



# Neurophysiological Measurements in Higher Education: A Systematic Literature Review

Ali Darvishi<sup>1</sup> · Hassan Khosravi<sup>1</sup>  · Shazia Sadiq<sup>1</sup> · Barbara Weber<sup>2,3</sup>

Accepted: 11 May 2021 / Published online: 16 June 2021  
© International Artificial Intelligence in Education Society 2021

## Abstract

The use of neurophysiological measurements to advance the design, development, use, acceptance, influence and adaptivity of information systems is receiving increasing attention. Within the field of education, neurophysiological measurements have commonly been used to capture a learner's psychological constructs such as cognitive load, attention and emotion, which play an important role in student learning. This paper systematically examines the literature on the use of neurophysiological measurements in higher education. In particular, using a well-established Systematic Literature Review (SLR) method, we identified 83 papers reporting empirical evidence about the outcome of employing neurophysiological measurements within educational technologies in higher education. The findings of the SLR are divided into three main themes discussing the employed measurements, experimental settings and constructs and outcomes. Our findings identify that (1) electroencephalography and facial expression recognition are the dominantly employed types of measurement, (2) the majority of the experiments used a pre-experimental design, (3) attention and emotion are the two foremost cognitive and non-cognitive constructs under investigation, while less emphasis is paid to meta-cognitive constructs and (4) the reported results mostly focus on monitoring learners' states, which are not always the same as the intended purpose, such as developing an adaptive system. On a broader term, the review of the literature provides evidence of the effective use of neurophysiological measurements by educational technologies to enhance learning; however, a number of challenges and concerns related to the accuracy and validity of the captured construct, the intrusiveness of the employed instruments as well as ethical and privacy considerations have surfaced, that need to be addressed before such technologies can be employed and adopted at scale.

**Keywords** Neurophysiological measurements · Psychological constructs · Neuro-information-systems · Educational technologies · Higher education

---

✉ Ali Darvishi  
[a.darvishi@uq.edu.au](mailto:a.darvishi@uq.edu.au)

## Introduction

Educational technologies have changed teaching and learning in higher education in many ways, transforming learning processes to be more effective both in formal and informal settings (Kirkwood and Price 2014; Lodge and Harrison 2019). In particular, educational technologies have been used to promote collaboration (Beldarrain 2006; Money and Dean 2019; So and Brush 2008), introduce gamification to education (Dicheva et al. 2015), provide better access and inquiry to learning resources (Hill and Hannafin 2001; MacKay 2019), develop novel supplementing curricula (Hawkins and Collins 1992; Thomas et al. 2016), engage students in peer learning (Boud et al. 2014), deliver adaptive instructions (Aleven et al. 2016), develop authentic ways of assessing students (Barber et al. 2015; Fluck 2019; McLoughlin and Luca 2002), provide rich and timely feedback (Ali et al. 2012; Pardo et al. 2019; Tempelaar et al. 2015) and much more. A side benefit of using educational technologies is that they provide rich digital traces about students' behaviours and interactions with learning activities, which are mined with the aim of discovering, monitoring, understanding and improving educational processes (Bogarín et al. 2018; Garcia et al. 2018). While these logs of digital traces paint a comprehensive picture of learner's behavioural engagement with learning, arguably they have a limited ability in terms of providing insights about a learner's psychological constructs such as cognitive load, attention and emotion (Fredricks et al. 2004; Dunn and Kennedy 2019), which also play an important role in student learning (Gašević et al. 2015).

The most common approach to acquiring information about psychological constructs, which are not directly observable (Fried 2017), is to use subjective measures by directly asking participants to complete questionnaires or surveys (Rubio et al. 2004). Subjective measures have been used heavily alongside behavioural measures in research from many fields as they are relatively easy and cost-effective to conduct at both small and large scales (Beg 2005; Saw et al. 2016; Zhou and Zhang 2019; Darvishi et al. 2020). Within educational technologies, subjective measures are the most common measures for studying cognitive (e.g., attention), non-cognitive (e.g., emotions) and meta-cognitive (e.g., self-regulation of cognition) constructs related to student engagement and learning (Greene 2015; Henrie et al. 2015; Sinatra et al. 2015). However, using subjective measures to acquire information about psychological constructs has two main drawbacks. Firstly, they are subject to concerns to cognitive biases and internal validity as the accuracy of the response cannot be easily verified (Jahedi and Méndez 2014). Secondly, unlike logs, they are unable to provide continuous and real-time information about users.

An alternative approach to measuring such constructs in educational technologies and more broadly information systems is by collecting and analysing neurophysiological data from participants. The use of neurophysiological measurements (denoted as neuro measurements in the remainder of the paper) in information systems has recently gained attention resulting in the development of a new interdisciplinary field of research called Neuro-Information-Systems (NeuroIS) that "relies on knowledge from disciplines related to neurobiology and behaviour, as well as knowledge from engineering disciplines" (Riedl and Léger 2016). At its core, NeuroIS uses neuro measurement instruments to collect and analyse neurophysiological data that are

commonly related to the Central Nervous System (CNS) or the Autonomic Nervous System (ANS) from participants. For example, neuro measures can be collected from various devices or instruments such as electroencephalogram (EEG), eye-tracker, or functional magnetic resonance imaging (fMRI). A particular device, in turn, can provide various neuro measures; for example, an eye-tracking device can measure pupil dilation and blink rate. These neurophysiological data are employed towards approximating constructs with the aim of advancing the design, development, use, acceptance, influence and adaptivity of information systems (Fischer et al. 2019; Brocke et al. 2013). For example, cognitive load and attention are commonly studied constructs (Fredricks et al. 2004; Dunn and Kennedy 2019). Although the relationship between neuro measurements and constructs is highly complex, NeuroIS research indicates that capturing psychological constructs overcomes the two drawbacks of subjective measures. In particular, they have the capacity to (1) quantify constructs that cannot be reliably measured on the basis of self-reporting techniques and (2) provide continuous and near real-time information about a user's psychological constructs (Riedl and Léger 2016).

Recent advancements in the development of neuro measurement instruments are making them increasingly more reliable, portable and affordable, thus providing a potential avenue for adoption in many new domains, including: medicine by gathering and visualising data for applications such as healthcare monitoring (Kim et al. 2019) and rehabilitation or remotely notifying the state of patients (Furtado and Trobec 2011); entertainment by detecting the affective impacts of video content (Fleureau et al. 2012); game industry by providing evaluation tools to help increase engagement in gameplay (Nacke 2011). The use of cognitive-state tracking technologies has received particular interest for notifying users when their attention decreases to a potentially dangerous level in high-risk activities (\*Derosière et al. 2014)<sup>1</sup>. For instance, detection of abnormal and hazardous activities in a timely manner such as recognising drowsy driving (Byong-Hoon 2008; Barr et al. 2009), reduction of human errors such as operational delays and fatigue-related accidents with the aim of increasing awareness and efficiency of aerial system operators (\*Mannaru et al. 2016), and enhancement in air travel safety using a portable brain-imaging device to avoid overload on the operators (\*Harrison et al. 2014).

Although the use of neuro measurements in the context of educational technologies is on the rise (Hofkens and Ruzek 2019; Ng and Ong 2018), a broad understanding of the involved measurements, methods, target constructs, outcomes and implications for educational technologies is largely unknown. This paper aims to synthesise recent developments in using neuro measurements in education, specifically in higher education. We conduct a systematic literature review (SLR) with a focus on the following three themes: measurements (the type of measurement instruments used within these systems), experimental settings with related considerations (e.g., participant recruitment, type of experiment, ethical issues, intrusiveness, and reproducibility) and finally constructs and intentional outcomes. These themes are chosen to shed light on the methodological and practical aspects associated with

---

<sup>1</sup>References marked with an asterisk (\*) are from the final articles retained for this SLR.

employing neurophysiological data to design, study, or enhance educational technologies. An interactive visualisation tool to support the SLR has been developed. This tool allows the reader to dynamically filter the figures and tables presented in the paper according to various parameters. The tool is available at the following URL <http://neuro-in-higher-education-slr.herokuapp.com/>.

In what follows, we first outline the research questions addressed in this SLR. Next, we describe the methods undertaken for conducting this review. Subsequent sections report our findings related to the three main themes of measurements, experimental Settings, and constructs and outcomes. Then, an overall discussion is provided about the opportunities and challenges of using neuro measurements and instruments in educational technologies. Finally, we provide concluding remarks.

## Research Questions

To provide insights on the different aspects of how neurophysiological data have been used in educational technologies, we have grouped the research questions to be investigated into three main categories: Measurements, Experimental Settings, and Constructs and Outcomes. Table 1 provides an overview of the research questions. The data type is classified into two classes in this table based on how the extraction process is done on each sub-question. The class “Predefined” refers to data categorisation based on an initial list derived from predefined codes or taxonomy in the literature, and the class “Exploratory” refers to data categorisation using a bottom-up approach relied only on the extracted data from the content of the selected papers. Also, a summary of the extracted data is given for each research question.

**Measurements** Questions related to this theme aim to provide an overview of the instruments applied to neuro measurements and accompanied non-neuro measurements employed in higher education. Q1.1. employs the neuro measurement categorisation introduced by Riedl and Léger (2016) to investigate which neuro measurement instruments are used in the selected articles; Q1.2. employs the measurement modalities introduced by Chen et al. (2016) to investigate which non-neuro measurements complement the neuro measurements employed in the selected articles; and finally, Q1.3. investigates the different neuro and non-neuro combination of measurement modalities that are used in the selected articles.

**Experimental Settings** Questions related to this theme aim to provide an overview of how studies that have used neuro measurements using educational technologies in higher education are conducted. Q2.1. investigates the setting ( field or lab experiment) of the selected studies. Q2.2. uses the experimental design categories introduced by Campbell and Stanley (2015) to investigate the type of experimental design in the selected studies; Q2.3. explores the types and methods of recruitment of participants for the selected studies; Q2.4. investigates the ethical clearances and considerations sought by the selected studies; Q2.5. explores the intrusiveness of the

Table 1 Research questions under investigation

Sub-question	Data Type	Extracted Data Summary
1. Measurement: What instruments are applied to neuro measurements and accompanied non-neuro measurements in higher education studies?		
“(Q1.1) Which Neuro Measurement Instruments are Used?”	Predefined	ANS, CNS, and hormone based measurement instruments.
“(Q1.2) Which Non-neuro Measurements are also Involved?”	Exploratory	Behavioural, performance, and subjective measures.
“(Q1.3) Which combinations of measurement modalities are used in these studies?”	Exploratory	Mapping neuro measurement instruments and non-neuro measurement modalities.
2. Experimental Settings: How studies are conducted using neuro measurements in higher education?		
“(Q2.1.) What is the Setting of the Study?”	Predefined	Field or lab experiment.
“(Q2.2.) What Type of Experimental Designs are Conducted?”	Predefined	Pre-, true-, and quasi-experimental.
“(Q2.3.) Who were the Participants and how were they Recruited?”	Exploratory	E.g., students or staffs who were volunteered or paid.
“(Q2.4) What Ethical Considerations were Sought or Reported by the Studies?”	Predefined	Ethics approval, consent form, or ethics not explicitly reported.
“(Q2.5.) How Intrusive were the Conducted Experiments?”	Exploratory	Four levels of low-, medium-, high-, and very high-intrusive.
“(Q2.6.) How many Participants were Involved in the Neuro Measurement Experiments of the Study?”	Exploratory	Distribution of participants’ size per neuro measurement instrument.
“(Q2.7.) How is the Reproducibility of the Conducted Experiments?”	Predefined	Availability of algorithms and accessibility of data sets.

**Table 1** (continued)

Sub-question	Data Type	Extracted Data Summary
3. Constructs and Outcomes: What psychological constructs are studied and what outcomes are achieved using neuro measurements in higher education?		
“Q3.1. Which Psychological Constructs are Studied?”	Exploratory	Cognitive, non-cognitive, and meta-cognitive constructs.
“(Q3.2.) Which Neuro Measurement Instruments are used to Capture Different Psychological Constructs?”	Exploratory	Mapping neuro measurement instruments and psychological constructs.
“(Q3.3.) What is the Purpose of the Study?”	Exploratory	Monitoring, performance estimation, feedback, and adaptive system.

experiments conducted by the selected studies; Q2.6. examines the number of participants involved in each of the selected studies; and finally, Q2.7. investigates the reproducibility of the conducted experiments of the selected studies.

**Constructs and Outcomes** Questions related to this theme aim to provide an overview of the constructs studied and outcomes achieved by the studies that have used neuro measurements using educational technologies in higher education. Q3.1. explores the psychological constructs studied in the selected papers; Q3.2. investigates which neuro measurement instruments are used for capturing different psychological constructs in the selected studies; and finally, Q3.3 investigates the purposes of the selected studies.

## Methods

SLRs have been undertaken in several fields and provide a reliable means of navigating large bodies of knowledge with the aim of understanding specific outcomes from literature through a systematic process of identifying, analysing and synthesising. A number of approaches to SLRs have been proposed (e.g., Higgins 2011; Kitchenham and Charters 2007; Moher et al. 2015). We rely on the SLR approach proposed by Kitchenham and Charters (2007), as it provides comprehensive guidelines adapted from social science to be utilised in software engineering which is very close to the scope of this review. The steps in the systematic literature review method are documented below.

## Queries and Search Strategy

Based on the aim of the SLR, we created a query to identify papers that have used neuro measurements in educational technologies to enhance teaching and learning. As such, our search query was designed to be inclusive, combining keywords using the Boolean AND operator between a set of query keywords for finding papers on educational technologies and another set of keywords identifying papers that have used a neuro measurement. The term used for identifying papers on educational technologies is (“education\*” OR “learn\*” OR “teach\*”) AND “technolog\*”, where \* represents a wildcard which can be replaced by zero or more non-space characters. For the term identifying the neurophysiological measurement instruments, we followed the terms provided by (Riedl et al. 2017), which is a logical OR separated list of the terms as follows: (“Brain”, “Diffusion Tensor”, “EEG”, “fMRI”, “Infrared”, “MEG”, “Morpho\*”, “NIRS”, “Positron emission”, “Transcranial”, “Dermal”, “ECG”, “EKG”, “Electrocardiogram”, “Electromyography”, “Eye”, “Facial”, “Galvan\*”, “Heart”, “Muscular”, “Oculo\*”, “Skin”, “Blood”, “Hormone”, “Saliva”, “Urine”). This query was run on 8 April 2019.

