

BCI-based Intelligent Tutoring Systems: A Literature Review and Meta-Analysis on Neuroadaptive ITS

J. Petit¹, M. J-M. Macé¹, A. Lécuyer², L. Pillette¹

¹Univ. Rennes, Inria, CNRS, IRISA, Rennes, France

²Inria, Univ. Rennes, IRISA, CNRS, Rennes, France

E-mail: jimmy.petit@irisa.fr

ABSTRACT: Intelligent tutoring systems (ITS) are artificial systems that provide personalised instruction, relying on student models to monitor cognitive states, such as mental workload. Recent studies have increasingly enhanced this monitoring through the use of passive brain-computer interfaces (BCI). In this review, we focused on ITS that use BCI metrics to adapt their interventions in real time. Following PRISMA 2020 guidelines we searched PubMed, Web of Science, Scopus, OpenAlex, and Semantic Scholar databases. We conducted a risk-of-bias assessment using RoB 2.0, and performed a meta-analysis to evaluate the impact of BCI-based ITS on learning outcomes. We found a moderate-to-large pooled effect size ($SMD = 0.79$) with a 95% confidence interval $[-0.08; 1.65]$ suggesting a potential benefit of BCI-based ITS. However, the wide confidence interval that includes zero indicates substantial uncertainty. As this field is still emerging, with most studies published after 2015, further research is needed to strengthen the evidence base and support more definitive conclusions.

INTRODUCTION

An intelligent tutoring system (ITS) is a computer-based instructional system that provides personalised guidance and feedback by adapting to a learner's needs, performance, and cognitive state, much like a human would in a learning context. It can provide abstract knowledge for formal education, e.g. mathematics [1] or biology [2], or applied skill sets for professional settings. For instance, the ITS called "CODES" provides a web-based environment for cooperative music prototyping [3]. An ITS consists of four basic components: *Domain model*, *Student model*, *Tutoring model* combined within a *User Interface* [4]. The *Domain model*, also known as *Expert model*, is a collection of concepts, rules, and problem-solving strategies of the domain to be learned. This *Domain model* can serve several purposes, for instance, it can be used as a standard to evaluate the trainee's performance or be used to detect errors. The *Student model* is a model that contains information on the trainee, such as their cognitive states and emotional states. It aims at monitoring these states during the step-by-step learning of the trainee. The model can also flag when the student model diverges from the domain model to provide a

correction. The *Student model* is often seen as the core model of the ITS. Finally, the *Tutor model* receives input from the two other models and provides decisions about tutoring strategies or interventions through the use of the *User interface*. To implement a *Student model*, an ITS designer often relies on monitoring in real-time the trainee's levels of mental workload or engagement [5, 6]. Various physiological sensors, including for instance heart rate variability, oculomotor and electrodermal activity, have been used to detect changes in these mental states [7, 8]. However, brain imaging techniques, such as electroencephalography (EEG), are considered to offer a more direct assessment of mental states in real-time [9]. As a result, passive BCI, hereafter simply referred to as BCI, have gained increasing attention within the ITS community over the years.

In a large meta-analysis comprising 107 effect sizes from 14,321 participants, Ma *et al.* [10] provided strong evidence that, in certain contexts, ITS can effectively complement and even substitute for other instructional modalities. Their findings indicated that ITS was associated with higher achievement compared with teacher-led large-group instruction, non-ITS computer-based instruction, and textbook- or workbook-based learning. Instructional formats that did not appear to benefit from ITS, however, were individualized human tutoring and small-group instruction.

Previous literature reviews at the intersection of neurophysiological assessment and learning have focused on different aspects. For instance, Xu *et al.* [11] have published a survey investigating EEG-based BCI for emotion recognition, such as happiness or sadness, in ITS. They focused on the accuracy of the system, however, its integration to actual ITS for inclusion in a *Student model* was not discussed. Another recent review of ITS focused on ADHD symptoms management through the use of serious-games [12]. The authors found that such ITS promote the acquisition of a wide range of cognitive and socioemotional skills among children with ADHD, including visuospatial working memory, attention and inhibition control, planning/organizing and problem solving. Overall, prior reviews have not examined how neurophysiological data have been incorporated into ITS nor whether such information yields additional benefits for learning outcomes.

