Published in final edited form as:
Epileptic Disord. 2024 August ; 26(4): 444–459. doi:10.1002/epd2.20229.

A randomized controlled educational pilot trial of interictal epileptiform discharge identification for neurology residents

Fábio A. Nascimento, MD[*,1,2], **Jin Jing, PhD**[*,1,3], **Christopher Traner, MD**[4], **Wan Yee Kong**[3], **Marcia Olandoski**[5], **Srishti Kapur**[6], **Erik Duhaime, PhD**[6], **Roy Strowd**[7], **Jeremy Moeller**[4], **M. Brandon Westover, MD, PhD**[1,3]

[1]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.

[2]Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA.

[3]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA.

[4]Department of Neurology, Yale School of Medicine, New Haven, CT, USA.

[5]School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brazil.

[6]Centaur Labs, Boston, MA, USA.

[7]Department of Neurology, Wake Forest University School of Medicine, Winston-Salem, NC, USA.

Abstract

Objective: To assess the effectiveness of an educational program leveraging technology-enhanced learning and retrieval practice to teach trainees how to correctly identify interictal epileptiform discharges (IEDs).

Methods: This was a bi-institutional prospective randomized controlled educational trial involving junior neurology residents. The intervention consisted of three video tutorials focused on the six IFCN criteria for IED identification and rating 500 candidate IEDs with instant feedback either on a web browser (intervention 1) or an iOS app (intervention 2). The control group underwent no educational intervention (“inactive control”). All residents completed a survey and a test at the onset and offset of the study. Performance metrics were calculated for each participant.

Results: Twenty-one residents completed the study: control (n=8); intervention 1 (n=6); intervention 2 (n=7). All but two had no prior EEG experience. Intervention 1 residents improved from baseline (mean) in multiple metrics including AUC (0.74; 0.85; $p<0.05$), sensitivity (0.53;

Corresponding author: Dr. Fábio A. Nascimento, Campus Box 8111, 660 South Euclid Avenue, St. Louis, MO, 63110, USA, fabion@wustl.edu.

[*] Co-first authors

Disclosures: F. Nascimento is Associate Editor for Epileptic Disorders. Centaur Labs developed and have financial interest in the DiagnosUs app. C. Traner is a paid consultant for Ceribell, Inc. R. Strowd serves a consultant for Monteris Medical and Novocure. He receives stipend from the American Academy of Neurology. He has received research/grant support from the American Academy of Neurology, American Society for Clinical Oncology, Jazz Pharmaceuticals, and the International Association for Medical Science Educators. He receives royalties from Elsevier, Lecturio, and Kaplan Medical. J. Moeller receives royalties from Wolters Klouwer (UpToDate). M. B. Westover is a co-founder and holds equity in Beacon Biosignals, and receives royalties for authoring Pocket Neurology from Wolters Kluwer and Atlas of Intensive Care Quantitative EEG by Demos Medical. F. Nascimento, J. Jing, W. Y. Kong, and M. Olandoski report no disclosures relevant to this manuscript.

0.75; $p<0.05$), and level of confidence (LOC) in identifying IEDs/committing patients to therapy (1.33; 2.33; $p<0.05$). Intervention 2 residents improved in multiple metrics including AUC (0.81; 0.86; $p<0.05$) and LOC in identifying IEDs (2.00; 3.14; $p<0.05$) and spike-wave discharges (2.00; 3.14; $p<0.05$). Controls had no significant improvements in any measure.

Significance: This program led to significant subjective and objective improvements in IED identification. Rating candidate IEDs with instant feedback on a web browser (intervention 1) generated greater objective improvement in comparison to rating candidate IEDs on an iOS app (intervention 2). This program can complement trainee education concerning IED identification.

Keywords

EEG; education; interictal epileptiform discharges; epileptiform discharges; International Federation of Clinical Neurophysiology criteria to define interictal epileptiform discharges

Introduction

Adult and child neurology residents need to be able to accurately and reliably interpret EEGs by the time they complete residency training. Achieving this goal is especially important in countries where neurologists without fellowship training in clinical neurophysiology or epilepsy typically read EEGs - such as the United States (U.S.) and many European countries [1–3]. Nevertheless, real-world resident-centered evidence suggests that a large portion of neurology trainees – including those in their last year of residency training – do not achieve EEG competency by residency graduation [4–7].

In the U.S., neurology residency is guided by the adult and child neurology milestone project formulated by the Accreditation Council for Graduate Medical Education (ACGME) [8, 9]. This document suggests Level 4 as a graduation goal, which in the EEG subcompetency denotes gaining critical skills in EEG including being able to “recognize common EEG abnormalities”. Accurately and reliably identifying IEDs is arguably one of the most important abilities within this skill set. In fact, it may be potentially one of the most important overall abilities required to obtain EEG competency because correct IED identification is necessary to provide optimal care to people with seizures and epilepsy [10–17]. In the realm of educational efforts targeted at IED identification, recent evidence supports that delivering a teaching module focused on IED identification is effective and a promising option to fill this particular gap in EEG education [18].