The electronic bibliographic databases searched included those that are indexed through Scopus and ProQuest. These databases were chosen due to their broad

coverage of peer-reviewed academic journals relevant to the topic of our study as well as convenient search engines for conducting an SLR.

## Document Selection

Running the query on abstracts resulted in 10,822 articles from the selected databases. In order to define the appropriate scope, we limited the timeline of our study to five years (2014 – 2018); a total of 4723 papers were found to be published in this period. Based on screening the title and abstract of these articles by one researcher, 396 papers that (1) were written in English, (2) referred to at least one neuro measurement instrument term and (3) had a focus on teaching and learning were included for further screening. The second criterion in the inclusion criteria was introduced because numerous articles that resulted from the initial query contained expressions such as “eye-opening experience” and “at the heart of education” or health-related terms such as “skin cancer” and “heart disease” in their abstracts that had no relation with neuro measurements. The following exclusion criteria were then applied while reviewing the full text of these articles.

**Exclusion Criteria 1** Articles that did not collect or analyse neurophysiological data in their study. This excluded papers that (1) referred to the use of neuro measurements, but only reported results from surveys or (2) used eye-tracking devices to solely collect and report results on visual fixation as eye movement, which is generally voluntarily and controlled by the user rather than the autonomous nervous system (Andreassi 2010), and hence not considered a neuro measurement. A total of 130 articles were removed based on this exclusion criteria.

**Exclusion Criteria 2** Articles that did not have a focus on higher education. This included papers where (1) participants of the study were not in tertiary education or (2) participants of the study were in tertiary education, but the aim of the paper was not to improve teaching and learning (e.g., the aim of the paper was to introduce a new face recognition algorithm, which used students from tertiary education in their study). Although incorporating non-tertiary groups would provide opportunities for more analysis to be made on a broader area of the learning environments, the scope would have increased by 113 articles if all levels of education were included. Therefore, we limited the focus of our SLR to higher education to improve the quality of the collected information. Also, we note that the focus of many of the studies in early childhood was on special needs and neurological issues which is out of the scope of this paper. Finally, the use of educational technologies in higher education seems to be contextualised differently on many aspects related to self-regulated learning, acceptance of technology and use of learning analytics. A focus on the difference of use cases in higher education vs younger age groups across the research questions would have significantly increased the length of the manuscript.

**Exclusion Criteria 3** Articles that were not a primary study (e.g., conceptual papers that included no neurophysiological data, duplicates, or papers where a progressed version of the same project from the same author(s) was published at a later date



between 2014 – 2018). A total of 96 articles were removed based on this exclusion criteria.

**Quality Assessment Criteria** To avoid any bias and subjective judgement, additional quality criteria beyond considering only articles that have been peer-reviewed were not considered.

The snowballing approach of Wohlin (2014), was applied to the remaining 57 papers. In particular, the 1981 references of the selected papers and the 353 citations to the selected papers (as of 29 May 2019) were analysed using the same inclusion and exclusion criteria, which resulted in the addition of 26 extra articles. The final number of selected articles is 83 of which 34 are published in conferences, 46 are published journal articles, and 3 are published as a book chapter. Figure 1 presents an overview of the selection process of papers in our SLR.

The initial screening based on titles and abstracts was conducted by one researcher. The screening based on the full text for the exclusion criteria and the snowballing procedure was also conducted by one researcher; however, as a reliability measure, a 10% sample was screened independently by a second researcher. The kappa agreement for these two parts between the independent screening of the two researchers were 0.92 and 0.85 for the exclusion criteria and the snowballing procedures, respectively. Data extraction from the final selected articles was conducted independently by two researchers, where any disagreement was resolved via a discussion between the two researchers or in consultation with the other authors, if necessary.

## SLR Findings Related to Measurements

### (Q1.1) Which Neuro Measurement Instruments are Used?

Neuro measurements are categorised into two major groups. First, measurements using neuro measurement instruments that are related to the autonomic nervous system (ANS) such as facial expression recognition, eye-based measurement, heart rate, skin response and blood pressure. The second group considers measurements using neuro measurement instruments that are directly related to the central nervous system (CNS; related to brain and spinal cord) such as electroencephalography

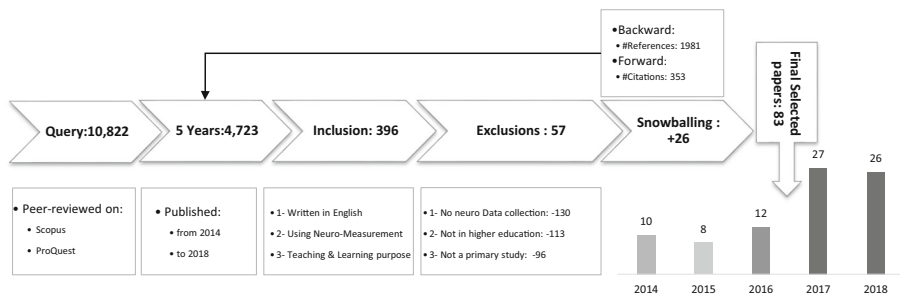


Fig. 1 Brief summary of the SLR procedure

(EEG), functional magnetic resonance imaging or functional MRI (fMRI), near-infrared spectroscopy (NIRS) and functional near-infrared spectroscopy (fNIR). In addition, hormone measurements (e.g., using blood and saliva) are also considered (Riedl et al. 2017).

From the selected papers, 58% of the reported measurements relate to ANS and 42% of the reported measurements relate to CNS. None of the selected papers reported the use of hormones. Figure 2 shows the number of studies that have used each neuro measurement instrument. Note that the sum of the total adoptions of neuro measurement instruments in Fig. 2 (i.e., 107) is greater than the total number of papers indicated in Fig. 1 (i.e., 83). This is because 21 out of 83 (25%) studies have reported the use of more than one neuro measurement instrument (see Table 4). 25 studies use **Facial** that refers to using cameras or webcams as neuro measurement instruments for collecting physiological data of facial expressions (e.g., \*Bian et al. 2018; \*Dimililer 2018; \*Nye et al. 2018; \*Sawyer et al. 2018; \*Wei et al. 2017). 15 studies use **Eye** that refers to using desktop eye-trackers or eye-tracking glasses as neuro measurement instruments for collecting physiological data of eye-related measures such as pupil dilation (e.g., \*Stuijzand et al. 2016; \*Mannaru et al. 2016; \*Menekse Dalveren and Cagiltay 2018; \*Prieto et al. 2018) and blink rate (e.g., \*Durall and Leinonen 2015; \*Liu et al. 2018a; \*Zlokazov et al. 2017). 11 studies use **Heart** that refers to using medical heart rate monitoring sensors or heart rate monitoring wristbands (e.g., smartwatch) as neuro measurement instruments for collecting physiological data of heart-related measures such as heart rate (e.g., \*Peng and Nagao 2018; \*Pham and Wang 2015) and heart rates variability (e.g., \*Chen and Wu 2015; \*Thompson and McGill 2017). 9 studies use **Skin** that refers to using galvanic skin response (GSR) sensors or temperature sensors as neuro measurement instruments for collecting physiological data of electrodermal activity-related measures such as skin conductance (e.g., \*Edwards et al. 2017; \*Medina et al. 2018) and skin temperature (e.g., \*Blanchard et al. 2014). 2 studies use **Blood** that refers to using blood pressure monitoring device as a neuro measurement instrument for collecting physiological data of blood pressure (\*Ray and Chakrabarti 2016; \*Siqueira et al. 2017). 38 studies use **EEG** that refers to using single EEG sensor headsets or head-mount multi-channel EEG sensors as neuro measurement instruments for collecting physiological data of EEG-related measures such as five primary bands of frequency known as

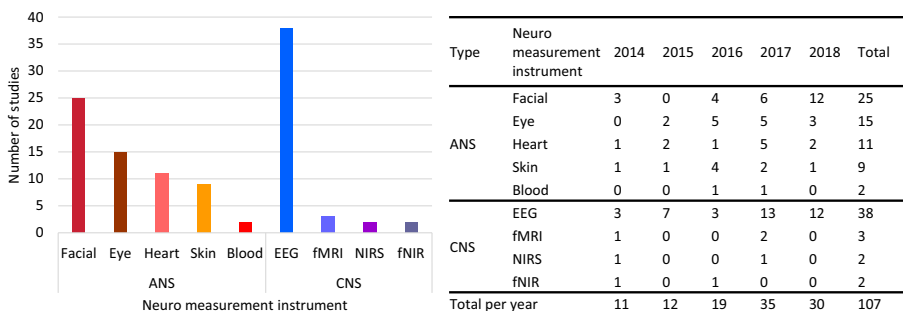


Fig. 2 The number of studies per neuro measurement instrument

alpha, beta, gamma, theta, and delta (e.g., \*Lin and Kao 2018; \*Nor and Salleh 2015; \*Qu et al. 2018b; \*Spüler et al. 2017; 2018), event-related potential (e.g., \*Batterink and Neville 2014; \*Varga and Bauer 2017; \*Zhang 2018), and power spectrum (e.g., \*Dan and Reiner 2018; \*Hu and Kuo 2017; \*Hubbard et al. 2017; \*Sethi et al. 2018). 7 studies use neuroimaging devices as neuro measurement instruments for collecting physiological data of brain activity measuring blood flow and blood / hemoglobin oxygenation level: 3 studies use *fMRI* that refers to using MRI scanners (\*Bridge et al. 2017; \*Gershman et al. 2017; \*Wang and Voss 2014); 2 studies use *NIRS* that refers to using near-infrared spectroscopy (\*Derosière et al. 2014; \*Tobita 2017); and 2 studies use *fNIR* that refers to using functional near-infrared spectroscopy (\*Harrison et al. 2014; \*Yuksel et al. 2016). Figure 2 also indicates that the most frequent neuro measurement instrument used is EEG with an adoption rate of 35%, followed by Facial (24%), and Eye (12%). An interesting observation is that the use of Facial has increased considerably during 2018, while the use of EEG has not.

### (Q1.2) Which Non-neuro Measurements are also Involved?

The literature indicates that neuro measurements are frequently complemented by a range of non-neuro measurements relating to learners. We use the measurement modalities introduced by Chen et al. (2016) to categorise non-neuro measurements into the following three groups:

- Behavioural measures are generally considered voluntarily controlled actions such as eye movements (e.g., saccades and fixation), body movements (e.g., head pose and gesture), and linguistic features.
- Performance measures are related to the accuracy and speed of the responses such as test scores or error rate and speed or reaction time.
- Subjective measures are self-reported assessments using research instruments such as questionnaires, surveys, and interviews.

We found 74 studies out of the 83 selected papers used non-neuro measurements alongside the use of neuro measurements. Table 2 shows the number of studies that have used each of the non-neuro measurement modalities in combination with neuro measurements. From the total selected papers, 35% use behavioural measures, 60% performance measures and 53% subjective measures.

Table 3 provides a drill-down into Table 2. It shows the number of studies that consider each sub-category of non-neuro measurements. As mentioned in Table 1, this sub-categorisation is based on an exploratory data extraction using a bottom-up

**Table 2** Number of studies per non-neuro measurement modality

Non-neuro measurement	2014	2015	2016	2017	2018	Total
Behavioural	4	1	5	11	8	29
Performance	10	3	8	16	13	50
Subjective	7	7	8	10	12	44

**Table 3** Reported sub-categories of non-neuro measurements in the selected papers

Non-neuro measurement	sub-category	2014	2015	2016	2017	2018	Total
Behavioural	Eye movements	2	1	4	6	4	17
	Body movements	1	0	2	4	1	8
	Linguistic features	2	0	0	5	3	10
Performance	Accuracy	9	3	8	16	13	49
	Speed	2	0	2	1	1	6
Subjective	Questionnaires/Surveys	7	7	6	10	12	42
	Interviews	1	2	3	2	2	10

approach where we only relied on the reported non-neuro measurements in the selected papers. Behavioural measures are related to collecting data of learners' interactions such as eye movement's parameters other than pupil dilation and blink rates like gaze or fixation (e.g., \*Menekse Dalveren and Cagiltay 2018; \*Muldner and Burleson 2015; \*Pantazos and Vatrpu 2016; \*Prieto et al. 2018; \*Rusák et al. 2016; \*Wang and Hsu 2014; \*Zlokazov et al. 2017), body movement's parameters like head pose and gesture (e.g., \*Chen et al. 2016; \*Kanimozhi and Raj 2017; \*Liu et al. 2018b; \*Monkaresi et al. 2017; \*Vail et al. 2016), and linguistic features like wordometer and language capabilities (e.g., \*Kise 2017; \*Batterink and Neville 2014; \*Kepinska et al. 2017; \*Prat et al. 2018; \*Qi et al. 2017; \*Zhang 2018). It is worth noting that applying only behavioural aspects of eye-tracking parameters such as fixation, saccades, and area of interests (AOI) have tended to increase in recent years, which is outside of the scope of this review. For more information, refer to reviews on the use of eye-tracking in different learning environments (Latif 2019; Alemdag and Cagiltay 2018; Ashraf et al. 2018; Chen et al. 2017; Leggette et al. 2018; Luo et al. 2017; O'Meara et al. 2015; Prichard and Atkins 2016; Mavrikis and et al. 2016; Wu 2012; Yang et al. 2018). Performance measures are related to considering learners' skills in terms of accuracy and speed during a task such as reaction time (e.g., \*Derosière et al. 2014; \*Stuijzand et al. 2016; \*Katona and Kovari 2016), test scores (e.g., \*Chen et al. 2017; \*Dan and Reiner 2018; \*Lin et al. 2014 \*Özek 2018), or error rates (e.g., \*Kublanov et al. 2017; \*Zhang 2018). Subjective measures are related to gathering self-reported data directly from learners through questionnaires and surveys such as National Aeronautical and Space Administration Task Load Index (NASA-TLX) (e.g., \*Dan and Reiner 2018; \*Grafsgaard et al. 2014), five or seven-point Likert-type scale surveys (e.g., \*Wu et al. 2014; \*Edwards et al. 2017; \*Kuo et al. 2017; \*Pantazos and Vatrpu 2016; \*Wang and Hsu 2014; \*Yang et al. 2018), and interviews (e.g., \*Enegi et al. 2018; \*Lin et al. 2014; \*Seugnet Blignaut and Matthew 2017; \*Yuksel et al. 2016).