In this literature review, we focus on studies of ITS-based learning that implement real-time adaptation driven by neurophysiological data reflecting the learner's state, such as engagement, attention, or mental workload. In particular, we examine the details of the adaptation strategies, signal processing methods and experimental design choices used to assess these systems. Through a meta-analysis, we seek to evaluate the effect of BCI-based adaptation in ITS on learning outcomes.

MATERIALS AND METHODS

Study selection In this literature review we included any ITS combined with a BCI. To that end, we incorporated ITS terminology from the classification of educational agents from Chou *et al.* [15] and main neurophysiological data acquisition methods. This resulted in the following query applied to title and abstract content, details of the exact query applied to each database are given in supplementary materials¹:

```
[ "Intelligent Tutoring"
OR "Adaptive Tutoring"
OR "Learning Companion*"
OR "Co-learning Companion*"
OR "Co-learning Agent*"
OR "Emotional Companion*"
OR "Educational Companion*"
OR "Pedagogical Companion*"
OR "Emotional Agent*"
OR "Educational Agent*"
OR "Pedagogical Agent*"
OR "Tangible Interface*"
OR "Tangible Interaction*" ]
AND
[ "Brain Activity"
OR "Electroencephalography" OR "EEG"
OR "Magnetoencephalography" OR "MEG"
OR "fNIRS" OR "fMRI"
OR "Neurofeedback" OR "NF"
OR "Brain-Computer Interface*" OR "BCI" ]
```

Following PRISMA 2020 guidelines, we conducted this search across PubMed, Web of Science, Scopus², OpenAlex, and Semantic Scholar² databases on 16/12/2025. The initial search yielded 325 references. After removing non-experimental studies and duplicates, 115 papers were screened, of which 8 studies met the inclusion criteria and were included in the analysis. Conference and journal articles were included if they: (i) involved both an ITS and measures of brain activity (ii) engaged participants in a learning task (iii) used indices derived from brain activity to adapt the ITS's behavior or pedagogical content. Studies were excluded if they: (i) were not published in English (ii) were proof-of-concept reports based on a single participant (iii) described experimental results

or ITS already presented in a more detailed article. In cases of duplicate reporting, only the most detailed and comprehensive publication was included. No restrictions were set with regards to study type or publication date. The details of the identification, screening, and study inclusion are presented in Figure 1. After inclusion in the review, we extracted the following information: protocol design, sample size, the names of the ITS, their embodiment, the adaptation strategy employed, the brain signal modalities and systems used, the number of available channels, the learning domain and outcome measures, as well as details of the signal-processing pipeline, including the types of temporal and spatial filters applied and the indices derived from brain activity.

Risk of Bias We assessed the risk of bias for all studies included in the literature review using the revised Cochrane Risk of Bias tool (RoB 2.0). When applicable, we applied either the version designed for randomised trials or the supplementary guidance for crossover trial designs. Our assessment focused on the per-protocol effect, examining adherence to the intervention as well as any deviations or implementation failures that could have influenced the reported outcomes.

Meta-analysis To be included in the meta-analysis, studies had to meet two additional criteria: (i) they reported a measurable learning outcome associated with the ITS, and (ii) they included both a control group and an intervention group. Among these studies, we distinguished two types of control groups based on how the control condition was defined. Passive control groups referred to studies comparing a BCI-based adaptive ITS (intervention) with either a static ITS that provided no additional adaptation or with no ITS at all (control). Active control groups, by contrast, compared the BCI-based adaptive ITS with an ITS delivering interventions similar in form to those of the adaptive system, but generated randomly or pseudo-randomly, and therefore unrelated with the neurophysiological data. This second type of control allows for a more balanced comparison by controlling for the amount of intervention and isolating the specific contribution of *adaptivity*. To account for this heterogeneity, the meta-analysis included control-group type as a moderator. This approach yields an overall effect size, subgroup-specific effects, and a test of between-group differences.