In this study, we sought to further address this EEG education gap by creating a novel educational program whose learning objective was to teach adult and child neurology residents how to correctly identify IEDs on EEG. The program was developed based on the educational premise that achieving this competency requires mastery of (i) knowledge about IED identification and (ii) skill related to the application of this knowledge to actual interpretation of candidate IEDs on EEGs.

The necessary knowledge to correctly identify IEDs was introduced via the delivery of three pre-recorded, online video tutorials. The translation of new knowledge into the skill of correctly identifying IEDs on EEG was facilitated with the exercise of rating multiple candidate IEDs without and with instant feedback. The development of this content followed the principles of technology-enhanced learning, which covers all uses of digital technology (such as online video tutorials) [19, 20], and retrieval practice (aka test-enhanced learning) [21–23].

We, therefore, leveraged two major learning theories - technology-enhanced learning and retrieval practice, which were tailored to EEG education and ultimately combined into one single conceptual framework upon which our educational program was developed. Herein, we tested the hypothesis that this program is educationally effective and superior to an inactive control group where no EEG teaching was delivered.

Methodology

This study was a prospective randomized controlled educational trial investigating the effectiveness of a novel EEG teaching program focused on identification of IEDs. A flowchart of the trial is shown in Figure 1.

Participants

All junior adult and child neurology residents (i.e., residents in their preliminary year or first year of neurology training) at Massachusetts General Brigham (MGB) and Yale were invited to participate. The decision to participate was voluntary and did not impact residents' standing in their respective programs. Participants were not offered any financial incentive.

Pre-randomization: pre-survey and pre-spike test

Upon entering the study, all participants were asked to complete a pre-survey and pre-test ("pre-spike test"). The survey was a 5-point Likert scale perception survey whose primary goal was to assess the confidence of learners in identifying IEDs. The pre-spike test was developed based on our prior work [24] and consisted of 500 10-second EEGs where a candidate IED was marked by a red 0.5-second vertical rectangle. Participants were asked to rate candidate IEDs as epileptiform or non-epileptiform, and no feedback was given. Importantly, participants were able to switch montages (bipolar, common average, and physical reference) and modify sensitivity at their discretion.

EEGs used in the test were taken from a pool of 13,262 candidate IEDs, originating from a total of 1,051 EEGs, that were rated as epileptiform or non-epileptiform in standard visual analysis and in an independent fashion [24] by eight neurologists experts in EEG (hereafter referred to as EEG experts). Experts had a median experience in reading EEGs of 9.5 years (range 4 to 16 years); all had at least one year of fellowship training in clinical neurophysiology. Candidate IEDs were considered epileptiform if more than four experts classified these waveforms as epileptiform. Conversely, candidate IEDs were considered non-epileptiform if fewer than four experts classified them as epileptiform. Candidate IEDs that were classified as epileptiform by precisely four experts were considered "unknown" as to being epileptiform or non-epileptiform. Therefore, expert consensus was considered the gold standard. Candidate IEDs were further categorized into nine bins based on the degree of agreement among the eight experts ranging from zero out of eight votes to eight out of eight votes. Additionally, we added normal EEG variants to the pool of 13,262 EEGs, which were classified as non-epileptiform by three authors (FN, JJ, MBW), and placed in a new bin for a grand total of ten bins. In the spike test, EEGs/candidate IEDs were selected randomly but evenly across all ten bins mentioned above: 50 EEGs from each bin for a total of 500 EEGs in each spike test. Each spike test was, therefore, unique as it presented a different combination of candidate IEDs.

Randomization: control, intervention 1, and intervention 2 study groups

Participants who completed both pre-survey and pre-spike test were then randomized to one of the following groups: control, intervention 1, and intervention 2. The first did not have any educational intervention ("inactive control"), whereas the last two included three pre-recorded video tutorials and a spike test with instant feedback. The spike test with instant feedback was built as the pre-spike test (additional details are described in the section above) thus containing 500 10-second EEGs where a candidate IED was marked by a red vertical rectangle. As previously discussed, correct answers were based upon expert consensus [24].

In the intervention 1 group, participants received positive feedback (happy smiley face) if their answers agreed with expert consensus, and negative feedback (sad smiley face) otherwise. Importantly, participants received neutral feedback (neutral smiley face) for those candidate IEDs considered "unknown" (i.e., those classified as epileptiform by precisely four experts). In the intervention 2 group, feedback was similar except for its esthetics: positive feedback was shown as shading the correct answer green, negative feedback as shading the incorrect answer red, and neutral feedback as shading the answer purple and triggering the following message: "Thanks! Your opinion contributes to AI!".