### (Q1.3) Which combinations of measurement modalities are used in these studies?

Table 4 shows the total number of studies that use a particular number of neuro measurement instruments with a particular combination of the non-neuro measurements

**Table 4** Number of studies that used neuro measurement instruments with different non-neuro measurement modality combinations

# neuro instrument/s	Non-neuro measurements								Total
	None	Behaviour	Performance	Subjective	B+P	B+S	P+S	B+P+S	
1	9	4	11	10	7	4	13	4	62
2	0	1	2	2	4	1	4	4	18
3	0	0	1	0	0	1	0	0	2
4	0	0	0	1	0	0	0	0	1
Sum	9	5	14	13	11	6	17	8	83

(i.e., B: Behavioural, P: Performance, and S: Subjective). The first row shows the number of papers that consider one neuro measurement instrument in their study, which is the most common approach with 63 studies out of 83 (76%). The first column in this table shows that 9 papers do not report the use of any non-neuro measurement and only rely on a single neuro measurement. The next three columns in Table 4 show the number of studies that consider only one of the non-neuro measurement modalities in their study and the next three columns after these show the number of those papers that consider two non-neuro measurement modalities in their study. As it can be seen from the last column in this group, 8 studies consider all three non-neuro measurement modalities in their experiments alongside neuro measurements.

Figure 3 provides a drill-down into Table 4. It provides an overview of the simultaneous use of neuro measurement instruments and non-neuro measurement modalities in the selected articles. For example, the upper-left corner cell demonstrates that the combination of facial and behavioural measurements has been used in 9 of the selected articles. As it is already expected from the results reported in Fig. 2, facial measurements and EEG are the most frequent ANS and CNS related neuro measurement instruments used in multimodal studies.

Neuro measurement instruments	ANS	Facial	9	13	17									
		Eye	9	8	5	1								
		Heart	3	9	8	3	2							
		Skin	4	6	7	6	1	2						
		Blood	0	1	2	1	0	2	1					
	CNS	EEG	15	21	17	2	4	2	1	0				
		fMRI	2	3	0	0	1	0	0	0	0			
		NIRS	0	2	1	0	0	0	0	0	0	0		
		fNIR	0	2	2	0	0	0	0	0	0	0	0	
	Behavioural	Performance	Subjective	Facial	Eye	Heart	Skin	Blood	EEG	fMRI	NIRS	fNIR		
				ANS						CNS				
	Non-neuro measurements				Neuro measurement instruments									

**Fig. 3** Simultaneous use of neuro measurement instruments and non-neuro measurement modalities

## SLR Findings Related to Experimental Settings

### (Q2.1.) What is the Setting of the Study?

The setting of a study is classified as either a field experiment (i.e., conducted in a natural setting) or a lab experiment (i.e., conducted in a tightly controlled environment). The controlled environment refers to the use of isolated laboratory settings in those studies that ran their experiments outside of the natural setting of the course where participants were asked to do a task individually while equipped with different measurement instruments. Merely 11 (13%) studies of this SLR report that their experiments were held in natural settings or traditional classrooms such as equipping a classroom with smart devices like cameras, microphones, tablets, and wearable sensors in (\*Liu et al. 2018b) to measure students' learning states using their heartbeats, blinks, facial expressions, and quiz scores. The remaining 72 are conducted in a variety of experimental settings in a controlled environment. Examples include investigating the impact of different factors in diverse learning methods, such as computer-based learning (\*Nye et al. 2018; \*Wang and Hsu 2014), online learning (\*Edwards et al. 2017; \*Kong and Li 2018; \*Lin and Kao 2018), and game-based learning (\*Samah et al. 2018; \*Wu et al. 2014).

### (Q2.2.) What Type of Experimental Designs are Conducted?

Campbell and Stanley (2015) categorise experimental designs into three general groups: true-experimental, quasi-experimental and pre-experimental. The main feature of a true-experiment is assigning a control group against the experimental group with the randomisation of the participants as the means of validating the results of the experiment. 13 studies from the selected papers used a true-experimental design. For example, \*Spüler et al. (2017) divided participants into two groups randomly where the experimental group used an adaptive learning environment based on a neuro measurement using EEG, and the control group used an adaptive learning environment based on a performance measurement using error-rate with similar difficulty level exercises. Quasi-experiments often aim to validate the results of the experiment using a controlled group; however, the assignment of participants to groups may use some criterion other than random. 28 studies from the selected papers used a quasi-experimental design. For example, \*Wong et al. (2016) used a non-randomised controlled experiment to evaluate the effectiveness of different reading strategies by measuring pupil diameter. Pre-experiments are often the results of passive observational case studies or static comparisons of pre- and post- results of a single group. 42 studies from the selected papers used a pre-experimental design. For example, \*Bian et al. (2018) collected Facial data from all participants while watching videos related to a MOOC.

### (Q2.3.) Who were the Participants and how were they Recruited?

53 studies out of 83 (64%) report recruiting university students for their experiments. Among them, 17 report using undergraduate and 5 report using graduate students. 9

studies use professionals such as university staffs, teachers, and employees, where 3 studies only use teachers as their subjects. 23 studies do not mention the type or profession of their participants. In terms of how participants were recruited, 14 studies report that their subjects have volunteered to participate in their experiments and 15 studies report that they have paid their participants with gifts or rewards. Only 3 studies mentioned recruiting via advertisements or flyers and 2 studies via Emails.

#### **(Q2.4) What Ethical Considerations were Sought or Reported by the Studies?**

Advancement in information technology has provided educational researchers with abundant data on students. Use of educational data has embraced many opportunities in higher education; however, even with the best intentions, data can be misinterpreted or misused. As such, there is an obligation that researchers handle educational data with care and to ensure that it is being used ethically and responsibly. The ethical considerations behind using student and educational data have been well studied in the field of learning analytics. A recent discussion paper from this field (Artífice et al. 2017) raises awareness on the importance of handling student data with care, providing insightful guidelines, protocols and processes for the ethical use of educational data.

Based on our review, in more than half of the selected papers (49 studies: 59%), no report of an ethical consideration was explicitly declared (we note that many of these studies might have received an institutional review board (IRB) approval, which is a requirement for conducting studies on human subjects in many countries, without explicitly mentioning it). In the remaining 34 (41%) of the studies, two main types of ethical considerations are reported: consent forms and committee of ethics approval. 18 studies (22%) have reported the use of consent forms without referencing ethical approvals. Two of these papers refer to following a standard code of ethics in their studies; \*Kepinska et al. (2017) clarified that their experiment was conforming to the ethical code of a university faculty of humanities and \*Poulsen et al. (2017) emphasised that their study was exempt from ethical committee processing by Danish law due to non-invasive experiments on healthy subjects. The remaining 16 studies have reported obtaining ethical clearance for their study. 13 of these studies mention providing a consent form for participants. In the other three cases, \*Thompson and McGill (2017) stated that participants were provided with required information before the task and ethics approval was also obtained, \*Liu et al. (2015) mentioned following a code of ethics from the medical association named declaration of Helsinki, and \*Seugnet Blignaut and Matthew (2017) just reported an ethics clearance number.

#### **(Q2.5.) How Intrusive were the Conducted Experiments?**

Intrusive measurement refers to the “use of devices or measurement procedures that affect the normal situation of the person, bringing significant impact on the mobility or comfort of the person involved” (Cruz-Cunha 2016). Similarly, “non-intrusive measurement refers to the use of devices or measurement procedures that induce

minimal impact on the person involved” (Cruz-Cunha 2016). In a review on physiological metrics of mental workload, Kramer (1991) proposes that intrusiveness is one of the main criteria for the functional utility of physiological measures that can be used in the selection process of suitable measures for different applications. He refers to intrusiveness as “the capability of measuring mental load without interfering with the operator’s performance on the primary task”. In his review, several techniques of CNS- and ANS-related measurements are evaluated based on the intrusiveness alongside other criteria such as sensitivity, diagnosticity, and reliability. Dealing with the issue of intrusiveness plays an essential role in increasing user acceptance and adoption of technology (Teo 2009). As such, we believed it would be important to provide information about the level of intrusiveness of the typical instruments that are used in the selected studies. Since we were unable to find a categorisation on the intrusiveness of the devices from the literature, we applied the following steps to reach a categorisation. First, to become familiar with the levels of intrusiveness, several types of neuro measurement devices including desktop eye-tracker, galvanic skin response (GSR) device, facial expression recognition using cameras, heart rate monitors, single and multi-channel electrodes devices, and fMRI were tried out by one of the authors in real experiments conducted within the schools of Psychology, Neuroscience, Business, and Computer science at The University of Queensland. Next, the following considerations were taken into account in the classification process: (1) the descriptions of how the device operates provided by the corresponding studies, (2) the impact on the mobility and comfort of the participant during the experiment as suggested by the provided definition for the intrusiveness (Cruz-Cunha 2016), and the level of performance degradation on the task of interest caused by the measurement device as recommended by Kramer (1991). Then, devices were classified into four levels, i.e., low-, medium-, high-, and very high-intrusive where “low” means no confinement of space and movement and “very high” means severe restrictions on both space and movement. In multimodal measurement studies, the device with higher impacts on mobility and comfort is considered for intrusiveness classification. For example, \*Muldner and Burleson (2015) used three commercial sensing devices for modelling student creativity in a digital learning environment: a desktop eye-tracker, a galvanic skin response bracelet, and a head-mount multi-channel EEG sensor. In this case, the intrusiveness of the head-mount device is considered for classification of the intrusiveness level of the experiment. This is because this device has a higher impact on the comfort level of a user in comparison with either a desktop eye-tracker device or a wristband. Table 5 reports the results.

**Low-intrusive** Camera, eye-tracker and wristband heart rate monitors are classified as low-intrusive. For camera and eye-tracking, the imposed intrusiveness is mostly based on the fact that participants are aware that they are being monitored. For the case of wearing a wristband heart rate monitor, the experience is very similar to wearing a smartwatch which has a very low limitation in terms of participant’s movement; however, as before, participants may feel a low level of discomfort because of being monitored. In total, 26 out of 83 (31%) of the papers were categorised as having a low level of intrusiveness. For studies that have used camera, all but one have captured learners’ facial emotion; the other study (\*Zhang and Shen 2017) used a camera



**Table 5** Intrusiveness of measurement devices

Intrusive	Device Type	Impacts on		# study	Total
		Mobility	Comfort		
Low	Camera	None	Low	16	26
	Desktop eye-tracker	Low	Low	6	
	Wristband heart rate monitor	Low	Low	4	
Medium	Single-channel EEG sensor	Medium	Medium	14	19
	Galvanic skin response bracelet	Medium	Low	3	
	Eye-tracking glasses	Low	Medium	1	
	Arm blood pressure monitor	Low	Medium	1	
High	Multi-channel EEG sensors	Medium	High	24	35
	Multi-channel/electrodes devices	Medium	High	8	
	Galvanic skin response with attached multi-sensors	High	Medium	3	
Very High	fMRI	Very High	High	3	3

of a mobile device for eye pupil diameter detection. For studies that have used eye-tracking, all but one have used a desktop eye-tracker. The other study is classified as medium-intrusive that uses mobile eye-tracking glasses to evaluate instructors' teaching skills in a traditional face to face classroom (\*Prieto et al. 2018).

**Medium-intrusive** Four types of devices including single-channel EEG sensor, galvanic skin response bracelet, arm blood pressure monitor, and Mobile eye-tracking glasses are classified as medium-intrusive. By using a single EEG sensor headset, researchers in the field of education try to reduce the cost and complexity of the multi-channel EEG sensors that are common in medical research. However, it would still affect participants' normal behaviour by limitation on head movements and feeling uncomfortable. For example, \*Zhai et al. (2018) even used a structure in their experiment setting as an adjustable head immobiliser to limit the participants' head movements. Although GSR bracelet is very similar to heart rate monitoring wristband, participants would not similarly be comfortable with it. It is sensible to movement, therefore, requires subjects to minimise their hand movement. While automatic arm blood pressure monitor does not enforce a limitation on movements and also made the blood pressure measurement easier nowadays, it would distract learner because the cuff inflates automatically during the measurement. In total, 19 out of 83 (23%) of the papers were categorised as having a medium-level of intrusiveness.

**High-intrusive** Devices in this class of intrusiveness consist of multi-sensors or -electrodes that attach to specific locations of the participants' body. They considerably affect both comfort and mobility. Another shared attribute among them is that their setup is more time-consuming than the previous classes, and an expert is generally required to attach each sensor correctly in the right place of the body. The most

commonly used device type in this class and also among the whole selected studies in this SLR is multi-channel EEG sensors (24 studies, 29%). They are typically head-mount multi-channel EEG sensors that are not comfortable and are sensitive to movements. In addition, a specialist (usually a neurologist) is needed to accurately attach electrodes to the scalp based on an internationally accepted standard named 10-20 system. There are some other multi-channel/electrodes devices with similar constraints in the selected studies: two studies (\*Derosière et al. 2014; \*Tobita 2017) used multi-channel NIRS; two studies (\*Harrison et al. 2014; \*Yuksel et al. 2016) used multi-channel fNIRS; two studies (\*Ray and Chakrabarti 2016; \*Thompson and McGill 2017) reported using multi-channel multipurpose devices where each channel records measurements of Skin, Heart and Blood; and one study (\*Monkaresi et al. 2017) used a multi-channel electrocardiogram electrodes for Heart. \*Kublanov et al. (2017) used a multi-channel neuro-electrostimulation device which is not a measurement instrument. However, compared to their heart rate monitoring, this stimulation device is considered for intrusiveness classification due to imposing more discomfort. Three studies reported using Skin measurement devices (\*Edwards et al. 2017; \*Landowska and Miler 2016; \*Medina et al. 2018) that are required to attach several electrodes to the fingers of participants' non-dominant hand. In total, 35 studies (42%) among the SLR selected papers are classified as high-intrusive.