One of the included studies had a cross-over design and was analysed as a parallel-group trial [16]. This conservative approach, treating the two conditions as independent groups, ignores the pairing and therefore inflates the variance rather than the effect size, resulting in an under-weighted contribution of this study rather than an artificial gain in precision. Another study included multiple control groups that shared the same experimental group [17]. In accordance with the recommendations of the *Cochrane Handbook for Systematic Reviews* [18], we divided the shared group size equally among the comparisons to avoid double weighting and maintain the statistical independence of the effects.

¹https://gitlab.inria.fr/jipetit/passive_bci_and_its_review

²Accessed via Publish or Perish 8 software.

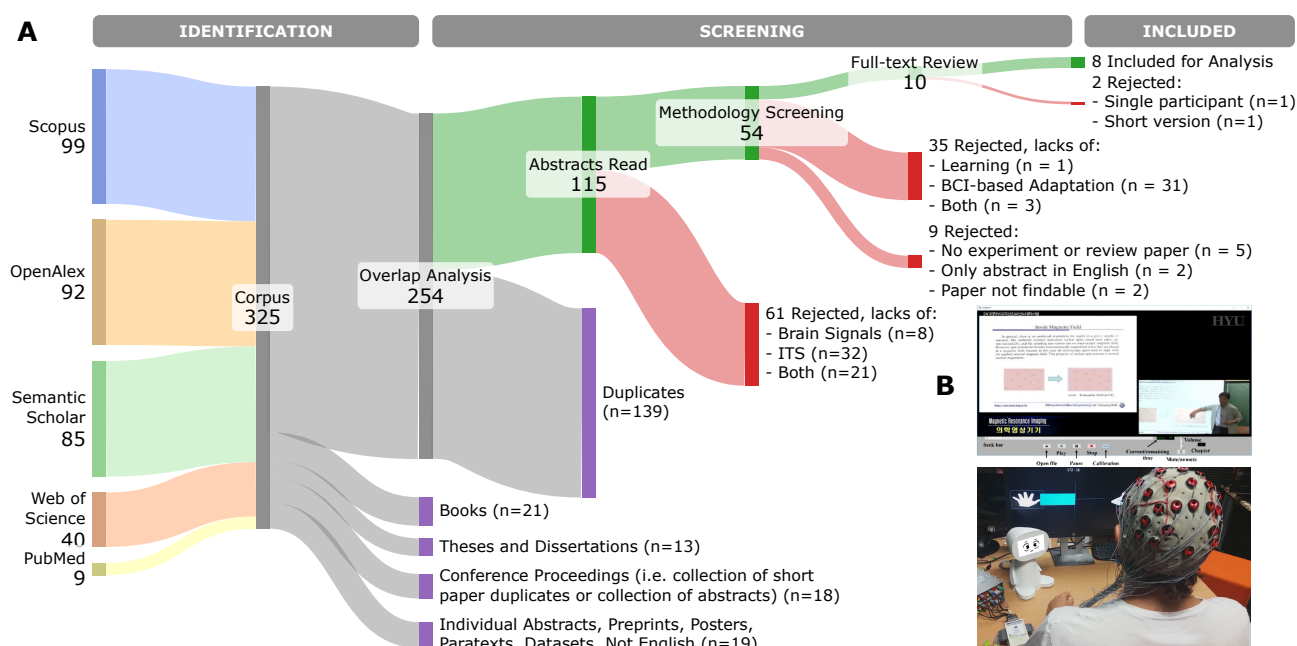


Figure 1: A. PRISMA diagram. B. Visual of two ITS: Top – User interface of the ITS in [13]. This ITS is not embodied. Down – ITS Peanut [14]. All figures have been reproduced with the appreciated authorisation of their respective authors. More visuals and details of the overlap analysis are presented as a Venn diagram in the supplementary material³.

To account for variability in assessment scales of the learning outcomes and variability in experimental conditions, we used the standardised mean difference (SMD) as the effect-size metric and conducted a meta-analysis using a random-effects model. Heterogeneity was assessed using I^2 and τ^2 . Given the small number of available effect sizes ($k = 5$ and $k = 3$ in subgroups), these estimates should be interpreted with caution.