The only difference between intervention groups was that the spike test with feedback was hosted on the AWS server and thus accessed through a web browser (intervention 1) or hosted on an iOS app named DiagnosUs thus accessed through a smartphone (intervention 2). The former allowed participants to switch montages (bipolar, common average, and physical reference) and modify sensitivity, whereas the latter did not offer these capabilities because candidate IEDs were shown as static EEG images, in bipolar montage.

Video tutorials were hosted on YouTube®, given by the authors (FN, JJ, MBW), and comprised a tutorial to the spike test (duration of 7 minutes and 39 seconds) (https://www.youtube.com/watch?v=3bUTMvQ-QdI), a video tutorial that introduced and explained the six operational criteria for defining an IED proposed by the IFCN [25] (duration of 15 minutes and 4 seconds) (https://www.youtube.com/watch?v=ZYviAGEaf18), and a video tutorial that applied the six IFCN criteria upon rating a set of 9 candidate IEDs from our prior work [24] and of varying levels of difficulty: 0 to 8/8 expert votes (duration of 19 minutes and 24 seconds) (https://www.youtube.com/watch?v=Z6A1OJ7_9cU). In these video tutorials, the authors (FN, JJ, MBW) instructed participants on how to analyze candidate IEDs using the six IFCN criteria [25] and to classify waveforms as epileptiform if they fulfill at least four of the six criteria.

Post-randomization: post-survey and post-spike test

Thirty days after randomization and as long as participants in either intervention group watched all three pre-recorded video tutorials and completed the spike test with instant feedback, they were asked to complete a post-survey and post-spike test. The post-survey was identical to the pre-survey, and the post-spike test was built as the pre-spike test (additional details are described in the sections above) thus containing 500 10-second EEGs where a candidate IED was marked by a red vertical rectangle.

Statistical analyses

Statistical analyses were performed using IBM (SPSS) Statistics Version 28.0 software and MATLAB R2020a. Non-parametric Wilcoxon test was used to compare metrics within study groups. Non-parametric Kruskal-Wallis test was used to preform comparisons between groups. Dunn’s test was used to perform post-hoc comparisons. Time between pre- and post-spike tests was analyzed using one-way ANOVA after logarithmic transformation of the data. Statistical significance was set at $p<0.05$. Bonferroni correction was performed for multiple pairwise comparisons.

Performance metrics including accuracy, true positive rate (TPR, aka sensitivity), false positive rate (FPR, aka 1-specificity), calibration index, area under receiver operating characteristic (AUC) curve, noise, and threshold/bias were computed to evaluate each participant’s performance in the pre- and post-spike tests. To investigate the factors underlying differences in performance between participants within and between groups, we created a latent trait model of the psychological process related to classifying candidate IEDs in a binary fashion (IED vs. non-IED) by each participant (Figure 2). Additional information concerning our methodology is summarized in Supplemental File 1.

Study protocol

The study was conducted in the first semester of the 2021–2022 academic year. Preparation of the data and public sharing of deidentified images of candidate IEDs on the spike test was conducted under an IRB-approved protocol. The study data was deidentified and obtained from adult and child neurology residents who volunteered to participate in the study. Therefore, the study did not require IRB approval based on our review of local IRB policies. Data and code to reproduce all results and figures are available through a public data sharing repository at https://github.com/bdsp-core/spike-test-pilot-trial.git.

Results

Forty-one adult and child neurology residents completed the pre-survey, 22 of whom also completed the pre-spike test. These 22 residents were subsequently randomized; all but one completed the entire study including the post-spike test and post-survey. Twenty-one residents were ultimately included in the analysis: control (n=8), intervention 1 (n=6), and intervention 2 (n=7). Participants’ data are summarized in Table 1 and Table 2.

Pre- and post-spike test scores (mean±SEM) related to accuracy, sensitivity, FPR, calibration, threshold/bias (θ), AUC, and noise (n) are summarized in Table 3. Post- and pre-spike test differences for each measure are plotted in Figure 3. Pre- and post-spike test individual scores related to threshold/bias (θ) and noise (n) are shown in Figure 4. Pre- and post-spike test ROC curves with individual operating points and calibration curves are shown in Figures 5–7. Pre- and post-survey level of confidence scores (mean; range 1–5) related to (i) identifying epileptiform discharges, (ii) identifying spike-wave discharges, and (iii) identifying epileptiform discharges and committing patient to an antiseizure drug are summarized in Table 5.

Control group (n=8)

There was no significant difference between post- and pre-spike test scores concerning any of the performance metrics. There was no statistically significant difference between post- and pre-survey level of confidence scores for any of the three variables assessed.