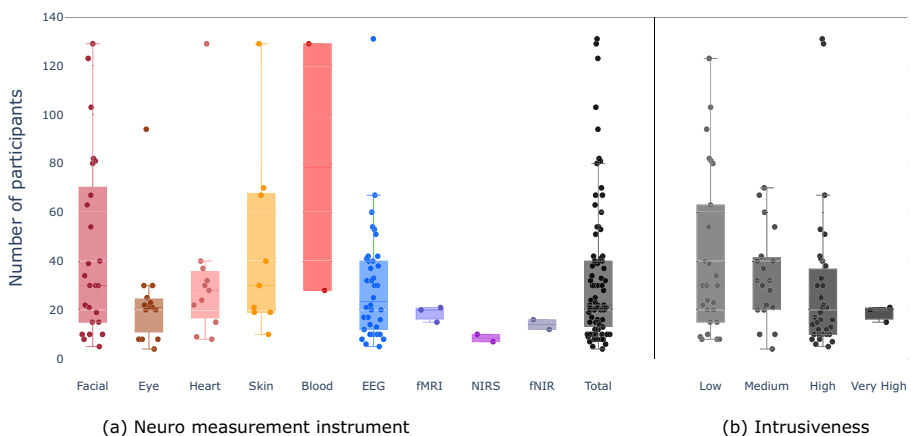
**Very high-intrusive** The fMRI device is classified to be the most intrusive device, considering mobility and comfort, among the applied measurement devices in the selected studies in this review. Participants of the fMRI studies in our selected studies were asked to perform learning tasks in a close-fitting medical device chamber while having minimal movement. Two of the three papers that have used fMRI have reported challenges in conducting experiments and collecting data as they had to exclude results of some of their participants due to their excessive movement during the experiment (\*Wang and Voss 2014; \*Bridge et al. 2017).

#### **(Q2.6.) How many Participants were Involved in the Neuro Measurement Experiments of the Study?**

In this section, we report on the number of participants who took part in a neuro measurement experiment and that their data was used in the reported results. To do so, the following three criteria are employed: First, in studies that reported a technical issue with data collected from some of the participants, the number of participants whose data was used in the final results are reported. For example, \*Stuijzand et al. (2016) stated that they had a total of 92 participants, of which only 67 granted permission for their data to be used for research. Only 10 of these participants were recruited for an experiment that collected neurophysiological data. Data from two of these participants were excluded due to technical issues in the data collection process. As a result, they only reported empirical outcomes using neuro measurements on 8 participants. Second, in multimodal studies where performance, behaviour, and subjective measures were utilised alongside the neuro measures, only the number of participants who took part in a neuro measurement experiment are reported. For example, \*Wang

and Hsu (2014) began their study with 189 participants. After excluding those with incomplete responses, the remaining 148 participated in a first experiment using a subjective measure, 7-point Likert scale questionnaire. However, only 20 were invited in their second experiment recording EEG brainwave signals. As a result, we report the number of participants to be 20 for this study. Third, in studies that used control and experimental groups, only the number of participants from the experimental group that was involved in a neuro measurement experiment is reported. For example, \*Zhai et al. (2018) recruited 106 participants; however, the number of participants is just reported for the experimental group with 54 students that neuro measurement was administered.

The average participant size for the 83 selected papers is 33, with a standard deviation of 29. In more than three-quarters of the selected studies, the number of participants is less than or equal to 40. Figure 4a reports the distribution of participants' size for each neuro measurement instrument. The maximum number of participants is reported as 131 for an EEG experiment in (\*Bin Abdul Rashid et al. 2015), and the minimum number of participants is reported as 4 for an Eye experiment using eye-tracking glasses to measure the pupil size and eye movement in (\*Prieto et al. 2018). Out of the 10 studies with the highest number of participants, 7 of them have used facial measurements. This is not surprising as instruments that use facial measurements are reasonably priced and generally have a low intrusiveness level. On the other hand, the size of participants for fMRI, NIRS, and fNIR studies in the SLR selected papers is generally smaller. The median number of participants in the fMRI, NIRS, and fNIR studies are 20, 9, and 14, respectively. This can be explained by the high cost associated with running these experiments as well as their high level of intrusiveness. Studies using Heart, Skin and EEG instruments have a medium size of participants with a median of 28, 30, and 24, respectively. Interestingly, only two studies used Blood measurements with very different participant sizes.



**Fig. 4** Distribution of participants' size in the selected papers: **a** for each neuro measurement instrument, **b** based on the intrusiveness of the applied neuro measurement instrument

**Table 6** Reproducibility of the studies in the SLR

Algorithms	Data set	
	Accessible	Not Accessible
Available	1	1
Clearly explained	2	13
Not available or explained	5	61

Figure 4b shows the association between the intrusiveness of the neuro measurement instruments and the number of participants in the selected studies. It demonstrates that the lower number of participants is significantly associated with the level of intrusiveness.

### (Q2.7.) How is the Reproducibility of the Conducted Experiments?

The reproducibility of the experiment is reported from the provided information in the selected papers of the SLR. Table 6 shows the number of studies in terms of availability of the data collected during the experiment and the processes or algorithms used for processing the data. Only two studies provided information for accessing their processes and algorithms. One of these studies, \*Prieto et al. (2018), shared parts of their data set that had been anonymised. In the other paper, the results of 10 open source algorithms on eye-based measures were used and compared (\*Menekse Dalveren and Cagiltay 2018). In total, 8 studies (10%) provide information on how to access their data, 15 studies (18%) provide clear explanations on their utilised algorithms, and there is no clear information on neither algorithms nor data of 61 studies (73%).

## SLR Findings Related to Constructs and Outcomes

The target outcome of using neuro measurements is to understand their relationship with psychological constructs and human behaviour (Andreassi 2010). In educational technologies, neuro measurements are used to improve different learning aspects regarding the psychological constructs of the learner that cannot be evidently observed or inferred by an academic expert due to human limitation (Lane and D'Mello 2019). In this section, we investigate the relationship between neuro measurements and psychological constructs in terms of what psychological constructs are captured using different neuro measurement instruments and how they are adopted to improve or support learning experiences.

### Q3.1. Which Psychological Constructs are Studied?

It is believed that student learning performance is influenced by different underlying psychological constructs such as emotion, attention and cognitive load (\*Durall and Leinonen 2015). However, quantifying this relationship is a challenging task.

Paas et al. (2003) state that traditional performance-based measures cannot necessarily reflect essential information for the mental efficiency of instructional methods while a combination with other types of measures such as the cognitive load has been acknowledged to provide a reliable estimate. Numerous different labels have been used to describe the intended target of collecting neurophysiological data in this SLR selected papers. Based on the reported results of the selected articles, these constructs are categorised into three high-level groups of cognitive, non-cognitive and meta-cognitive constructs. Each of these constructs with references to sample studies is described below.

**Cognitive constructs** The label cognitive applies to skills or states that are predictive of the learning achievements such as intellective abilities, information-processing skills, and subject-matter knowledge (Messick 1979). 51 studies in total among the SLR selected papers explore cognitive-related factors in a learning environment. Attention, the capability of choosing specific data from the enormous and continuous array of sensory inputs (Robins et al. 2019), is the most common cognitive construct studied in 23 articles in our SLR (e.g., \*Chen et al. 2017; \*Derosière et al. 2014; \*Kublanov et al. 2017; \*Pham and Wang 2015; \*Sethi et al. 2018). Cognitive load, which relates to limitations in capacity and duration of working memory while dealing with new information (Robins et al. 2019), is also a well-studied construct which was studied by 14 of our SLR papers (e.g., \*Wu et al. 2014; \*Gazdi et al. 2017; \*Stuijfsand et al. 2016; \*Harrison et al. 2014). Furthermore, different types of skills are also examined in 19 studies that are closely associated with cognitive constructs such as cognitive performance (\*Peng and Nagao 2018; \*Siqueira et al. 2017), language learning proficiency (\*Kepinska et al. 2017; \*Prat et al. 2018; \*Qi et al. 2017), reading skills (\*Qu et al. 2018a; \*Rusák et al. 2016; \*Wong et al. 2016; \*Zhai et al. 2018). In addition, studies that have used terms such as cognitive states (\*Hubbard et al. 2017; \*Pham and Wang 2015), cognitive activity (\*Qu et al. 2018a; \*Zlokazov et al. 2017), and cognitive workload (\*Harrison et al. 2014; \*Mannaru et al. 2016; \*Yuksel et al. 2016) are grouped in this category.

**Non-cognitive constructs** Lipnevich et al. (2013) believe that non-cognitive constructs such as beliefs and emotions equally impacts students' academic achievement as the cognitive ones. They group the inconsistent non-cognitive construct labels in different research disciplines into four domains: (1) attitudes and beliefs such as self-confidence and beliefs on difficulty level of different disciplines, (2) social and emotional qualities such as anxiety, (3) habits and processes such as study habits and time management, and (4) personality traits which refer to individuals' steady behaviours and emotions in different situations such as tendency to often experience negative emotions (Neuroticism), tendency to be kind (Agreeableness), and tendency to be open to new thoughts and experiences (Openness). 39 studies use neuro measurements to find non-cognitive factors of the users' psychological constructs. The social and emotional qualities were extensively investigated in 27 studies considering the relationship between emotions and learning activities (e.g., \*Kanimozhi and Raj 2017; \*Landowska and Miler 2016; \*Liu et al. 2015; \*Özek 2018; \*Ray and Chakrabarti 2016; \*Wu 2017). Many of the papers grouped in this category (e.g.,

\*Bian et al. 2018; \*Chen et al. 2016; \*Grafsgaard et al. 2014; \*Lin et al. 2014; \*Manseras et al. 2018) relate their work to affective computing, which generally focuses on development of information systems that can recognise and understand human emotions (Picard 2000). Habits and processes were studied by 8 papers (e.g., \*El-Abbasy et al.; 2018; \*Liu et al.; 2018b; \*Manseras et al.; 2018; \*Monkaresi et al.; 2017; \*Whitehill et al. 2014). Attitude and beliefs such as learning styles are examined in 3 studies (\*Bin Abdul Rashid et al. 2015; \*Edwards et al. 2017; \*Tobita 2017). Finally, 3 studies (\*Gershman et al. 2017; \*Muldner and Bursleson 2015; \*Yang et al. 2018) explored personality traits such as creativity and imagination.

**Meta-cognitive constructs** Two general aspects of metacognition are knowledge about cognition and self-regulation of cognition (Pintrich 1999). “Knowledge of cognition refers to how much learners understand about their own memories and the way they learn and regulation of cognition refers to how well learners can regulate their own memory and learning” (Sperling et al. 2004). 5 studies consider the knowledge about cognition. For example, \*Lin and Kao (2018) try to facilitate users’ self-awareness of mental efforts in online learning contexts using EEG. They have tried to enable automatic feedback in synchronous and asynchronous learning contexts, especially for MOOCs. \*Durall and Leinonen (2015) present a tool to support awareness about learning activity using EEG data. 12 studies in total among the SLR selected papers explore meta-cognitive-related factors in a learning environment. 7 studies consider the self-regulation of cognition. For example, \*Wang and Voss (2014) attempt to link strategic exploration decisions during learning to quantifiable information. They aim to advance the understanding of adaptive behaviour by identifying the distinct and interactive nature of brain-network contributions to decisions using fMRI. \*Wong et al. (2016) assess whether pupil diameter can be used to distinguish between the uses of different reading strategies and whether it is linked to the quality and effectiveness of the strategy in terms of learning gains.

Table 7 shows the number of studies on each of the psychological constructs in our SLR. Note that the sum of the total studies reported in this table (i.e., 109) is greater than the total number of selected papers. This difference is because of the several selected articles that studied more than one construct. Figure 5 provides an overview of the simultaneous capturing of the psychological constructs in the selected articles. For example, the upper-left corner cell demonstrates that capturing the attention alongside the cognitive load has been in 5 of the selected articles, which is also the most frequent one. It is followed by capturing cognitive load alongside social and emotional qualities as well as capturing attention alongside habits and processes.

### (Q3.2.) Which Neuro Measurement Instruments are used to Capture Different Psychological Constructs?

Table 8 shows the number of times each neuro measurement instrument is applied to capture the above-mentioned constructs in the selected papers. For the multi-modal measurements, all the applied neuro measurement instruments are taken into account for the corresponding construct. For example, \*Liu et al. (2018a) applied

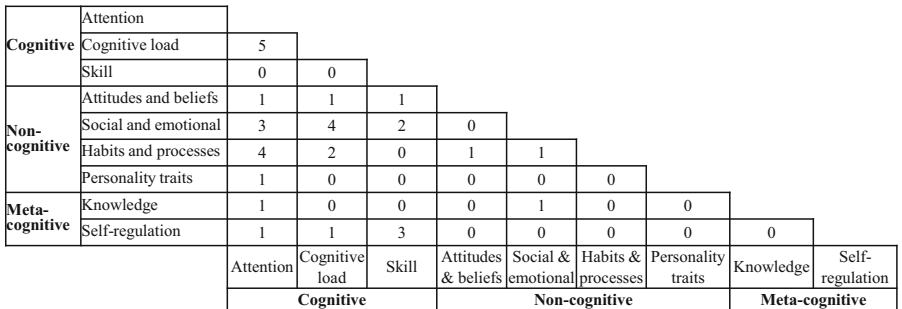
**Table 7** Number of studies per psychological construct

Psychological construct	Construct labels	# Study
Cognitive	Attention	23
	Cognitive load	14
	Skills	19
Non-cognitive	Attitudes & beliefs	3
	Social & emotional	27
	Habits & processes	8
	Personality traits	3
Meta-cognitive	Knowledge about cognition	5
	Self-regulation of cognition	7

facial expressions features, blink rates and heartbeats as the neuro measures to identify learners’ emotional state, so this study will contribute to the count of all three of their corresponding neuro measurement instruments (i.e., Facial, Eye and Heart). As can be seen, EEG is most frequently used in capturing both cognitive constructs such as the attention and meta-cognitive constructs such as the demonstrable learner knowledge about cognition or self-awareness of mental effort. The facial expression recognition has the highest adoption rate in capturing the non-cognitive constructs, specifically the emotions. Meta-cognitive constructs have been attempted to be captured with both ANS and CNS neuro measurement instruments.

