi, O., Abbas, S. A., Shah, S. A., & Butt, W. H. (2016). Inverse correlation between heart rate variability and heart rate demonstrated by linear and nonlinear analysis. *PLoS One*, *11*(6). <https://doi.org/10.1371/journal.pone.0157557>
- Kmet, L. M., Lee, R. C., & Cook, L. S. (2004, February). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields*. (HTA Initiative #13). Alberta Heritage Foundation for Medical Research. <https://doi.org/10.7939/R37M04F16>
- Kofman, O., Meiran, N., Greenberg, E., Balas, M., & Cohen, H. (2006). Enhanced performance on executive functions associated with examination stress: Evidence from task-switching and Stroop paradigms. *Cognition & Emotion*, *20*(5), 577–595. <https://doi.org/10.1080/02699930500270913>
- Krumpal, I. (2011). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>
- Kwon, K., Hanrahan, A. R., & Kupzyk, K. A. (2017). Emotional expressivity and emotion regulation: relation to academic functioning among elementary school children. *School Psychology Quarterly*, *32*(1), 75–88. <https://doi.org/10.1037/spq0000166>
- Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, *8*. <https://doi.org/10.3389/fpsyg.2017.00213>
- Lee, L., Packer, T. L., Tang, S. H., & Girdler, S. (2008). Self-management education programs for age-related macular degeneration: a systematic review. *Australasian Journal on Ageing*, *27*, 170–176. <https://doi.org/10.1111/j.1741-6612.2008.00298.x>
- Lee, J., Kim, H., Kim, K. H., Jung, D., Jowsey, T., & Webster, C. S. (2020). Effective virtual patient simulators for medical communication training: a systematic review. *Medical Education*, *54*(9), 786–795. <https://doi.org/10.1111/medu.14152>
- Lehrer, P., Kaur, K., Sharma, A., Shah, K., Huseby, R., Bhavsar, J., Sgobba, P., & Zhang, Y. (2020). Heart rate variability biofeedback improves emotional and physical health and performance: a systematic review and meta-analysis. *Applied Psychophysiology and Biofeedback*, *45*(3), 109–129. <https://doi.org/10.1007/s10484-020-09466-z>
- Lisnjj, K. T., Russell, R., & Papadopoulos, A. (2020). Risk and protective factors for anxiety impacting academic performance in post-secondary students. *Canadian Journal of Higher Education/revue Canadienne D'enseignement Supérieur*, *50*(2), 71–88. <https://doi.org/10.47678/cjhe.v50i2.188715>
- Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience*, *13*. <https://doi.org/10.3389/fnhum.2019.00057>
- Lu, K., Dahlman, A. S., Karlsson, J., & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A systematic review. *Accident Analysis & Prevention*, *178*. <https://doi.org/10.1016/j.aap.2022.106830>
- Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445. <https://doi.org/10.1038/nrn2639>
- Malik, M. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Annals of Noninvasive Electrocardiology*, *1*, 151–181. <https://doi.org/10.1111/j.1542-474X.1996.tb00275.x>
- Martin, R. D., & Naziruddin, Z. (2020). Systematic review of student anxiety and performance during objective structured clinical examinations. *Currents in Pharmacy Teaching and Learning*, *12*(12), 1491–1497. <https://doi.org/10.1016/j.cptl.2020.07.007>
- Mathewson, M. H., Eimer, M. S. J., & van Wouwe, J. C. F. (2007). Heart rate variability as an index of attentional load. *Biological Psychology*, *75*(1), 73–78.