Intervention 1 group (n=6)

There were significant differences between post- and pre-spike test scores concerning accuracy (mean post-spike test = 0.76, mean pre-spike test = 0.67, mean difference = 0.08, p=0.028), sensitivity (0.75, 0.53, 0.22, p=0.028), calibration (3.3, −33, 36, p=0.046), threshold/bias (θ) (−0.01, 1.5, −1.5, p=0.046), AUC (0.85, 0.74, 0.11, p=0.028), and noise (n) (0.42, 1.0, −0.62, p=0.028); but not FPR (0.24, 0.21, 0.03, p=0.344). There were significant differences between post- and pre-survey level of confidence scores concerning participants' (iii) confidence in identifying epileptiform discharges and committing patient to an antiseizure drug (2.33–1.33=1.00; p=0.034).

Intervention 2 (n=7)

There were significant differences between post- and pre-spike test scores for accuracy (mean post-spike test = 0.77, mean pre-spike test = 0.72, mean difference = 0.05, p=0.028), AUC (0.86, 0.81, 0.06, p=0.028) and noise (n) (0.32, 0.65, −0.33, p=0.028); but not sensitivity (0.75, 0.71, 0.04, p=0.463), FPR (0.21, 0.27, −0.06, p=0.237), calibration (0.37, 3.8, −3.4, p=0.735), or threshold/bias (θ) (0.12, 0.04, 0.08, p=0.612). There were significant differences between post- and pre-survey level of confidence scores for (i) identifying epileptiform discharges (3.14–2.00=1.14; p=0.023) and (ii) identifying spike-wave discharges (3.14–2.00=1.14; p=0.011).

Inter-group comparisons

There was no statistically significant difference in time between completion of pre- and post-spike tests across the three groups (p=0.052) (Table 2). Concerning neurology residency institution, there were more MGB residents in the control and intervention 2 groups and fewer in the intervention 1 group. Pediatric residents were present only in the intervention 2 group. Control and intervention 1 groups had predominantly PGY2s and PGY1s, respectively; intervention 2 group had a more homogenous distribution. We were unable to statistically compare the frequencies of the three categorical values across the groups given the small sample size.

We compared pre-spike test scores (accuracy, sensitivity, FPR, calibration, threshold/bias (θ), AUC, and noise (n)) across the three groups. There were no statistically significant differences except for sensitivity (p=0.012). Pairwise comparison concerning sensitivity showed a statistically significant difference only in the comparison between control and intervention 2 (p=0.013). Moreover, we compared pre-survey level of confidence scores ((i) identifying epileptiform discharges, (ii) identifying spike-wave discharges, and (iii) identifying epileptiform discharges and committing patient to an antiseizure drug) across the three groups. There were no statistically significant differences.

In terms of pre-post changes in spike test scores across the three groups (Table 3), there were statistically significant differences in every measure but FPR. Upon performing pairwise group comparisons in all measures but FPR (Table 4), there were no statistically significant differences between the control and intervention 2 groups. As far as comparisons between the control and intervention 1 groups, there were statistically significant differences in all measures but calibration. Upon comparing the intervention 1 and 2 groups, there were no statistically significant differences except for calibration. Lastly, there were no statistically significant differences in pre-post changes in survey levels of confidence across the three groups (Table 5).

Discussion

This study shows that our educational program consisting of three pre-recorded, online video tutorials and rating 500 candidate IEDs without and with instant feedback is effective in teaching adult and child neurology residents with minimal-or-no prior EEG experience how to correctly identify IEDs on EEG. The program’s effectiveness was reflected in residents’ improvement in confidence and objectively measured skill, contrasted with a lack of improvement among residents who participated as controls.

The educational benefit of our program may be explained in light of a conceptual framework based upon technology-enhanced learning and retrieval practice. The former, which encompasses the use of computer-based technologies to support and mediate educational activities, has been shown to be educationally advantageous by providing better accessibility to the learner and leading to improvement in learners’ knowledge and skill [19, 20]. We used computer-based learning in the development of the three pre-recorded, online video tutorials and the spike test with and without feedback in its computer device/web browser version. Moreover, we used mobile learning in the development of the spike test with feedback in its mobile application-based version. We believe leveraging technology in our educational activities provided learners with accessibility and flexibility thereby facilitating just-in-time learning and conceivably optimizing motivation to learn.

Retrieval practice is defined as the process of retrieving facts, concepts, or events from memory to enhance learning. Evidence suggests that retrieval practice generally outperforms alternative strategies of learning such as restudying and leads to improvement in long-term retention [22, 23]. Notably, not only does this practice improve memory of factual data but also of procedural skills [23]. Further, retrieval opportunities that require schema formation (in other words, those that require learners to generate an organizational structure around the information) rather than retrieval of only isolated facts are preferable for long-term retention. Coupling retrieval practice with feedback amplifies the educational experience as learners use feedback to correct errors and make changes to the retrieval strategies thus allowing them to identify the most effective retrieval mechanisms [23].