**(Q3.3.) What is the Purpose of the Study?**

Based on the findings from the intended objectives and the actual reported results of the selected articles, the purposes of the studies are categorised into four main groups: (1) monitoring learners’ psychological constructs, (2) estimating learners’ performance based on neuro measurements, (3) providing feedback/notifications of learners’ current psychological constructs, and (4) developing an adaptive system that changes pedagogical decisions based on the current captured psychological construct from learners. The first row of Table 9 categorises the studies based on their



**Fig. 5** Simultaneous capturing of the psychological constructs

**Table 8** Mapping psychological constructs and corresponding neuro measurement instruments: number of studies capturing each construct using different neuro measurement instruments

Constructs	Neuro measurement instrument									
	ANS					CNS				
	Facial	Eye	Heart	Skin	Blood	EEG	fMRI	NIRS	fNIR	
Cognitive	3	3	4	2	0	15	0	1	0	0
	3	4	3	2	0	5	0	0	2	2
	3	4	3	1	1	10	1	1	0	0
Non-cognitive	1	0	0	1	0	1	0	1	0	0
	18	1	4	5	1	9	0	0	0	0
	5	0	2	1	0	3	0	0	0	0
	0	1	0	1	0	2	1	0	0	0
	1	2	0	0	0	3	0	0	0	0
Meta-cognitive	1	2	1	0	0	4	1	0	0	0



**Table 9** Purposes of the studies in the SLR

	Monitoring	Performance Estimation	Feedback/ Notification	Adaptive System
Intended purpose	32(39%)	17(20%)	15(18%)	19(23%)
Reported outcome	59 (71%)	5(6%)	10(12%)	9(11%)

intended purpose of the study tool. We noticed, however, that the actual reported outcomes did not always align with the intended purpose of the study; therefore, we have also categorised the studies based on their actual reported outcomes within the same table. Results demonstrate that 27 of papers that have the ultimate goal of estimating performance, providing feedback or employing adaptive systems have only reported results of monitoring student psychological constructs while learning. This result is not surprising as having the ability to monitor student learning is a requirement for the development of more advanced tools that can estimate performance, provide feedback or adapt to a user's mental state. However, it does suggest that much of the work presented in the literature is still in the early stages of its development. Each of these general categories with references to sample studies is described below.

**Monitoring** Studies with three similar objectives are classified in this category. First, a group of studies merely records learners' neurophysiological data to capture their psychological constructs. For instance, \*Bueno-Palomeque et al. (2018) use EEG to identify the level of attention in a second language class; \*Landowska et al. (2017) suggest using Facial for automatic emotion recognition in monitoring e-learning processes; \*Poulsen et al. (2017) undertake a study to quantify real-time engagement from EEG recorded in a classroom; another study using various sensors (Eye, Skin and EEG) is performed to model the creativity of students in a digital environment (\*Muldner and Burleson 2015). Second, the majority of this type of research aims to observe or evaluate the impact of applying different methods in education through capturing learner's cognitive constructs. For example, \*Sezer et al. (2015), using EEG, imply that different teaching methods or different course materials such as PowerPoint presentations, digital maps, graphs, and internet would affect the level of attention in students. Similarly, \*Dan and Reiner (2018) examine the effects of 2D displays versus a 3D scenario in reducing the cognitive load using EEG. Third, a number of studies step forward to evaluate or recommend different teaching methods, facilities, and learning environments based on neuro measures analysis. As a case in point, \*Pi and Hong (2016), measuring blink duration using Eye to estimate the mental fatigue, suggest more effective learning outcomes will be achieved presenting both the instructor and PPT slides in video podcasts in comparison with other presentation methods such as instructor without PPT slides, only PPT slides, and the whole classroom; \*Tobita (2017) compares learners' brain activities by analysing blood flow and change in Oxy-Hemoglobin using NIRS to develop effective course design for improving their skills in English as a foreign language; in another study, the quality of instruction delivery is assessed via measuring student engagement using

Facial (\*Manseras et al. 2018). The common feature among them is using neuro measurements to only differentiate between learners' psychological constructs in various conditions without providing any feedback or intervention in a system or content.

**Performance Estimation** Studies in this group consider neuro measurements to estimate users' performance while interacting with learning environments. It is hypothesised that matching the challenges of learning contents with the learner's skills may improve performance and satisfaction (\*Wang and Hsu 2014). For example, an eye-tracking experiment is conducted to understand different skill levels of surgical residents with the aim of developing suitable assessment tools and instructional systems to enhance their skills (\*Menekse Dalveren and Cagiltay 2018); \*Qi et al. (2017) imply that using EEG power in particular frequency bands may predict the performance before the novel language-learning initiates; \*Whitehill et al. (2014) predict a post-test performance using automatic engagement judgements by capturing students' facial expressions.

**Feedback/Notification** Studies in this category aim to establish a system to provide feedback or notification based on neuro measurements to maintain the engagement of learners at a satisfactory level. Riedl and Léger (2016) define the concept of a biofeedback system as (1) recording biological signals from the user, (2) presenting these signals visually/acoustically, and (3) changing the behaviour by the user to control the biological signals. It is assumed that neuro-technology enhanced learning should enable students to realise the different aspects that influence their learning performance (\*Durall and Leinonen 2015). For example, \*Serrhini and Dargham (2017), using an EEG-based attention alerting system, state that personalised feedback plays an essential role to enhance students' attention level; \*Özek (2018) proposes an emotion aware learning management system (LMS) for the instructor to provide effective feedback for students in distance education; and \*Zhai et al. (2018), using eye-trackers and EEG devices, consider biofeedback of learner's mental mechanisms in self-regulated online learning as a replacement for the beneficial learning feedback in a traditional environment by a teacher.

**Adaptive system** Adapting pedagogical decisions based on the current learner state in a learning system is the common goal of researchers in this group. Riedl and Léger (2016) define the concept of a neuroadaptive system as (1) recording biological signals from the user, (2) deriving a mental state by analysing these signals, and (3) the system adapts using the derived mental state. Neuro measurements can be utilised to show the difficulty level of the delivered materials and also help instructors to adjust the content to enhance students' learning (\*Medina et al. 2018). For instance, a system is designed that uses a brainwave signal-based attention promoting mechanism by providing timely assistance in an English listening course (\*Kuo et al. 2017); \*Thompson and McGill (2017) propose an affective platform to understand affective activation and valence for providing real-time support to deal with negative states and also providing guidelines while the learner interacts with the system; and \*Yuksel et al. (2016) present a brain-based adaptive system that changes the difficulty level in a musical learning task based on the users' cognitive workload. In addition, only one

study among the selected papers aims to improve the learning efficiency by directly influence the nervous system, a method called neuro-electrostimulation is proposed to dynamically correct the cognitive abilities using a special field of current pulses (\*Kublanov et al. 2017).

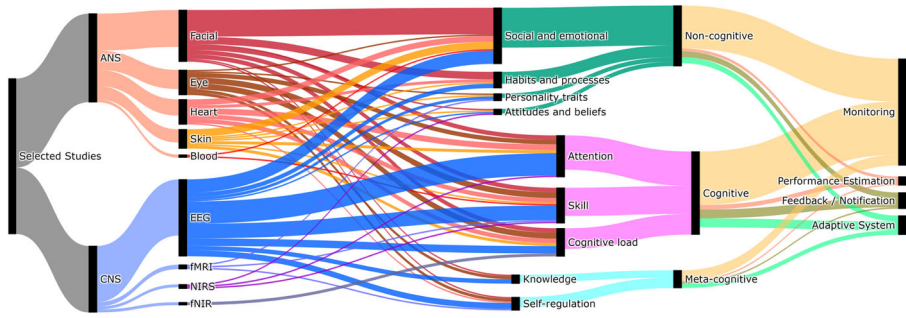
Table 10 shows the association between the type of neuro measurement instrument and the purposes of the selected studies. Numbers in brackets are for the actual reported outcomes of collecting the neurophysiological data in these studies. The number of studies that merely reported monitoring without feedback or interventions is considerably more than those who had this intention. This increase suggests that the majority of the selected studies are in an early stage of examining the reliability of applied instruments and the validity of captured constructs. In contrast, performance estimation has the least number of the actual reported outcomes. The most frequent neuro measurement instruments for the intention of monitoring, estimating performance, and providing feedback/notification are EEG, Facial, and Eye, respectively. However, for developing adaptive systems, Facial has a higher adoption rate than EEG. In general, ANS-related measurement instruments were more commonly used or intended to be used for adaptive systems, which could have been the result of the more affordable cost of implementation and the low intrusiveness.

Discussion

Our SLR findings suggest that the use of neuro measurements in higher education has the potential to make meaningful contributions to teaching and learning in higher education, filling in the current gap of providing insights about the psychological constructs and mental states of a learner and how that can be used to enhance learning outcomes and educational technology design. In order to provide an interconnected understanding of neuro measurements, constructs and outcomes, we represent in Fig. 6, a Sankey diagram that shows the flow of collecting neurophysiological data by ANS and CNS related neuro measurement instruments, capturing

**Table 10** Neuro measurement instruments and the purposes of the studies (numbers in brackets show the actual reported outcomes)

Purpose of the study- Intended and (Reported)	Neuro measurement instrument									
	ANS					CNS				
	Facial	Eye	Heart	Skin	Blood	EEG	fMRI	NIRS	fNIR	
Monitoring	10(18)	6(11)	2(7)	2(7)	1(2)	15(24)	1(1)	1(2)	0(1)	
Performance Estimation	4(1)	4(2)	3(0)	2(0)	0(0)	8(3)	1(1)	0(0)	0(0)	
Feedback / Notification	3(2)	2(1)	1(1)	0(0)	0(0)	9(8)	1(1)	1(0)	0(0)	
Adaptive System	8(4)	3(1)	5(3)	5(2)	1(0)	6(3)	0(0)	0(0)	2(1)	



**Fig. 6** Sankey diagram of findings from our SLR

different psychological constructs of learners for monitoring, performance estimation, providing feedback, and developing adaptive systems in higher education. This figure summarises much of the findings on measurements, constructs and outcomes.

In terms of neuro measurements, as illustrated, there seems to have been a near close balance between the use of ANS and CNS related measurements. EEG is the only neuro measurement instrument that has a significant contribution in capturing all sub-categories of the constructs. This is followed by Facial, which also plays a considerable role in identifying different types of constructs. This trend seems likely to follow in the future as advances in facial recognition, powered by machine learning algorithms, provide the ability for running larger, less intrusive studies at a lower cost. From the measurements point of view, each neuro measurement instrument is more associated with detecting some specific psychological constructs as follows: Facial with the social and emotional qualities; Eye with the cognitive load and skills; Heart with the attention and emotion; Skin with the emotion; EEG with the attention, skill, and emotion; fNIR with the cognitive load. For some neuro measurement instruments with the limited number of studies (e.g., Blood, fMRI, NIRS), there is no significant focus on a specific construct where each study focuses on one construct as for: 2 Blood applied studies: 1 on skill and 1 emotion; 3 fMRI: 1 on skill, 1 on personality, and 1 on self-regulation; 2 NIRS: 1 on the attention, and 1 on skill and attitude. Among the studies in our SLR, there is no report of using hormone-based measurements to detect learners' internal conditions in an educational setting. Measuring hormones through saliva seems to be a common approach in other domains like psychiatry, medicine, clinical and basic research (Gröschl 2008). However, our findings suggest that this method unsurprisingly does not seem to lend itself well to the educational setting, and it has not been used in any of the studies in our SLR. As suggested by a recent survey by Fischer et al. (2019), there seems to be a growing research interest in NeuroIS community towards using saliva measurement due to less required cost and effort in comparison with the CNS-related measurements. However, implementation of a hormone-based measurement in a learning environment is still a challenging task due to being intrusive and inconvenient for learners as well as not being suitable for real-time scenarios.

From the psychological constructs point of view, the associations of some sub-categories of the constructs with some specific neuro measurement instruments are

stronger. For example, the cognitive constructs (e.g., attention, cognitive load, and skill) are investigated in more studies using EEG. However, Eye-based measurements and Heart-based measurements have also contributed considerably in detecting the cognitive constructs. The highest portion of the studies is dedicated to approximating cognitive constructs with all its sub-categories receiving a fair share of the focus. Noticeably, CNS-related measurements have generally been used the most for capturing cognitive constructs. The second-highest portion of studies is dedicated to approximating non-cognitive constructs, where the social and emotional sub-category has received the most focus. Noticeably, ANS-related measurements have generally been used the most for capturing non-cognitive constructs. Meta-cognitive constructs attract the least number of research attentions, which seems to demonstrate a gap that can be focused on in future work. Finally, the figure indicates that the majority of the papers reported just monitoring the state of the learner and that there is a reasonable presence from all three types of constructs across all four types of outcomes.

In terms of experimental settings, designing and conducting reliable experiments that use neuro measurements in education seems to be quite challenging. For the design, there have been fiery debates about the opportunities and challenges of using true-experiments using randomised controlled trials in education (Sullivan 2011; Sung et al. 2005). While true-experiments remain a gold-standard test for establishing causality in many domains, they are often subject to threats to ethical considerations and fairness in providing equal opportunities for all students. Consequently, many of the selected papers have used non-true experiments in their studies. A potential method to mitigate the ethical challenges of using true-experiments in education, as utilised by many of the SLR's studies, is to conduct quasi-experimental studies where students self-select whether or not to engage with an intervention. Quasi-experiments are often subject to threats to internal validity: self-selected engagement with an intervention might be influenced by specific traits or needs, meaning that students in the control group are not comparable to those in the experimental group at baseline. A potential solution to reducing self-selection bias in a quasi-experimental study is to use Propensity Score Matching (PSM) (Rosenbaum and Rubin 1983). This method matches each participant in the experiment group with a similar participant from the control group based on a set of covariates. For an example of conducting an educational quasi-experimental study with PSM, please see the work of Khosravi et al. (2019).