RESULTS

Eight studies were included in the literature review, of which six met the inclusion criteria. At the same time, two did not because they lacked a control group (criterion (ii)) [19, 20]. The methodological characteristics of these studies are summarised below, with additional details provided in the supplementary material³.

ITS Characteristics ITS representations varied widely, ranging from humanoid robots and tangible companions [14, 16, 17, 20, 21] to static facial images animated through lip-sync software [19]. In [13] and [9], the authors did not embody the ITS with either a physical or virtual avatar. The former ITS is a video player that provides a lecture on MRI; the latter is an ITS called MENTOR designed to teach mathematical notation called reverse Polish notation through a series of pedagogical activities. Despite this diversity, their shared purpose is to support learning by adapting instructions to the learner's cognitive or emotional state. Most systems rely on an engagement index computed from the ratio $\beta/(\alpha + \theta)$ [9, 16, 17, 19, 21], sometimes complemented by proprietary metrics of stress [21]. In [13], the

authors developed an *ad-hoc* “attention index” in a preliminary study involving 32 participants. Candidate indices included $\beta/(\alpha + \theta)$, β/α , $(\beta + \theta)/\alpha$, and θ/α , computed for each available channel (Fp1, Fpz, Fp2, F3, Fz, F4, C3, Cz, and C4). Participants were instructed to rapidly identify a “d2” target (the letter *d* with two dots above or below it) among visually similar distractors (a “d” with only one dot, or a “p” replacing the “d”). High task performance was interpreted as reflecting a high attentional state. Based on these performances, the authors selected the fourth-highest candidate, θ/α at Cz, even though stronger candidates, such as $\beta/(\alpha + \theta)$ at C3, were identified, Cz was ultimately chosen due to manufacturing considerations. Additionally, moving average techniques were commonly applied to smooth the index over time and reduce false positive ITS adaptations caused by transient fluctuations. For example, [16] combined an exponentially weighted moving average with a 4-s moving average window, while [9] used a 40-s moving-average window. Only 2 studies did not use an engagement or attention index. In [20], the authors implemented four types of filters, each designed to highlight specific aspects of the user's brain activity on a humanoid tangible companion. For example, to emphasise vision-related activity, they applied a narrow band-pass filter centred on the α frequency band (8–12 Hz), followed by spatial filtering that retained channels located over and around the occipital cortex (P3, Pz, P4, PO3, POz, PO4, O1, Oz, and O2). The resulting α -band amplitude topomaps were then projected onto the tangible companion. In [14], the authors implemented a standard MI-based BCI pipeline consisting of temporal filtering (8–30 Hz band-pass) combined with spatial filtering using common spatial

³https://gitlab.inria.fr/jipetit/passive_bci_and_its_review

pattern filters. Band-power features were estimated via log-variance, and classification was performed using a shrinkage linear discriminant analysis.

All of these neurophysiological signals drive a variety of adaptation strategies: repeating or simplifying content when engagement drops [19, 21], inserting multimedia segments to enliven attention or provides supplementary information [13], modulating a robot's gestures or prosody [16, 17], or selecting between worked examples and problem-solving tasks [9]. Two systems stand apart from these cognitive-state-driven approaches. Frey *et al.* [20] implemented a tangible augmented-reality tutor that does not adapt instructional content but instead visualizes the learner's brain activity in real time through multiple EEG-based filters, aiming to foster conceptual understanding of brain processes. Similarly, Pillette *et al.* [14] adapted their system's behavior based on BCI performance during a mental-imagery tasks, providing emotional support and social presence through a rule-based mechanism rather than monitoring cognitive states. The learning domains addressed range from language learning [16, 19] and mathematics [9] to computer science [21], folk tale [17], mental imagery [14], and neuroscience [13, 20], with outcome measures including questionnaires [9, 13, 16, 17, 20, 21] or classification accuracy [14, 19].