We used the principles of retrieval practice to create our spike test and use it as a cornerstone of our educational program. Participants retrieved information about how to correctly identify IEDs (delivered by the video tutorials) as they completed each question in the spike test. This process required learners to repeatedly generate and refine an organizational structure around the six IFCN criteria to define IEDs before answering each question. This organizational structure derived from (i) retrieving detailed information about each of the six IFCN criteria and (ii) applying this knowledge in the evaluation of the candidate IED as to the absence/presence of each of the criteria before finally classifying the waveform as epileptiform or non-epileptiform. The spike test in its version with instant feedback also offered learners the opportunity to correct any errors and dynamically fine-tune their retrieval mechanisms and organizational structures around the six IFCN criteria.

Our results showed that residents who completed the study in either intervention group had statistically significant improvement in IED identification both objectively (increase in accuracy and AUC, and decrease in noise) and subjectively (level of confidence). Further analysis showed that residents in the intervention 1 group (vs. those in the intervention 2 group) also had significant improvement in sensitivity, calibration, and threshold/bias. In terms of residents' subjective improvement, it is unclear why intervention 1 residents had significant improvement in level of confidence regarding identifying epileptiform discharges and committing patients to antiseizure drugs, whereas intervention 2 residents had improvement in level of confidence regarding identifying epileptiform discharges as well as identifying spike-wave discharges.

We speculate that differences in objective IED identification improvement between both intervention groups may be due to better baseline/pre-spike test scores in most parameters (e.g., accuracy, sensitivity, calibration, threshold/bias, AUC, and noise) among residents randomized to the intervention 2 group. This may be because residents in the intervention 2 group were at different levels of training (postgraduate years (PGY)), whereas residents in the intervention 1 group were all in their first year of training (PGY1s). Residents who are more senior in neurology residency may have acquired epilepsy/EEG knowledge along their training even without having formally had prior EEG experience. An alternative, but not mutually exclusive, explanation would be related to how they practiced rating IEDs with instant feedback.

Intervention 1 group residents accessed the spike test with feedback through a web browser, which allowed participants to view EEGs on a larger screen (i.e., standard computer screen) and modify montages and adjust sensitivity. These features may have better reflected the "real-life" exercise of reading EEGs therefore maximizing residents' engagement and promoting optimized consequential impact. Conversely, intervention 2 group residents accessed the spike test with feedback through smartphones only. This method offered more flexibility and versatility, as accessing mobile devices is "quick and easy", as well as convenience, as learners were able to practice in their own environments and on their own time (e.g., at home or during commute) [26]. However, this method offered a smaller viewing screen (mobile device screen) and did not allow changes in montage or sensitivity - as candidate IEDs were shown as static EEG epochs in bipolar montage.

We hypothesize that completing the spike test with feedback on a computer/web browser (intervention 1 group) – as opposed to on a mobile application (intervention 2 group) - was more conducive to successful retrieval practice. The characteristics of the former method, as detailed in the paragraph above, conceivably better allowed residents to generate and refine an organizational structure around the six IFCN criteria and progressively adjust retrieval mechanisms in response to feedback. Another explanation would be that our cohort of residents was too small to capture statistically significant improvements in all performance measures and in both groups.

We believe that the measured improvement in IED identification seen in our study is exclusively due to our educational intervention because (a) participants had minimal-to-no prior EEG exposure thus removing this potential confounding factor and (b) participants who were controls had no statistically significant improvement - neither objectively nor subjectively. The latter observation also suggests that exclusively completing the spike test at two different time points, without feedback or any other teaching intervention, does not have educational benefit. Consequently, future studies using the spike test as an assessment tool may not require an “inactive” control group.

Our educational program has many strengths. First, it includes a pool of thousands of expert-labeled candidate IEDs from which samples are collected to build unique 500-EEG spike tests - without feedback (for assessment purposes) and with feedback (for teaching purposes). This aspect allows residents to be exposed to a large number of candidate IEDs, from a large and representative group of patients, in a relatively short period of time. This is valuable as time dedicated to EEG learning during residency training is limited both in the U.S. and many European countries [27, 28]. Furthermore, insufficient EEG exposure is a pervasive issue and a well-known barrier to resident EEG education [7, 27, 28]. This aspect is also valuable since preliminary evidence suggests that a higher number of EEGs read is associated with greater long-term retention [29]. Comparatively, prior published EEG teaching and assessment tools have included a significantly lower number of EEGs/ candidate IEDs [6, 18, 29–37]. Additionally, our educational EEG content brings input from multiple experts since all candidate IEDs had been labeled by eight experts. This feature contrasts with the standard method of teaching EEG (apprenticeship) where learners typically receive one opinion through reviewing studies with a single expert. The latter approach may not be ideal from an educational standpoint given that the same EEG may be interpreted differently by experts - primarily due to their idiosyncratic over- or under-calling preferences [24, 38].