In terms of conducting reliable experiments that use neuro measurements in education, there are three main types of challenges: data collection, number of participants, and response bias.

- *Data collection.* Many articles report challenges in data collection where they had to remove a significant portion of their collected data due to some technical issues, ethical issues such as lack of consent, or participants' discomfort, which led to moving or removing the measurement device. For example, \*Muldner and Burleson (2015) report that they had to remove data from two participants on the skin measurement due to dropped connectivity of the skin conductance sensor. They also report a similar problem with the EEG sensor resulting in loss of

- data on four participants. Similarly, \*Harrison et al. (2014) had to remove some results due to poor quality of eye-tracker data because some participants moved the device due to discomfort.
- *Number of participants.* Given the high cost and the high level of intrusiveness associated with neuro measurement instruments, many of the experiments had a relatively low number of participants in their studies. This limitation of participants' size may have an impact on the reliability and generalisability of the findings of many of these studies. For example, \*Landowska and Miler (2016) recognise the sample size of their study (10 users) as a threat to the validity due to deficiency in dealing with usability issues. Many of the studies had initially recruited an adequate number of participants, but only conducted and reported empirical outcomes that used neuro measurements on a small fraction of their participants. For example, \*Wang and Hsu (2014) had initially recruited 189 participants, but of whom only 20 participants were hired in their experiment that used EEG.
  - *Response bias.* As discussed in "[Introduction](#)", the use of subjective measures for capturing psychological constructs raises cognitive biases and internal validity concerns. While the use of neuro measurements has been recognised as a powerful alternative that can reduce these concerns, response bias challenges still need to be considered when using neuro measurement instruments. One of the reported challenges is due to the distraction that the measurement devices themselves cause on the participants' normal behaviour in the learning environment. This bias in response has been explained in two ways. First, some studies claim the undesirable results of their experiments can be a result of the intrusiveness of the device. For example, \*Edwards et al. (2017) state that conducting experiments using Eye and Skin requires participants to remain still, which impacts natural behaviour and leads to more stoic and inexpressive faces. In another study, \*Siqueira et al. (2017) used automatic arm blood pressure monitor and wristband heart rate monitor in their experiments to investigate the association of the air temperature changes in a learning environment and cognitive performance and comfort of students. While students may feel comfortable with the wristband heart rate monitor, which is very similar to a smartwatch, the automatic arm blood pressure monitor may be distracting and discomforting them during the process in addition to the thermal discomfort. The second reason for bias in response can be because of the Hawthorne effect which relates to the different behaviours of the participants while being watched in a controlled environment compared to their normal acts in a natural setting (\*Landowska and Miler 2016). For example, \*Seugnet Blignaut and Matthew (2017) suspect the emotional state influences on their participants could be a result of the experimental procedures such as being blocked in behind a strange machine, accompanied by an examiner, and wearing new experimental apparatus.

While much has been achieved in successfully developing neuro measurement tools, they have only been slowly embraced by higher education, with adoption mostly restricted to research projects. A few independent factors may be contributing to the current low adoption of neuro measurement tools. Xie et al. (2019) highlight

that using wearable devices in learning environments is not common due to the lack of up-to-date IT skills and knowledge. Henrie et al. (2015) point two main issues of using neurophysiological technologies as cost and complexity. In general, technology acceptance, defined as willingness towards using the technology designed to support tasks (Teo 2011), is one of the ongoing challenges in implementing neuro measurements tools in the learning environments. Limited investigations of technology acceptance models are reported in the education compared to other disciplines (e.g., business) due to more autonomy of the educational users, especially teachers (Teo 2011). Scherer et al. (2019) considered several variables that affect the technology acceptance model of teachers' adoption of digital technology in education such as perceived usefulness, perceived ease of use, attitudes toward technology, behavioural intention, and technology use. In this regard, dealing with the intrusiveness and the influence of technical support on students' satisfaction, as an important variable in technology acceptance model (Teo 2009), would promote the perceived ease of use whereas the ethical consideration and privacy would enhance the attitudes toward neuro measurements technology.

Several shreds of evidence suggest that compared to some other fields such as neuroscience and clinical domains, the use of neuro measurements in higher education is still in the early stages of development. Firstly, in terms of the data collection and analysis, educational studies often utilise simple commercial tools, with minimal reliability that tradeoff data quality for lowering the cost of the device or convenience of students. For example, a single-channel EEG sensor has been applied in a considerable amount of studies to detect the level of attention, which compromises the quality of the data compared to more advanced multi-channel EEG devices that are used in clinical domains (Alotaiby et al. 2015). Second, in terms of the reproducibility of the conducted experiments, there are only a few studies in the field of education with either available code or accessible dataset. This lack of availability could be a result of the ethical considerations and privacy issues. Nevertheless, reproducibility is an important aspect for developing benchmarks of public data sets and reliable results (Munafò et al. 2017). In comparison, there are many openly accessible datasets (e.g., Banaee et al. 2013) and available algorithms (e.g., Rajeswari and Jagannath 2017) from clinical domains that enable conducting reproducible studies. Finally, the vast majority (71%) of the papers reported in the SLR solely monitor a learner's state during a learning task without providing feedback or adapting to a learner's need. In comparison, many tools used in the clinical setting support adaptive interventions (e.g., Hardeman et al. 2019; Wang and Miller 2020) or provide various types of biofeedback (e.g., Schoenberg and David 2014).

The maturity level of using neuro measurements in higher education seems to be comparable to its use in other subfields of information systems in that automated interventions are theorised more often than pursued. For example, in a systematic review on stress management interventions using psychophysiological components, De Witte et al. (2019) posited that the limited variety of interventions using physiological measures heretofore might be a result of practical issues such as delayed feedback associated with hardware or big data, and Fischer et al. (2019) appreciated the relatively high number of methodological papers as the necessity of becoming



more familiar with the neuroscience tools in the field of NeuroIS. There is, however, some work that can be mentioned. The recent review by Lux et al. (2018) provides examples of the development of live biofeedback systems from non-clinical application domains such as decision making and computer games.

## Conclusion

To the best of our knowledge, this is the first systematic literature review that focuses on the use of neuro measurements in higher education. We examined the literature based on three main themes of measurements, experimental design, as well as constructs and outcomes. An interactive visualisation tool, available at <http://neuro-in-higher-education-slr.herokuapp.com/>, has been developed to support readers in pursuing further investigations and dynamically drilling down into the reported findings. The review confirms that there is empirical evidence for the development of educational technologies that employ neurophysiological measurements to enhance teaching and learning practices. However, the review indicates that at this time, a number of challenges and concerns exist at both technical and empirical level, and hence the adoption of educational technologies augmented with neurophysiological measurement is limited and in its early stages of development. At the same time, the review provides evidence that the use of neurophysiological measurement in the context of educational technologies is on the rise. We hope that the findings of the SLR in terms of the type of measurements used within these systems, the experimental settings used in these studies and their outcomes and intentional uses can provide insights for researchers that are interested in the field or technologists that are involved in the implementation of educational tools and technologies.

## References

- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428.
- Aleven, V., McLaughlin, E.A., Glenn, R.A., Koedinger, K.R. (2016). Instruction based on adaptive learning technologies. *Handbook of Research on Learning and Instruction* 522–560.
- Ali, L., Hatala, M., Gašević D., Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1), 470–489.
- Alotaiby, T., Abd El-Samie, F.E., Alshebeili, S.A., Ahmad, I. (2015). A review of channel selection algorithms for eeg signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1), 66.
- Andreassi, J.L. (2010). *Psychophysiology: Human behavior and physiological response*. Psychology Press.
- Artifice, A., Sarraipa, J., Jardim-Goncalves, R., Guevara, J.C., Kadar, M. (2017). Ethical issues of a smart system to enhance students' attention. In *2017 International Conference on Engineering, Technology and Innovation, ICE/ITMC, IEEE*, pp. 1564–1571.
- Ashraf, H., Sodergren, M.H., Merali, N., Mylonas, G., Singh, H., Darzi, A. (2018). Eye-tracking technology in medical education: A systematic review. *Medical Teacher*, 40(1), 62–69.
- Banaee, H., Ahmed, M.U., Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, 13(12), 17472–17500.



- Barber, W., King, S., Buchanan, S. (2015). Problem based learning and authentic assessment in digital pedagogy: Embracing the role of collaborative communities. *Electronic Journal of E-Learning*, 13(2), 59–67.
- Barr, L. et al. (2009). An evaluation of emerging driver fatigue detection measures and technologies. Tech. rep., United States. Federal Motor Carrier Safety Administration.
- \*Batterink, L., & Neville, H. (2014). ERPs recorded during early second language exposure predict syntactic learning. *Journal of Cognitive Neuroscience*, 26(9), 2005–2020. <https://doi.org/10.1162/jocn.a.00618>.
- Beg, M.S. (2005). A subjective measure of web search quality. *Information Sciences*, 169(3-4), 365–381.
- Beldarrain, Y. (2006). Distance education trends: Integrating new technologies to foster student interaction and collaboration. *Distance education*, 27(2), 139–153.
- \*Bian, C., Zhang, Y., Wang, D., Liang, Y., Wu, B., Lu, W. (2018). An academic emotion database and the baseline evaluation. In *2018 13th International Conference on Computer Science & Education, ICCSE, IEEE*, pp. 1–6.
- \*Bin Abdul Rashid, N., bin Taib, M.N., bin Lias, S., bin Sulaiman, N., et al. (2015). Appraisal of eeg beta summative power towards learning style classification. In *2015 2nd International Conference on Biomedical Engineering, ICoBE, IEEE* (pp. 1–5).
- \*Blanchard, N., Bixler, R., Joyce, T., D'Mello, S. (2014). Automated physiological-based detection of mind wandering during learning. In Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (Eds.) *Intelligent Tutoring Systems, Springer International Publishing, Lecture Notes in Computer Science*, pp. 55–60.
- Bogarín, A., Cerezo, R., Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1), e1230.
- Boud, D., Cohen, R., Sampson, J. (2014). Peer learning in higher education, Learning from and with each other. Routledge.
- \*Bridge, D., Cohen, N., Voss, J. (2017). Distinct hippocampal versus frontoparietal network contributions to retrieval and memory-guided exploration. *Journal of Cognitive Neuroscience*, 29(8), 1324–1338. <https://doi.org/10.1162/jocn.a.01143>.
- Brocke, J.V., Riedl, R., Léger, P.M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems*, 53(3), 1–13.
- \*Bueno-Palomeque, F., Lema-Condo, E., Castro-Villalobos, S., Serpa-Andrade, L., Ordoñez-Morales, E. (2018). Variation of the EEG-energy in a second language class. In *NEUROTECHNIX 2018- Proceedings of the 6th International Congress on Neurotechnology, Electronics and Informatics*, pp 41–45.
- Byong-Hoon, J. (2008). Preventive terminal device and internet system from drowsy and distracted driving on motorways using facial recognition technology. *US Patent Application*, 12(021), 120.
- Campbell, D.T., & Stanley, J.C. (2015). Experimental and quasi-experimental designs for research. Ravenio Books.
- \*Chen, C.M., & Wu, C.H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, 80, 108–121. <https://doi.org/10.1016/j.compedu.2014.08.015>.
- \*Chen, C.M., Wang, J.Y., Yu, C.M. (2017). Assessing the Attention Levels of Students by Using a Novel Attention Aware System Based on Brainwave Signals. *British Journal of Educational Technology*, 48(2), 348–369. <https://doi.org/10.1111/bjet.12359>.
- Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S.Z., Khawaji, A., Conway, D. (2016). *Robust multimodal cognitive load measurement*. Cham: Springer International Publishing.
- \*Chen, J., Luo, N., Liu, Y., Liu, L., Zhang, K., Kolodziej, J. (2016). A hybrid intelligence-aided approach to affect-sensitive e-learning. *Computing*, 98(1-2), 215–233. <https://doi.org/10.1007/s00607-014-0430-9>.
- Chen, O.T.C., Chen, P.C., Tsai, Y.T. (2017). Attention estimation system via smart glasses. In *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB, IEEE*, pp. 1–5.
- Cruz-Cunha, M.M. (2016). Encyclopedia of E-health and Telemedicine. IGI global.
- \*Dan, A., & Reiner, M. (2018). Reduced mental load in learning a motor visual task with virtual 3d method. *Journal of Computer Assisted Learning Oxford*, 34(1), 84–93. <https://doi.org/10.1111/jcal.12216>.