All studies relied on EEG to measure brain activity, with an average of 14 channels, ranging from 1 [13, 17] to 32 [20]. Some studies did not specify the types of electrodes used, but based on the referenced hardware, we inferred that 6 out of 8 studies relied on wet electrodes [9, 13, 14, 19–21], while one study used dry electrodes [17]. Interestingly, [19] explicitly refers to dry channels despite using the Emotive Epoc+ headset, which is a saline-based system. In [16], the authors employed a Unicorn Hybrid Black system (g.tec medical engineering, Austria), which can use either wet or dry electrodes, but they did not specify which configuration was used.

Risk of Bias The methodological quality of the eight included studies is summarised in Table 1. Overall, evidence quality was moderate: three studies showed low risk of bias, two showed some concerns, and only two were rated as high risk. For the randomisation domain (D1), all studies were at most single-blind and generally provided limited information about their randomisation procedures. For example, [17] did not report group homogeneity tests despite visible baseline differences, and in [13], one passive control group was recorded before the treatment group for calibration purposes, preventing proper random allocation and resulting in “high risk” in D1 for both studies. In [14], the control group consisted of previously collected data, which also compromised randomisation. However, because the study aimed to evaluate the social presence of an ITS, reusing earlier data reduced the risk of experimenter and participant awareness to intervention, resulting in an overall assessment of “some concerns”. Carryover effects (DS) were unlikely in [16], as distinct pedagogical materials were

Table 1: Traffic light plot for Risk of Bias (RoB 2.0).

	D1	DS	D2	D3	D4	D5	Overall
[19]	N/A	N/A	N/A	+	!	!	!
[20]	N/A	N/A	N/A	+	!	!	!
[17]	−	N/A	+	+	+	!	−
[13]	−	N/A	+	+	+	!	−
[14]	!	N/A	+	+	+	!	!
[9]	+	N/A	+	+	+	!	+
[21]	+	N/A	+	+	+	!	+
[16]	+	+	+	+	+	!	+

Legend. D1: Randomisation process; DS: Bias arising from period and carryover effects (only in within-subject study); D2: Deviations from the intended interventions; D3: Missing outcome data; D4: Measurement of the outcome; D5: Selection of the reported result.

⊕ : Low risk; ⊕! : Some concern; ⊖ : High risk.

used across sequences. Deviations from intended interventions (D2) and missing outcome data (D3) were minimal, resulting in low risk for both domains. For the outcome measurement (D4), most studies relied on objective assessments of learning outcomes, including multiple choice questionnaires [9, 16, 21], open-ended and essay questions [13, 20], or system performance metrics [14]. In [17], the authors did not specify the type of questions used, other than noting that they assessed participants' ability to recall details of the story. In [19], no learning outcome assessment was provided, instead, the authors reported only the system's accuracy in recognising the engagement index. The absence of control groups in [19, 20] limited between-group comparisons, led to “some concerns”, while being of “low risk” for the other studies. Finally, selective reporting (D5) was generally transparent, but the lack of preregistered protocols resulted in an overall rating of “some concerns”.

Meta-analysis Six of the 8 studies included in the literature review met the inclusion criteria of the meta-analysis [9, 13, 14, 16, 17, 21]. Collectively, these studies provided 8 effect sizes because 2 studies included multiple control groups or treatment groups [13, 17]. Figure 2 presents the forest plot of the effect sizes measuring the learning outcomes from the adaptive ITS. Studies including a passive control group were the largest (treatment $n = 52$ and control $n = 65$), resulting in a pooled effect size SMD of 0.89 [−0.77; 2.54] (95%-CI) and a statistically significant heterogeneity ($I^2 = 77.2\%$ [44.8%; 90.6%], $\tau^2 = 1.2$ [0.2; 17.4], $p = 0.0015$). Three studies included an active control group (treatment $n = 28$ and control $n = 33$), resulting in an pooled effect size SMD of 0.77 [−1.24; 2.78] and a slightly smaller heterogeneity ($I^2 = 66\%$ [0%; 90.2%], $\tau^2 = 0.4$ [0.0; 24.5], $p = 0.053$). Overall, the pooled effect size SMD of 0.79 [−0.08; 1.65] showed a statistical trend in favour of neuroadaptive ITS ($t_7 = 2.15$, $p = 0.0689$) and a large statistically significant heterogeneity ($I^2 = 71.0\%$ [40.1%; 86.0%], $\tau^2 = 0.6$ [0.1; 5.4], $p = 0.0011$) calculated over a large set of participants (treatment $n = 80$