Second, our program includes three pre-recorded, online video tutorials, which were given by experts (FAN, JJ, MBW). In addition to being founded upon technology-enhanced learning, the video tutorials incorporated principles of the cognitive theory of multimedia learning [39]. Tutorials were a combination of pictures – including an infographic showing all six IFCN criteria to define IEDs [25] (organization and explanative graphic) and examples of candidate IEDs - and key words (multimedia principle). This infographic served as a blueprint for learners to generate an organizational structure around the six IFCN criteria, which was refined during the retrieval practice stage of our program (spike test with and without feedback). Further, we excluded decorative elements to avoid competing for learners' processing capacity (coherence principle) and avoided on-screen text that was redundant with our narration (redundancy principle). We added vocal and visual cues to draw learners' attention to key concepts (signaling principle) and divided the tutorials into three videos that were short thus allowing learners to progress through at their own pace (segmenting principle). Lastly, the tone of the tutorials was casual and conversational rather than a formal expository narration (personalization principle).

As noted above, our video tutorials focused on teaching and applying the IFCN criteria to define IEDs on EEG [25]. This standardized set of operational criteria has the major advantage of decomposing the complex judgement involved in IED identification into a series of six binary decisions about the presence or absence of elementary waveform features. Interestingly, this strategic decomposition of complex judgements is also effective in other fields [40]. In addition to being useful in clinical practice by EEG experts [41], teaching the IFCN criteria to a group of seven neurology trainees/recent graduates has been shown to improve diagnostic performance and inter-rater agreement [18].

Third, we evaluated curricular outcomes by assessing residents' level of confidence (Level 1 on Kirkpatrick model) as well as knowledge/skills (Level 2 on Kirkpatrick model) [42]. The former was measured by pre-/post-surveys, and the latter by pre-/post-spike tests (without feedback). A major advantage of the spike test as an assessment tool is that each test is unique as it presents a different set of 500 candidate IEDs. This ensured that residents did not memorize or share test questions. A second major advantage relates to the large number of candidate IEDs included in each spike test. The length of the test, in fact, was determined based on our prior work showing that rating approximately 500 candidate IEDs is needed to precisely categorize a test taker's skill level in IED identification [43].

We believe that this conceptual framework founded upon technology-enhanced learning and retrieval practice may be used to teach other abilities within the skill set of reading EEGs. For instance, additional educational programs may be developed to teach learners how to accurately and reliably identify other "must-know" routine EEG findings [13], critical care EEG findings, and intra-cranial EEG patterns. Further, we hypothesize that this framework may also be valuable in the education of other visual pattern-based skills within clinical neurophysiology such as evoked potentials studies, nerve conduction studies, and electromyography, and in other areas such as electrocardiograms.

Our study has several limitations. Our educational program focused only on identification of IEDs. This skill is arguably the most important in EEG reading given the critical role of the presence of IEDs in the diagnosis and management of epilepsy [10, 11, 14]. Nonetheless, EEG reading proficiency also requires skills in other domains, which were not covered in our program. These include identification of other pertinent normal and abnormal EEG findings [13], the ability to read an EEG from beginning to end and select regions of interest for further analysis, and the mastery of generating a report. Further, candidate IEDs used in our educational program were presented in short 10-second EEG segments. This characteristic precluded participants from assessing additional background EEG activity and potentially other similar candidate IEDs. The latter is particularly relevant as evidence suggests that candidate IEDs that repeat themselves may require fewer of the IFCN criteria to qualify as epileptiform [44]. Lastly, there may have been a selection bias insofar that residents who agreed to participate may have been inherently more interested in EEG.

As far as our expert consensus-based gold standard, it is grounded in expert experience rather than video-EEG-confirmed epilepsy. Nevertheless, an expert consensus gold standard has the advantage of allowing the collection of candidate IEDs from a large and representative patient population given that a large portion of patients with (epileptic or non-epileptic) seizures do not undergo video-EEG monitoring. Moreover, we defined IEDs as those candidate IEDs that received more than four votes from the group of eight experts [24]. This threshold was determined by consensus among the authors in light of the inherent tradeoff between sensitivity and FPR associated with varying thresholds. Lastly, all but one of the eight experts either trained or currently practice at the same institution (Massachusetts General Hospital). This may represent a bias where these experts may share a similar reasoning upon rating candidate IEDs on EEG.