- Darvishi, A., Khosravi, H., Sadiq, S. (2020). Utilising learnersourcing to inform design loop adaptivity. In *European Conference on Technology Enhanced Learning*, Springer, pp. 332–346.
- De Witte, N.A., Buyck, I., Van Daele, T. (2019). Combining biofeedback with stress management interventions: A systematic review of physiological and psychological effects. *Applied Psychophysiology and Biofeedback*, 44(2), 71–82.
- \*Derosière, G., Dalhoumi, S., Perrey, S., Dray, G., Ward, T. (2014). Towards a near infrared spectroscopy-based estimation of operator attentional state. *PLoS ONE*, 9(3). <https://doi.org/10.1371/journal.pone.0092045>.
- Dicheva, D., Dichev, C., Agre, G., Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, 18(3), 75–88.
- \*Dimililer, K. (2018). Use of Intelligent Student Mood Classification System (ISMCS) to achieve high quality in education. *Quality & Quantity*, 52, 651–662. <https://doi.org/10.1007/s11135-017-0644-y>.
- Dunn, T.J., & Kennedy, M. (2019). Technology enhanced learning in higher education; motivations, engagement and academic achievement. *Computers & Education*, 137, 104–113.
- \*Durall, E., & Leinonen, T. (2015). Feeler: Supporting awareness and reflection about learning through EEG data. *Awareness and Reflection in Technology-Enhanced Learning*, 1465, 67–73.
- \*Edwards, A., Massicci, A., Sridharan, S., Geigel, J., Wang, L., Bailey, R., Alm, C. (2017). Sensor-based methodological observations for studying online learning. In *Proceedings of the 2017 ACM Workshop on Intelligent Interfaces for Ubiquitous and Smart Learning*, pp 25–30. <https://doi.org/10.1145/3038535.3038536>.
- \*El-Abbasy, K., Angelopoulou, A., Towell, T. (2018). Measuring the engagement of the learner in a controlled environment using three different biosensors. In *10th International Conference on Computer Supported Education*, (Vol. 2 pp. 278–284).
- \*Enegi, I., Hamada, M., Adeshina, S. (2018). Adaptive multimedia learning framework with facial recognition system. In *2017 13th International Conference on Electronics, Computer and Computation (ICECCO)*, vol 2018-January, pp. 1–6. <https://doi.org/10.1109/ICECCO.2017.8333315>.
- Fischer, T., Davis, F.D., Riedl, R. (2019). Neurois: A survey on the status of the field. In *Information Systems and Neuroscience*, Springer, pp. 1–10.
- Fleureau, J., Guillotel, P., Huynh-Thu, Q. (2012). Physiological-based affect event detector for entertainment video applications. *IEEE Transactions on Affective Computing*, 3(3), 379–385.
- Fluck, A.E. (2019). An international review of eExam technologies and impact. *Computers & Education*, 132, 1–15.
- Fredricks, J.A., Blumenfeld, P.C., Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Fried, E.I. (2017). What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review*, 11(2), 130–134.
- Furtado, H., & Trobec, R. (2011). Applications of wireless sensors in medicine. In *2011 Proceedings of the 34th International Convention MIPRO, IEEE*, pp 257–261.
- Gašević, D., Dawson, S., Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71.
- Garcia, R., Falkner, K., Vivian, R. (2018). Systematic literature review: Self-Regulated Learning strategies using e-learning tools for computer science. *Computers & Education*, 123, 150–163.
- \*Gazdi, L., Pomazi, K., Radostyan, B., Szabo, M., Szegletes, L., Forstner, B. (2017). Experimenting with classifiers in biofeedback-based mental effort measurement. In *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 331–335. <https://doi.org/10.1109/CogInfoCom.2016.7804571>.
- \*Gershman, S., Zhou, J., Kommers, C. (2017). Imaginative reinforcement learning: Computational principles and neural mechanisms. *Journal of Cognitive Neuroscience*, 29(12), 2103–2113. <https://doi.org/10.1162/jocn.a.01170>.
- \*Grafsgaard, J.F., Wiggins, J.B., Vail, A.K., Boyer, K.E., Wiebe, E.N., Lester, J.C. (2014). The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring. In *Proceedings of the 16th International Conference on Multimodal Interaction, ACM, New York, NY, USA, ICM '14*, pp 42–49. <https://doi.org/10.1145/2663204.2663264>.
- Greene, B.A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist*, 50(1), 14–30.
- Gröschl, M. (2008). Current status of salivary hormone analysis. *Clinical Chemistry*, 54(11), 1759–1769.

- Hardeman, W., Houghton, J., Lane, K., Jones, A., Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (jitais) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 31.
- \*Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P., Bunce, S., Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Transactions on Human-Machine Systems*, 44(4), 429–440. <https://doi.org/10.1109/THMS.2014.2319822>.
- Hawkins, J., & Collins, A. (1992). Design-experiments for infusing technology into learning. *Educational Technology*, 32(9), 63–67.
- Henrie, C.R., Halverson, L.R., Graham, C.R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, 90, 36–53.
- Higgins, J.P. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.1(0).
- Hill, J.R., & Hannafin, M.J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development*, 49(3), 37–52.
- Hofkens, T.L., & Ruzek, E. (2019). Measuring student engagement to inform effective interventions in schools. In *Handbook of Student Engagement Interventions*, Elsevier, pp 309–324.
- \*Hu, P.C., & Kuo, P.C. (2017). Adaptive learning system for E-learning based on EEG brain signals. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, vol 2017-January, pp 1–2, <https://doi.org/10.1109/GCCE.2017.8229382>.
- \*Hubbard, R., Sipolins, A., Zhou, L. (2017). Enhancing learning through virtual reality and neurofeedback: A first step. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp 398–403. <https://doi.org/10.1145/3027385.3027390>.
- Jahedi, S., & Méndez, F. (2014). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior & Organization*, 98, 97–114.
- \*Kanimozhi, A., & Raj, V. (2017). An adaptive e-learning environment centred on learner's emotional behaviour. In *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, vol 2017-January, pp 1–5, <https://doi.org/10.1109/ICAMMAET.2017.8186752>.
- \*Katona, J., & Kovari, A. (2016). A brain-computer interface project applied in computer engineering. *IEEE Transactions on Education*, 59(4), 319–326. <https://doi.org/10.1109/TE.2016.2558163>.
- \*Kepinska, O., Pereda, E., Caspers, J., Schiller, N. (2017). Neural oscillatory mechanisms during novel grammar learning underlying language analytical abilities. *Brain and Language*, 175, 99–110. <https://doi.org/10.1016/j.bandl.2017.10.003>.
- Khosravi, H., Kitto, K., Williams, J.J. (2019). Ripple: A crowdsourced adaptive platform for recommendation of learning activities. arXiv:1910.05522.
- Kim, J., Campbell, A.S., de Ávila, B.E.F., Wang, J. (2019). Wearable biosensors for healthcare monitoring. *Nature biotechnology* 1.
- Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: what is 'enhanced' and how do we know? a critical literature review. *Learning, media and technology*, 39(1), 6–36.
- \*Kise, K. (2017). Deeply sensing learners for better assistance: Towards distribution of learning experiences. In *Positive Learning in the Age of Information: A Blessing or a Curse?*, Springer, pp 373–385, [https://doi.org/10.1007/978-3-658-19567-0\\_22](https://doi.org/10.1007/978-3-658-19567-0_22).
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Citeseer.
- \*Kong, Y., & Li, W. (2018). Research on recognition method of learning concentration based on face feature. In *2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, vol 2018-January, pp 334–338, <https://doi.org/10.1109/ICCIS.2017.8274797>.
- Kramer, A.F. (1991). Physiological metrics of mental workload: A review of recent progress. Multiple-task performance 279–328.
- \*Kublanov, V., Petrenko, A., Nabiullina, A. (2017). Correction of attention in a learning ability task with using non-invasive neurostimulation of peripheral nervous system. In *BIOSIGNALS*, (Vol. 4 pp. 269–275).

- \*Kuo, Y.C., Chu, H.C., Tsai, M.C. (2017). Effects of an integrated physiological signal-based attention-promoting and English listening system on students' learning performance and behavioral patterns. *Computers in Human Behavior*, 75, 218–227. <https://doi.org/10.1016/j.chb.2017.05.017>.
- \*Landowska, A., & Miler, J. (2016). Limitations of Emotion Recognition in Software User Experience Evaluation Context. In *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp 1631–1640. <https://doi.org/10.15439/2016F535>.
- \*Landowska, A., Brodny, G., Wrobel, M. (2017). Limitations of emotion recognition from facial expressions in e-learning context. In *CSEDU*, (Vol. 2 pp. 383–389).
- Lane, H.C., & D'Mello, S.K. (2019). Uses of physiological monitoring in intelligent learning environments: A review of research, evidence, and technologies. In *Mind, Brain and Technology*, Springer. pp 67–86.
- Latif, M.M.A. (2019). Eye-tracking in recent l2 learner process research: A review of areas, issues, and methodological approaches. *System*, 83, 25–35.
- Leggette, H.R., Rice, A.H., Carraway, C., Baker, M.A., Conner, N. (2018). Applying eye-tracking research in education and communication to agricultural education and communication: A review of literature. *Journal of Agricultural Education*, 59(2), 79–108.
- \*Lin, F.R., & Kao, C.M. (2018). Mental effort detection using EEG data in E-learning contexts. *Computers and Education*, 122, 63–79. <https://doi.org/10.1016/j.compedu.2018.03.020>.
- \*Lin, H.C., Wu, C.H., Hsueh, Y.P. (2014). The influence of using affective tutoring system in accounting remedial instruction on learning performance and usability. *Computers in Human Behavior*, 41, 514–522. <https://doi.org/10.1016/j.chb.2014.09.052>.
- Lipnevich, A.A., MacCann, C., Roberts, R.D. (2013). Assessing noncognitive constructs in education. A review of traditional and innovative approaches. Oxford handbook of child psychological assessment.
- \*Liu, C.J., Huang, C.F., Liu, M.C., Chien, Y.C., Lai, C.H., Huang, Y.M. (2015). Does gender influence emotions resulting from positive applause feedback in self-assessment testing? evidence from neuroscience. *Educational Technology & Society*, 18(1), 337–350.
- \*Liu, S., Chen, Y., Huang, H., Xiao, L., Hei, X. (2018a). Towards smart educational recommendations with reinforcement learning in classroom. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pp 1079–1084, <https://doi.org/10.1109/TALE.2018.8615217>.
- \*Liu, Y., Chen, J., Zhang, M., Rao, C. (2018b). Student engagement study based on multi-cue detection and recognition in an intelligent learning environment. *Multimedia Tools and Applications*, 77(21), 28749–28775.
- Lodge, J.M., & Harrison, W.J. (2019). Focus: Attention science: The role of attention in learning in the digital age. *The Yale Journal of Biology and Medicine*, 92(1), 21.
- Luo, L., Kiewra, K.A., Peteranetz, M.S., Flanigan, A.E. (2017). Using eye-tracking technology to understand how graphic organizers aid student learning. In *Eye-tracking Technology Applications in Educational Research*, IGI Global, pp 220–238.
- Lux, E., Adam, M.T., Dorner, V., Helming, S., Knierim, M.T., Weinhardt, C. (2018). Live biofeedback as a user interface design element: A review of the literature. *Communications of the Association for Information Systems*, 43(1), 18.
- MacKay, J.R. (2019). Show and 'tool': How lecture recording transforms staff and student perspectives on lectures in higher education. *Computers & Education*.
- \*Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., Coyne, J. (2016). Cognitive context detection using pupillary measurements. In *Next-Generation Analyst IV, International Society for Optics and Photonics*, vol 9851, p 98510Q. <https://doi.org/10.1117/12.2224176>.
- \*Manseras, R., Eugenio, F., Palaoag, T. (2018). Millennial filipino student engagement analyzer using facial feature classification. In *IOP Conference Series: Materials Science and Engineering*, vol 325, <https://doi.org/10.1088/1757-899X/325/1/012006>.
- Mavrikis, M., et al. (2016). A study on eye fixation patterns of students in higher education using an online learning system. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, pp 408–416.
- McLoughlin, C., & Luca, J. (2002). A learner-centred approach to developing team skills through web-based learning and assessment. *British Journal of Educational Technology*, 33(5), 571–582.

- \*Medina, R., Carpenter, D., Geigel, J., Bailey, R., Wang, L., Alm, C.O. (2018). Sensing behaviors of students in online vs. face-to-face lecturing contexts. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp 77–82. <https://doi.org/10.1109/PERCOMW.2018.8480398>.
- \*Menekse Dalveren, G., & Cagiltay, N. (2018). Insights from surgeons' eye-movement data in a virtual simulation surgical training environment: effect of experience level and hand conditions. *Behaviour and Information Technology*, 37(5), 517–537. <https://doi.org/10.1080/0144929X.2018.1460399>.
- Messick, S. (1979). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology*, 71(3), 281.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement. *Systematic Reviews*, 4(1), 1.
- Money, W.H., & Dean, B.P. (2019). Incorporating student population differences for effective online education: A content-based review and integrative model. *Computers & Education*, 138, 57–82.
- \*Monkarese, H., Bosch, N., Calvo, R.A., D'Mello, S.K. (2017). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, 8(1), 15–28. <https://doi.org/10.1109/TAFFC.2016.2515084>.
- \*Muldner, K., & Bureson, W. (2015). Utilizing sensor data to model students' creativity in a digital environment. *Computers in Human Behavior*, 42, 127–137. <https://doi.org/10.1016/j.chb.2013.10.060>.
- Munafò, M. R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Du Sert, N.P., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., Ioannidis, J.P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021.
- Nacke, L.E. (2011). Directions in physiological game evaluation and interaction. In *CHI 2011 BBI Workshop Proceedings, Citeseer*.
- Ng, B., & Ong, A.K.K. (2018). Neuroscience and digital learning environment in universities: what does current research tell us?. *Journal of the Scholarship of Teaching and Learning*, 18(3), 116–131.
- \*Nor, N., & Salleh, S. (2015). Correlation between precursor emotion and human stress by using EEG signals. *ARPN Journal of Engineering and Applied Sciences*, 10(23), 17881–17889.
- \*Nye, B., Karumbaiah, S., Tokel, S., Core, M., Stratou, G., Auerbach, D., Georgila, K. (2018). Engaging with the scenario: Affect and facial patterns from a scenario-based intelligent tutoring system Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10947 LNAI:352–366, [https://doi.org/10.1007/978-3-319-93843-1\\_26](https://doi.org/10.1007/978-3-319-93843-1_26).
- O'Meara, P., Munro, G., Williams, B., Cooper, S., Bogossian, F., Ross, L., Sparkes, L., Browning, M., McClounan, M. (2015). Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: A proof of concept paper. *International emergency nursing*, 23(2), 94–99.
- \*Özek, M.B. (2018). The effects of merging student emotion recognition with learning management systems on learners' motivation and academic achievements. *Computer Applications in Engineering Education*, 26(5), 1862–1872. <https://doi.org/10.1002/cae.22000>.
- Paas, F., Tuovinen, J.E., Tabbers, H., Gerven, V., W, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63–71.
- \*Pantazos, K., & Vatrappu, R. (2016). Enhancing the professional vision of teachers: A physiological study of teaching analytics dashboards of students' repertory grid exercises in business education. In *2016 49th Hawaii international conference on System sciences (HICSS)*, vol 2016-March, pp. 41–50. <https://doi.org/10.1109/HICSS.2016.14>.
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 50(1), 128–138.
- \*Peng, S., & Nagao, K. (2018). Automatic evaluation of presenters' discussion performance based on their heart rate. In *CSEDU*, (Vol. 2 pp. 27–34).
- \*Pham, P., & Wang, J. (2015). AttentiveLearner: Improving mobile MOOC learning via implicit heart rate tracking. In Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M.F. (Eds.) *Artificial Intelligence in Education*, (Vol. 9112 pp. 367–376). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-19773-9\\_37](https://doi.org/10.1007/978-3-319-19773-9_37).
- \*Pi, Z., & Hong, J. (2016). Learning process and learning outcomes of video podcasts including the instructor and PPT slides: a Chinese case. *Innovations in Education and Teaching International*, 53(2), 135–144. <https://doi.org/10.1080/14703297.2015.1060133>.
- Picard, R.W. (2000). Affective computing. MIT press.