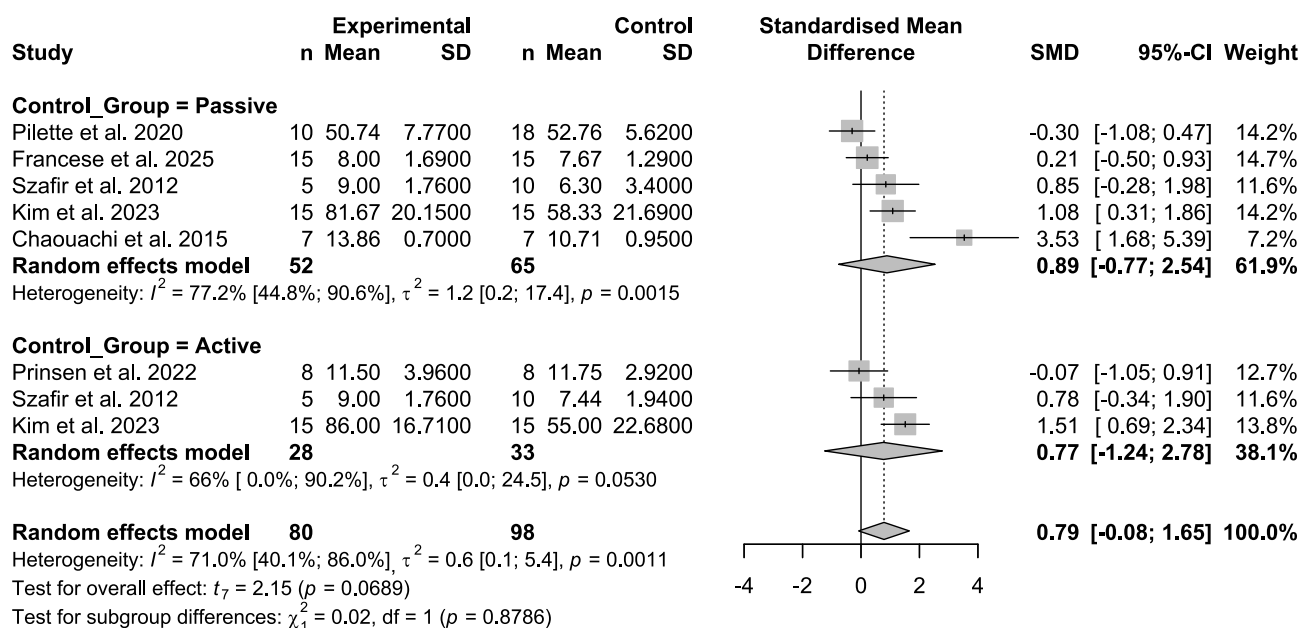


Figure 2: Forest plot of the meta-analysis comparing two subgroups of control groups versus passive BCI-based ITS.

and control $n = 98$). Subgroup differences between the two types of control group were not statistically significant ($\chi^2_1 = 0.02$, $p = 0.88$).

DISCUSSION

Our meta-analysis revealed a statistical trend toward a moderate-to-large positive effect ($SMD = 0.79$, 95% CI $[-0.08; 1.65]$, $p = 0.07$), suggesting a potential benefit of neuroadaptive ITS over standard ITS. However, wide confidence intervals crossing zero, indicate substantial remaining uncertainty. This uncertainty likely arises from several factors, including small sample sizes, heterogeneous study designs, variable outcome measures, and inconsistent signal processing pipelines. As the non-significant subgroup-difference test suggests, the type of control group alone does not account for this variability. Interestingly, studies using active control groups, i.e., where participants received a balanced amount of feedback, isolated more precisely the specific contribution of *adaptivity* within the ITS. These studies exhibited somewhat lower heterogeneity and therefore provide more conclusive empirical evidence in favour of BCI-based adaptive ITS. Greater standardisation in outcome measures, assessment scales, included electrodes, signal processing pipelines, and experimental protocols would help reduce these discrepancies in future work and improve comparability across studies. Finally, this research area remains in its early stages, with all but one study published after 2015. As additional studies accumulate, uncertainty surrounding the true effect size should decrease, allowing for more robust conclusions about the effectiveness of neuroadaptive ITS.