Future directions for our program include expanding our cohort of participants by recruiting trainees from programs at other institutions, nationally and internationally. These trainees may include not only residents but also fellows, medical students, advanced practice providers, neurophysiologists, and EEG technologists. In fact, we anticipate that our program – particularly its video tutorial component – may even be more effective for residents with prior EEG exposure [45]. A larger study cohort may allow us to further our understanding on how trainees learn to identify IEDs, and to better understand the educational advantages and disadvantages of using mobile- and computer-based EEG learning. A larger study may also allow us to re-test learners to assess long-term retention. Further, we may incorporate our automated algorithm that quantifies the six IFCN criteria into our educational program as this user-friendly model may enhance learning and application of the IFCN criteria by trainees [46]. Additionally, we plan to enrich the evaluation of our program by measuring how much it influences residents' behavior in identifying IEDs "on the job" (Level 3 on Kirkpatrick model) and the translation of this behavior change upon clinical outcomes (Level 4 on Kirkpatrick model) [42]. Furthermore, we plan to design a head-to-head randomized controlled educational trial comparing our program with standard methods of EEG teaching in neurology residency training [27, 28]. Finally, we hope to expand our sample of candidate IEDs and develop a robust gold standard merging expert consensus and an externally validated source based on video-EEG data capturing patients' habitual spells [38].

In conclusion, we describe our educational program and report its effectiveness in teaching bi-institutional adult and child neurology residents with minimal-to-no prior EEG experience how to correctly identify IEDs on EEG. Future studies should investigate our program's effectiveness in larger cohorts of multi-institutional trainees at different levels of training. We believe our program has the potential to complement trainee EEG education concerning

IED identification thus helping ensure that the ACGME Level 4 EEG milestone regarding “interpreting common EEG abnormalities” is universally met. Ensuring that trainees are competent in accurate and reliable IED identification is imperative because it directly translates into better care of patients with seizures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

Dr. Westover was supported by grants from the NIH (R01NS102190, R01NS102574, R01NS107291, RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598), and NSF (2014431).

This work has been previously presented at the American Academy of Neurology 2022 Annual Meeting and American Clinical Neurophysiology 2022 Annual Meeting.

TEST YOURSELF

1. Our educational program, which was developed upon a conceptual framework based on technology-enhanced learning and retrieval practice, is effective in teaching adult and child neurology residents with minimal-or-no prior EEG experience how to correctly identify IEDs on EEG.
   - **A.** True
   - **B.** False
2. Technology-enhanced learning encompasses the use of computer-based technologies to support and mediate educational activities.
   - **A.** True
   - **B.** False
3. Retrieval practice is defined as the process of retrieving facts, concepts, or events from memory to enhance learning.
   - **A.** True
   - **B.** False

References

1. Adornato BT, Drogan O, Thoresen P, Coleman M, Henderson VW, Henry KA et al. The practice of neurology, 2000–2010: report of the AAN Member Research Subcommittee. Neurology 2011 Nov 22;77(21):1921–8. [PubMed: 22031533]
2. Benbadis SR. “Just like EKGs!” Should EEGs undergo a confirmatory interpretation by a clinical neurophysiologist? Neurology 2013 Jan 1;80(1 Suppl 1):S47–51. [PubMed: 23267045]
3. Kang PB, Bale JF Jr., Mintz M, Joshi SM, Gilbert DL, Radabaugh C et al. The child neurology clinical workforce in 2015: Report of the AAP/CNS Joint Taskforce. Neurology 2016 Sep 27;87(13):1384–92. [PubMed: 27566740]