- Pintrich, P.R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research*, 31(6), 459–470.
- \*Poulsen, A.T., Kamronn, S., Dmochowski, J., Parra, L.C., Hansen, L.K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports* 7. <https://doi.org/10.1038/srep43916>.
- \*Prat, C., Yamasaki, B., Peterson, E. (2018). Individual differences in resting-state brain rhythms uniquely predict second language learning rate and willingness to communicate in adults. *Journal of Cognitive Neuroscience*, 31(1), 78–94. <https://doi.org/10.1162/jocn.a.01337>.
- Prichard, C., & Atkins, A. (2016). Evaluating 12 readers' previewing strategies using eye tracking. *The Reading Matrix: An International Online Journal*, 16(2), 110–130.
- \*Prieto, L.P., Sharma, K., Kidzinski, L., Dillenbourg, P. (2018). Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. *IEEE Transactions on Learning Technologies*, 11(2), 216–229. <https://doi.org/10.1109/TLT.2017.2690687>.
- \*Qi, Z., Beach, S.D., Finn, A.S., Minas, J., Goetz, C., Chan, B., Gabrieli, J.D.E. (2017). Native-language N400 and P600 predict dissociable language-learning abilities in adults. *Neuropsychologia*, 98, 177–191. <https://doi.org/10.1016/j.neuropsychologia.2016.10.005>.
- \*Qu, X., Hall, M., Sun, Y., Sekuler, R., Hickey, T. (2018a). A personalized reading coach using wearable EEG sensors: A pilot study of brainwave learning analytics. In *CSEDU*, (Vol. 2 pp. 501–507).
- \*Qu, X., Sun, Y., Sekuler, R., Hickey, T. (2018b). EEG markers of STEM learning. In *2018 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9. <https://doi.org/10.1109/FIE.2018.8659031>.
- Rajeswari, J., & Jagannath, M. (2017). Advances in biomedical signal and image processing—a systematic review. *Informatics in Medicine Unlocked*, 8, 13–19.
- \*Ray, A., & Chakrabarti, A. (2016). Design and implementation of technology enabled affective learning using fusion of bio-physical and facial expression. *Journal of Educational Technology & Society; Palmerston North*, 19(4), 112–125.
- Riedl, R., & Léger, P.M. (2016). Fundamentals of NeuroIS. *Studies in Neuroscience, Psychology and Behavioral Economics*.
- Riedl, R., Fischer, T., Léger, P. M. (2017). A decade of NeuroIS research: status quo, challenges and future directions. Thirty Eighth International Conference on Information Systems, South Korea.
- Robins, A.V., Margulieux, L., Morrison, B.B. (2019). Cognitive sciences for computing education. *Learning Sciences Faculty Publications* 22.
- Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rubio, S., Díaz, E., Martín, J., Puente, J.M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-tlx, and workload profile methods. *Applied Psychology*, 53(1), 61–86.
- \*Rusák, Z., Van De Water, N., De Smit, B., Horváth, I., Van Der Vegte, W. (2016). Smart reading aid for detecting problems with reading fluency and comprehension. In *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol 1B-2016, <https://doi.org/10.1115/DETC2016-59130>.
- \*Samah, N., Sha' Ameri, A., Daliman, S., Safri, N., Qusai, S. (2018). Using electroencephalogram signals to determine differences in brain functional connectivity during game-based problem solving task. In *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, pp. 7–12. <https://doi.org/10.1109/ICBAPS.2018.8527422>.
- Saw, A.E., Main, L.C., Gastin, P.B. (2016). Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *British Journal of Sports Medicine*, 50(5), 281–291.
- \*Sawyer, R., Mudrick, N., Azevedo, R., Lester, J. (2018). Impact of learner-centered affective dynamics on metacognitive judgements and performance in advanced learning technologies. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10948 LNAI:312–316. [https://doi.org/10.1007/978-3-319-93846-2\\_58](https://doi.org/10.1007/978-3-319-93846-2_58).
- Scherer, R., Siddiq, F., Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13–35.
- Schoenberg, P.L., & David, A.S. (2014). Biofeedback for psychiatric disorders: a systematic review. *Applied Psychophysiology and Biofeedback*, 39(2), 109–135.



- \*Serrhini, M., & Dargham, A. (2017). Toward incorporating bio-signals in online education case of assessing student attention with BCI. *Advances in Intelligent Systems and Computing*, 520, 135–146. [https://doi.org/10.1007/978-3-319-46568-5\\_14](https://doi.org/10.1007/978-3-319-46568-5_14).
- \*Sethi, C., Dabas, H., Dua, C., Dalawat, M., Sethia, D. (2018). EEG-based attention feedback to improve focus in e-learning. In *Proceedings of the 2018 2Nd International Conference on Computer Science and Artificial Intelligence, ACM, New York, NY, USA, CSAI '18*, pp 321–326.
- \*Seugnet Blignaut, A., & Matthew, G. (2017). Part II: Survive with vuvu on the vaal electroencephalography results of a gameplay experience evaluation of a mobile serious game for statistics education. In *Proceedings of the 16th World Conference on Mobile and Contextual Learning*. <https://doi.org/10.1145/3136907.3136911>.
- \*Sezer, A., Inel, Y., Seçkin, A., Uluçınar, U. (2015). An investigation of university students' attention levels in real classroom settings with Neurosky's MindWave mobile (EEG) device 2. *Turkish Online Journal of Educational Technology* 2015:61–69.
- Sinatra, G.M., Heddy, B.C., Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. Taylor & Francis.
- \*Siqueira, J., Da Silva, L., Coutinho, A., Rodrigues, R. (2017). Analysis of air temperature changes on blood pressure and heart rate and performance of undergraduate students. *Work*, 57(1), 43–54. <https://doi.org/10.3233/WOR-172533>.
- So, H.J., & Brush, T.A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, 51(1), 318–336.
- Sperling, R.A., Howard, B.C., Staley, R., DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation*, 10(2), 117–139.
- \*Spüler, M., Krumpel, T., Walter, C., Scharinger, C., Rosenstiel, W., Gerjets, P. (2017). Brain-computer interfaces for educational applications. In *Informational Environments: Effects of use, Effective Designs*, Springer, pp 177–201. [https://doi.org/10.1007/978-3-319-64274-1\\_8](https://doi.org/10.1007/978-3-319-64274-1_8).
- \*Stuijzand, B.G., Van Der Schaaf, M.F., Kirschner, F.C., Ravesloot, C.J., Van Der Gijp, A., Vincken, K.L. (2016). Medical students' cognitive load in volumetric image interpretation: Insights from human-computer interaction and eye movements. *Computers in Human Behavior*, 62, 394–403.
- Sullivan, G.M. (2011). Getting off the “gold standard”: randomized controlled trials and education research. *Journal of graduate medical education*, 3(3), 285–289.
- Sung, Y.T., Chang, K.E., Chiou, S.K., Hou, H.T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers & Education*, 45(2), 187–202.
- Tempelaar, D.T., Rienties, B., Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167.
- Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education*, 52(2), 302–312.
- Teo, T. (2011). *Technology acceptance in education*. Berlin: Springer Science & Business Media.
- Thomas, P.A., Kern, D.E., Hughes, M.T., Chen, B.Y. (2016). Curriculum development for medical education: a six-step approach. JHU Press.
- \*Thompson, N., & McGill, T. (2017). Genetics with Jean: the design, development and evaluation of an affective tutoring system. *Educational Technology Research and Development*, 65(2), 279–299. <https://doi.org/10.1007/s11423-016-9470-5>.
- \*Tobita, R. (2017). Development of effective ESP course design integrating analysis with near-infrared spectroscopy. In *Proceedings of The 8th International Conference on Society and Information Technologies (ICSIT 2017)*, vol 2017-March, pp 183–186.
- \*Vail, A.K., Grafsgaard, J.F., Boyer, K.E., Wiebe, E., Lester, J.C. (2016). Predicting learning from student affective response to tutor questions. In *International Conference on Intelligent Tutoring Systems*, pp 154–164. [https://doi.org/10.1007/978-3-319-39583-8\\_15](https://doi.org/10.1007/978-3-319-39583-8_15).
- \*Varga, N., & Bauer, P. (2017). Using event-related potentials to inform the neurocognitive processes underlying knowledge extension through memory integration. *Journal of Cognitive Neuroscience*, 29(11), 1932–1949. <https://doi.org/10.1162/jocn.a.01168>.
- \*Wang, C.C., & Hsu, M.C. (2014). An exploratory study using inexpensive electroencephalography (EEG) to understand flow experience in computer-based instruction. *Information & Management*, 51(7), 912–923. <https://doi.org/10.1016/j.im.2014.05.010>.

- \*Wang, J.X., & Voss, J.L. (2014). Brain networks for exploration decisions utilizing distinct modeled information types during contextual learning. *Neuron*, 82(5), 1171–1182. <https://doi.org/10.1016/j.neuron.2014.04.028>.
- Wang, L., & Miller, L.C. (2020). Just-in-the-moment adaptive interventions (jitai): a meta-analytical review. *Health Communication*, 35(12), 1531–1544.
- \*Wei, Q., Sun, B., He, J., Yu, L. (2017). BNU-LSVED 2.0: Spontaneous multimodal student affect database with multi-dimensional labels. *Signal Processing: Image Communication*, 59, 168–181. <https://doi.org/10.1016/j.image.2017.08.012>.
- \*Whitehill, J., Serpell, Z., Lin, Y., Foster, A., Movellan, J.R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 5(1), 86–98. <https://doi.org/10.1109/TAFFC.2014.2316163>.
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering.
- \*Wong, A., Moss, J., Schunn, C. (2016). Tracking reading strategy utilisation through pupillometry. *Australasian Journal of Educational Technology*, 32(6), 45–57. <https://doi.org/10.14742/ajet.3096>.
- \*Wu, C.H. (2017). New technology for developing facial expression recognition in e-learning. In *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*, pp 1719–1722. <https://doi.org/10.1109/PICMET.2016.7806648>.
- \*Wu, C.H., Tzeng, Y.L., Huang, Y.M. (2014). Understanding the relationship between physiological signals and digital game-based learning outcome. *Journal of Computers in Education*, 1(1), 81–97.
- Wu, C.I. (2012). Hci and eye tracking technology for learning effect. *Procedia-Social and Behavioral Sciences*, 64, 626–632.
- Xie, H., Chu, H.C., Hwang, G.J., Wang, C.C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education* 103599.
- Yang, F.Y., Tsai, M.J., Chiou, G.L., Lee, S.W.Y., Chang, C.C., Chen, L.L. (2018). Instructional suggestions supporting science learning in digital environments based on a review of eye tracking studies. *Journal of Educational Technology & Society*, 21(2), 28–45.
- \*Yang, X., Lin, L., Cheng, P.Y., Yang, X., Ren, Y., Huang, Y.M. (2018). Examining creativity through a virtual reality support system. *Educational Technology Research and Development*, 66(5), 1231–1254. <https://doi.org/10.1007/s11423-018-9604-z>.
- \*Yuksel, B.F., Oleson, K.B., Harrison, L., Peck, E.M., Afergan, D., Chang, R., Jacob, R.J. (2016). Learn Piano with BACH: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, CHI '16, pp 5372–5384.
- \*Zhai, X., Fang, Q., Dong, Y., Wei, Z., Yuan, J., Cacciollatti, L., Yang, Y. (2018). The effects of biofeedback-based stimulated recall on self-regulated online learning: A gender and cognitive taxonomy perspective. *Journal of Computer Assisted Learning; Oxford*, 34(6), 775–786. <https://doi.org/10.1111/jcal.12284>.
- \*Zhang, C., & Shen, R. (2017). Automatic detection of mind wandering using mobile device in the scenario of online and mobile education. In *Computer Science and Technology, WORLD SCIENTIFIC, Shenzhen, China*, pp 153–162. [https://doi.org/10.1142/9789813146426\\_0018](https://doi.org/10.1142/9789813146426_0018).
- \*Zhang, Z. (2018). A cognitive study of college students' english vocabulary based on electroencephalogram. *NeuroQuantology*, 16(5), 363–368. <https://doi.org/10.14704/nq.2018.16.5.1299>.
- Zhou, M., & Zhang, X. (2019). Online social networking and subjective well-being: Mediating effects of envy and fatigue. *Computers & Education* 103598.
- \*Zlokazov, K., Voroshilova, M., Pirozhkova, I., Lapenok, M. (2017). Eye tracking technology for assessment of electronic hybrid text perception by students. *Smart Innovation, Systems and Technologies* 75:245–252, [https://doi.org/10.1007/978-3-319-59451-4\\_24](https://doi.org/10.1007/978-3-319-59451-4_24).



## Affiliations

Ali Darvishi<sup>1</sup> · Hassan Khosravi<sup>1</sup>  · Shazia Sadiq<sup>1</sup> · Barbara Weber<sup>2,3</sup>

Hassan Khosravi  
h.khosravi@uq.edu.au

Shazia Sadiq  
shazia@itee.uq.edu.au

Barbara Weber  
barbara.weber@unisg.ch

- <sup>1</sup> School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Brisbane, QLD, 4072, Australia
- <sup>2</sup> Institute of Computer Science, University of St. Gallen, 9000, St. Gallen, Switzerland
- <sup>3</sup> Software and Process Engineering, Technical University of Denmark, 2800, Kgs. Lyngby, Denmark