The RoB assessment indicated moderate evidence quality, mainly due to limited use of preregistered protocols and, more importantly, randomisation weaknesses. Unlike clinical research, where preregistration tools are

well-established, this field currently lacks identifiable infrastructure, making preregistration less common. However, as more studies are needed to draw firm conclusions on neuroadaptive ITS effectiveness, adopting more rigorous practices, particularly stronger randomisation procedures and, when feasible, double-blind designs, would produce more robust evidence.

EEG signals are particularly relevant to measure cognitive states. However, EEG-based monitoring of cognitive and affective states should be cautiously included in ITS. In [22], the authors show that EEG-based BCI remain limited for cognitive state monitoring due to low spatial resolution, strong sensitivity to ocular and muscular artifacts, and fluctuations in electrode-skin impedance that degrade signal quality. In turns, degraded signal quality may result in poorer and unstable cognitive state monitoring, like mental workload estimates. These factors should therefore be explicitly controlled, through appropriate preprocessing, artefact rejection, impedance monitoring, or predefined exclusion criteria, as they directly affect the reliability of the cognitive state monitoring. Also, EEG-based BCI suffers from pronounced inter- and intra-individual variabilities, combined with the non-stationarity of EEG, reduce model generalisation across users and sessions [22]. Many cognitive and affective states also share overlapping spectral signatures, constraining classifier specificity and increasing misclassification risks. In this context, the use of proprietary algorithms for which little or no methodological information is available should be avoided, as they prevent any assessment of validity, reproducibility, or bias. By contrast, mental workload is one of the most studied cognitive states in BCI research [23], with robust and reproducible EEG markers, most notably, increased frontal theta and decreased parietal alpha, making it a more reliable and promising target for future ITS applications. Finally, even though dry electrodes are improving, wet

electrodes should still be preferred, as they provide higher signal quality with fewer artefacts [24, 25]. Finally, in this review, we focused on ITS that adapt their pedagogical content or feedback based on brain activity. In future work, it would be relevant to compare the added value of these brain-signal input with other physiological measures, such as heart-rate variability, oculomotor activity, or electrodermal responses [7, 8], which have been used as reliable indicators of arousal, attention, and cognitive effort. Integrating these modalities, alongside EEG, could help disentangle overlapping cognitive states and improve the robustness of adaptive ITS.

CONCLUSION

BCI-based ITS show promising potential to enhance learning by adapting feedback and instructional behaviours to learners' cognitive states. Despite heterogeneous embodiment, learning domains, and signal processing methods, most systems relied on engagement or attention-related indices to guide real-time adaptation. Our meta-analysis suggests a moderate-to-large positive effect, though the wide confidence intervals and limited number of studies highlight substantial uncertainty. As the field is still emerging, with generally small samples and variable methodological reporting, more rigorous and standardised studies are needed to reach definitive conclusions.

Acknowledgements Work supported by the French National Research Agency (ANR-25-CE33-3211).