- McEwen, B. S., & Gianaros, P. J. (2011). Central role of the brain in stress and adaptation: Links to socioeconomic status, health, and disease. *Annals of the New York Academy of Sciences*, *1231*(1), 1–15. <https://doi.org/10.1111/j.1749-6632.2009.05331.x>
- Melillo, P., Bracale, M., & Pecchia, L. (2011). Nonlinear heart rate variability features for real-life stress detection. Case study: Students under stress due to university examination. *BioMedical Engineering OnLine*, *10*(1), 96. <https://doi.org/10.1186/1475-925x-10-96>

- Melis, C., & van Boxtel, A. (2001). Differences in autonomic physiological responses between good and poor inductive reasoners. *Biological Psychology*, 58(2), 121–146. [https://doi.org/10.1016/s0301-0511\(01\)00112-0](https://doi.org/10.1016/s0301-0511(01)00112-0)
- Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M., & Sarlo, M. (2019). Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*, 56(11). <https://doi.org/10.1111/psyp.13441>
- Michels, N., Clays, E., De Buyzere, M., Huybrechts, I., Marild, S., Vanaelst, B., De Henauw, S., & Sioen, I. (2013). Determinants and reference values of short-term heart rate variability in children. *European Journal of Applied Physiology*, 113(6), 1477–1488. <https://doi.org/10.1007/s00421-012-2572-9>
- Moreira, P. S., Almeida, P. R., Leite-Almeida, H., Sousa, N., & Costa, P. (2016). Impact of chronic stress protocols in learning and memory in rodents: Systematic review and meta-analysis. *PLoS One*, 11(9). <https://doi.org/10.1371/journal.pone.0163245>
- Nunan, D., Sandercock, G., & Brodie, D. (2010). A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing and Clinical Electrophysiology*, 33(11), 1407–1417. <https://doi.org/10.1111/j.1540-8159.2010.02841.x>
- Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology*, 86(1), 122–133. <https://doi.org/10.1037/0022-0663.86.1.122>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2020). The Prisma 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 1–11. <https://doi.org/10.31222/osf.io/v7gm2>
- 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>
- Prichard, R. S., O'Neill, C. J., Oucharek, J. J., Colinda, Y. H., Delbridge, L. W., & Sywak, M. S. (2012). A prospective study of heart rate variability in endocrine surgery: Surgical training increases consultant's mental strain. *Journal of Surgical Education*, 69(4), 453–458. <https://doi.org/10.1016/j.jsurg.2012.04.002>
- Prokofieva, V., Kostromina, S., Polevaia, S., & Fenouillet, F. (2019). Understanding emotion-related processes in classroom activities through functional measurements. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02263>
- Quintana, D. S., Guastella, A. J., Outhred, T., Hickie, I. B., & Kemp, A. H. (2012). Heart rate variability is associated with emotion recognition: direct evidence for a relationship between the autonomic nervous system and social cognition. *International Journal of Psychophysiology*, 86(2), 168–172. <https://doi.org/10.1016/j.ijpsycho.2012.08.012>
- Ramírez, E., Ortega, A. R., & Del Paso, G. A. R. (2015). Anxiety, attention, and decision making: the moderating role of heart rate variability. *International Journal of Psychophysiology*, 98(3), 490–496. <https://doi.org/10.1016/j.ijpsycho.2015.10.007>
- Ramírez-Adrados, A., Fernández-Martínez, S., Martínez-Pascual, B., Gonzalez-de-Ramos, C., Fernández-Elías, V. E., & Clemente-Suárez, V. J. (2020). Psychophysiological stress response of physiotherapy last year students in his final degree dissertation. *Physiology & Behavior*, 222. <https://doi.org/10.1016/j.physbeh.2020.112928>
- Rangaswamy, D. T. S., Silverman, E. K., & Rangaswamy, M. (2016). Comparison of heart rate variability measured by ECG and PPG: a systematic review. *Journal of Clinical Monitoring and Computing*, 30(2), 195–202.
- Redondo, B., Vera, J., Luque-Casado, A., García-Ramos, A., & Jiménez, R. (2019). Associations between accommodative dynamics, heart rate variability and behavioural performance during sustained attention: A test-retest study. *Vision Research*, 163, 24–32. <https://doi.org/10.1016/j.visres.2019.07.001>
- Redondo-Flórez, L. A. U. R. A., Tornero-Aguilera, J. F., & Clemente-Suárez, V. J. (2020). Could academic experience modulate psychophysiological stress response of biomedical sciences students in laboratory? *Physiology & Behavior*, 223. <https://doi.org/10.1016/j.physbeh.2020.113017>
- Ribeiro, Í. J. S., Pereira, R., Freire, I. V., de Oliveira, B. G., Casotti, C. A., & Boery, E. N. (2018). Stress and quality of life among university students: a systematic literature review. *Health Professions Education*, 4(2), 70–77. <https://doi.org/10.1016/j.hpe.2017.03.002>
- Sacha, J. (2014). Interaction between heart rate and heart rate variability. *Annals of Noninvasive Electrocardiology*, 19(3), 207–216. <https://doi.org/10.1111/anec.12148>

- Salinsky, M. C., Storzach, D., Dodrill, C. B., & Binder, L. M. (2001). Test–retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12–16-week period. *Journal of the International Neuropsychological Society*, 7(5), 597–605. <https://doi.org/10.1017/s1355617701755075>
- Sassi, R., Cerutti, S., Lombardi, F., Malik, M., Huikuri, H. V., Peng, C. K., & Macfadyen, R. (2015). Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. *Ep Europace*, 17(9), 1341–1353. <https://doi.org/10.1093/europace/euv015>
- Schmier, J. K., & Halpern, M. T. (2004). Patient recall and recall bias of health state and health status. *Expert Review of Pharmacoeconomics & Outcomes Research*, 4(2), 159–163. <https://doi.org/10.1586/14737167.4.2.159>
- Schoenberg, P. L., & David, A. S. (2014). Biofeedback for psychiatric disorders: a systematic review. *Applied Psychophysiology and Biofeedback*, 39(2), 109–135. <https://doi.org/10.1007/s10484-014-9246-9>
- Scrimin, S., Peruzza, M., Mastromatteo, L. Y., & Patron, E. (2021). Psychophysiological regulation and classroom climate influence First and second graders' well-being: the role of body mass index. *European Journal of Investigation in Health, Psychology and Education*, 11(4), 1581–1598. <https://doi.org/10.3390/ejihpe11040112>
- Scrimin, S., Moscardino, U., & Mason, L. (2018). First-graders' allocation of attentional resources in an emotional Stroop task: The role of heart period variability and classroom climate. *British Journal of Educational Psychology*, 89(1), 146–164. <https://doi.org/10.1111/bjep.12228>
- Scrimin, S., Moscardino, U., Finos, L., & Mason, L. (2019). Effects of psychophysiological reactivity to a school-related stressor and temperament on early adolescents' academic performance. *The Journal of Early Adolescence*, 39(6), 904–931. <https://doi.org/10.1177/0272431618797008>
- Spellenberg, C., Heusser, P., Büssing, A., Savelsbergh, A., & Cysarz, D. (2020). Binary symbolic dynamics analysis to detect stress-associated changes of nonstationary heart rate variability. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-72034-2>
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. Jossey-Bass. <https://psycnet.apa.org/record/1995-98746-000>
- Tharion, E., Parthasarathy, S., & Neelakantan, N. (2009). Short-term heart rate variability measures in students during examinations. *National Medical Journal of India*, 22(2), 63–66.