REFERENCES

- [1] A. Tato, R. N'Kambou, and R. Ghali, "Towards predicting attention and workload during math problem solving," in *Intelligent Tutoring Systems*, 2019.
- [2] W. Li, F. Wang, R. E. Mayer, and T. Liu, "Animated pedagogical agents enhance learning outcomes and brain activity during learning," *JCAL*, 2022.
- [3] E. M. Miletto, M. S. Pimenta, R. M. Vicari, and L. V. Flores, "Codes: A web-based environment for cooperative music prototyping," *Organised Sound*, 2005.
- [4] R. Nkambou, J. Bourdeau, and R. Mizoguchi, Eds., *Advances in Intelligent Tutoring Systems*, 2010.
- [5] R. H. Stevens, T. Galloway, and C. Berka, "EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills," in *User Modeling*, 2007.
- [6] A. Zammouri, A. A. Moussa, and S. Chevallier, "Use of cognitive load measurements to design a new architecture of intelligent learning systems," *Expert Syst. Appl.*, 2024.
- [7] K. F. Van Orden, W. Limbert, S. Makeig, and T. P. Jung, "Eye activity correlates of workload during a visuospatial memory task," *Human Factors*, 2001.
- [8] G. F. Wilson, "An analysis of mental workload in pilots during flight using multiple psychophysiological measures," *Int. J. Aerosp. Psychol.*, 2002.
- [9] M. Chaouachi, I. Jraid, and C. Frasson, "MENTOR: A physiologically controlled tutoring system," in *User Modeling, Adaptation and Personalization*, 2015.
- [10] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, "Intelligent tutoring systems and learning outcomes: A meta-analysis," *Journal of Educational Psychology*, 2014.
- [11] T. Xu, Y. Zhou, Z. Wang, and Y. Peng, "Learning emotions EEG-based recognition and brain activity: A survey study on BCI for intelligent tutoring system," *Procedia Computer Science*, 2018.
- [12] A. Doulou, P. Pergantis, A. Drigas, and C. Skianis, "Managing ADHD symptoms in children through the use of various technology-driven serious games: A systematic review," *Multimodal Technol. Interact.*, 2025.
- [13] H. Kim, Y. Chae, S. Kim, and C.-H. Im, "Development of a computer-aided education system inspired by face-to-face learning by incorporating EEG-based neurofeedback into online video lectures," *IEEE Trans. Learn. Technol.*, 2023.
- [14] L. Pillette, C. Jeunet, B. Mansencal, R. N'Kambou, B. N'Kaoua, and F. Lotte, "A physical learning companion for mental-imagery BCI user training," *IJHCS*, 2020.
- [15] C.-Y. Chou, T.-W. Chan, and C.-J. Lin, "Redefining the learning companion: The past, present, and future of educational agents," *Computers & Education*, 2003.
- [16] J. Prinsen, E. Pruss, A. Vrans, C. Ceccato, and M. Alimardani, "A passive brain-computer interface for monitoring engagement during robot-assisted language learning," in *IEEE SMC*, 2022.
- [17] D. Szafir and B. Mutlu, "Pay attention! designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012.
- [18] J. Higgins, S. Eldridge, and L. T., "Chapter 23: Including variants on randomized trials," *Cochrane Handbook for Systematic Reviews of Interventions - version 6.5*, 2019.
- [19] T. Xu, X. Wang, J. Wang, and Y. Zhou, "From textbook to teacher: An adaptive intelligent tutoring system based on BCI," in *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, 2021.
- [20] J. Frey, R. Gervais, S. Fleck, F. Lotte, and M. Hachet, "Teegi: Tangible EEG interface," in *27th ACM symposium on User interface software and technology*, 2014.
- [21] R. Francese, M. Buccino, F. Blaso, and L. De Santis, "Adapting educational content to student emotions: A user study with the furhat social robot," *IEEE TCSS*, 2025.
- [22] N. Gudikandula, R. Janapati, R. Sengupta, and S. Chintala, "Brain computer interface based emotion recognition with error analysis and challenges: An interdisciplinary review," *Discover Applied Sciences*, 2025.
- [23] J. Hassan, S. Reza, S. Ahmed, N. H. Anik, and M. Obaydullah, "EEG workload estimation and classification: A systematic review," *JNE*, 2025.
- [24] H. Hinrichs, M. Scholz, A. Baum, J. Kam, R. Knight, and H.-J. Heinze, "Comparison between a wireless dry electrode EEG system with a conventional wired wet electrode EEG system for clinical applications," *Scientific Reports*, 2020.
- [25] N. M. Ehrhardt, C. Niehoff, A.-C. Oßwald, D. Antonenko, G. Lucchese, and R. Fleischmann, "Comparison of dry and wet electroencephalography for the assessment of cognitive evoked potentials and sensor-level connectivity," *Frontiers in Neuroscience*, 2024.