- Tracy, L. M., Ioannou, L., Baker, K. S., Gibson, S. J., Georgiou-Karistianis, N., & Giummarra, M. J. (2016). Meta-analytic evidence for decreased heart rate variability in chronic pain implicating parasympathetic nervous system dysregulation. *Pain*, 157(1), 7–29. <https://doi.org/10.1097/j.pain.0000000000000360>
- von der Embse, N., Barterian, J., & Segool, N. (2012). Test anxiety interventions for children and adolescents: a systematic review of treatment studies from 2000–2010. *Psychology in the Schools*, 50(1), 57–71. <https://doi.org/10.1002/pits.21660>
- Voss, A., Schroeder, R., Heitmann, A., Peters, A., & Perz, S. (2015). Short-term heart rate variability—Influence of gender and age in healthy subjects. *PLoS One*, 10(3). <https://doi.org/10.1371/journal.pone.0118308>
- Wagenaar, W. A., & Keren, G. B. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. *NATO ASI Series*, 87–103. https://doi.org/10.1007/978-3-642-50329-0_6
- Weenk, M., Alken, A. P. B., Engelen, L. J. L. P. G., Bredie, S. J. H., van de Belt, T. H., & van Goor, H. (2018). Stress measurement in surgeons and residents using a smart patch. *The American Journal of Surgery*, 216(2), 361–368. <https://doi.org/10.1016/j.amjsurg.2017.05.015>
- Wetzel, C. M., Black, S. A., Hanna, G. B., Athanasiou, T., Kneebone, R. L., Nestel, D., Wolfe, J. H., & Woloshynowych, M. (2010). The effects of stress and coping on surgical performance during simulations. *Annals of Surgery*, 251(1), 171–176. <https://doi.org/10.1097/sla.0b013e3181b3b2be>
- Williamson, A. (2007). Using self-report measures in neurobehavioural toxicology: Can they be trusted? *Neurotoxicology*, 28(2), 227–234. <https://doi.org/10.1016/j.neuro.2006.03.009>
- Wimmer, S., Lackner, H. K., Papousek, I., & Paechter, M. (2019). Influences of different dimensions of academic self-concept on students' cardiac recovery after giving a stressful presentation. *Psychology Research and Behavior Management*, 12, 1031–1040. <https://doi.org/10.2147/prbm.s219784>
- Xhyheri, B., Manfrini, O., Mazzolini, M., Pizzi, C., & Bugiardini, R. (2012). Heart rate variability today. *Progress in Cardiovascular Diseases*, 55(3), 321–331. <https://doi.org/10.1016/j.pcad.2012.09.001>

- Xu, J., & Zhong, B. (2018). Review on portable EEG technology in educational research. *Computers in Human Behavior*, *81*, 340–349. <https://doi.org/10.1016/j.chb.2017.12.037>
- Yee Chung, J. W., Fuk So, H. C., Tak Choi, M. M., Man Yan, V. C., & Shing Wong, T. K. (2021). Artificial intelligence in education: using heart rate variability (HRV) as a biomarker to assess emotions objectively. *Computers and Education: Artificial Intelligence*, *2*, 100011. <https://doi.org/10.1016/j.caeai.2021.100011>
- Zeidner, M. (2007). Test anxiety in educational contexts. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in Education* (pp. 165–184). Cambridge, MA: Academic Press. <https://doi.org/10.1016/b978-012372545-5/50011-3>
- Zeidner, M. (2014). Anxiety in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 265–288). New York, NY: Routledge. <https://doi.org/10.4324/9780203148211.ch14>
- Zhang, J. (2007). Effect of age and sex on heart rate variability in healthy subjects. *Journal of Manipulative and Physiological Therapeutics*, *30*(5), 374–379. <https://doi.org/10.1016/j.jmpt.2007.04.001>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.