4. Mahajan A, Cahill C, Scharf E, Gupta S, Ahrens S, Joe E et al. Neurology residency training in 2017: A survey of preparation, perspectives, and plans. Neurology 2019 Jan 8;92(2):76–83. [PubMed: 30518554]
5. Daniello KM, Weber DJ. Education Research: The current state of neurophysiology education in selected neurology residency programs. Neurology 2018 Apr 10;90(15):708–11. [PubMed: 29632112]
6. Nascimento FA, Maheshwari A, Chu J, Gavvala JR. EEG education in neurology residency: background knowledge and focal challenges. Epileptic Disord2020 Dec 1;22(6):769–74. [PubMed: 33399093]
7. Lourenco ES, Kowacs DP, Gavvala JR, Kowacs PA, Nascimento FA. EEG education in Brazil: a national survey of adult neurology residents. Arq Neuropsiquiatr2022 Jan;80(1):43–47. [PubMed: 34755770]
8. Barratt DLS, Chiota-McCollum N, McClean J, Dewey J, Potrebic S, et al. Neurology Milestones. The Accreditation Council for Graduate Medical Educaion (ACGME). Second revision.2020: Available from: https://www.acgme.org/what-we-do/accreditation/milestones/milestones-by-specialty/.
9. Albert DVFGA, Bass N, Lotze T, Bodensteiner J, Mink J, et al. Child Neurology Milestones. The Accreditation Council for Graduate Medical Education (ACGME). Second revision. 2020.
10. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017 Apr;58(4):522–30. [PubMed: 28276060]
11. Pillai J, Sperling MR. Interictal EEG and the diagnosis of epilepsy. Epilepsia 2006;47 Suppl 1:14–22.
12. van Donselaar CA, Schimsheimer RJ, Geerts AT, Declerck AC. Value of the electroencephalogram in adult patients with untreated idiopathic first seizures. Arch Neurol1992 Mar;49(3):231–7. [PubMed: 1536624]
13. Nascimento FA, Jing J, Strowd R, Sheikh IS, Weber D, Gavvala JR et al. Competency-based EEG education: a list of “must-know” EEG findings for adult and child neurology residents. Epileptic Disord2022 Oct 1;24(5):979–82. [PubMed: 35904042]
14. Goodin DS, Aminoff MJ. Does the interictal EEG have a role in the diagnosis of epilepsy? Lancet 1984 Apr 14;1(8381):837–9. [PubMed: 6143148]
15. Benbadis SR, Tatum WO. Overintepretation of EEGs and misdiagnosis of epilepsy. J Clin Neurophysiol2003 Feb;20(1):42–4. [PubMed: 12684557]
16. Oto MM. The misdiagnosis of epilepsy: Appraising risks and managing uncertainty. Seizure 2017 Jan;44:143–46. [PubMed: 28017581]
17. Smith D, Defalla BA, Chadwick DW. The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic. QJM 1999 Jan;92(1):15–23. [PubMed: 10209668]
18. Kural MA, Aydemir ST, Levent HC, Olmez B, Ozer IS, Vlachou M et al. The operational definition of epileptiform discharges significantly improves diagnostic accuracy and inter-rater agreement of trainees in EEG reading. Epileptic Disord 2022 Apr 1;24(2):353–58. [PubMed: 34903504]
19. Cook DA, Ellaway RH. Evaluating technology-enhanced learning: A comprehensive framework. Med Teach2015;37(10):961–70. [PubMed: 25782599]
20. Sen ALC. Technology-enhanced learning. T A, editor: Springer, Cham; 2020.
21. Green ML, Moeller JJ, Spak JM. Test-enhanced learning in health professions education: A systematic review: BEME Guide No. 48. Med Teach2018 Apr;40(4):337–50. [PubMed: 29390949]
22. Larsen DP. When I say ... test-enhanced learning. Med Educ2013 Oct;47(10):961. [PubMed: 24016165]
23. Larsen DP. Planning Education for Long-Term Retention: The Cognitive Science and Implementation of Retrieval Practice. Semin Neurol2018 Aug;38(4):449–56. [PubMed: 30125899]

24. Jing J, Herlopian A, Karakis I, Ng M, Halford JJ, Lam A et al. Interrater Reliability of Experts in Identifying Interictal Epileptiform Discharges in Electroencephalograms. JAMA Neurol2020 Jan 1;77(1):49–57. [PubMed: 31633742]
25. Kane N, Acharya J, Benickzy S, Caboclo L, Finnigan S, Kaplan PW et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. Clin Neurophysiol Pract2017;2:170–85. [PubMed: 30214992]
26. Walsh K. Mobile Learning in Medical Education: Review. Ethiop J Health Sci2015 Oct;25(4):363–6. [PubMed: 26949301]
27. Nascimento FA, Gavvala JR. Education Research: Neurology Resident EEG Education: A Survey of US Neurology Residency Program Directors. Neurology 2021 Apr 27;96(17):821–24. [PubMed: 33310878]
28. Nascimento FA, Gavvala JR, Tankisi H, Beniczky S. Neurology resident EEG training in Europe. Clin Neurophysiol Pract2022;7:252–59. [PubMed: 36133398]
29. Fahy BG, Chau DF, Ozrazgat-Baslanti T, Owen MB. Evaluating the long-term retention of a multidisciplinary electroencephalography instructional model. Anesth Analg2014 Mar;118(3):651–6. [PubMed: 24557110]
30. Dericioglu N, Ozdemir P. The Success Rate of Neurology Residents in EEG Interpretation After Formal Training. Clin EEG Neurosci2018 Mar;49(2):136–40. [PubMed: 29017369]
31. Weber D, McCarthy D, Pathmanathan J. An effective automated method for teaching EEG interpretation to neurology residents. Seizure 2016 Aug;40:10–2. [PubMed: 27295562]
32. Chau DF, Bensalem-Owen MK, Fahy BG. The effectiveness of an interdisciplinary approach to EEG instruction for residents(r). J Clin Neurophysiol2010 Apr;27(2):106–9. [PubMed: 20505373]
33. Chau D, Bensalem-Owen M, Fahy BG. The impact of an interdisciplinary electroencephalogram educational initiative for critical care trainees. J Crit Care2014 Dec;29(6):1107–10. [PubMed: 25056845]
34. Fahy BG, Chau DF, Bensalem-Owen M. Evaluating the requirements of electroencephalograph instruction for anesthesiology residents. Anesth Analg2009 Aug;109(2):535–8. [PubMed: 19608829]
35. Fahy BG, Vasilopoulos T, Chau DF. Use of Flipped Classroom and Screen-Based Simulation for Interdisciplinary Critical Care Fellow Teaching of Electroencephalogram Interpretation. Neurocrit Care2020 Aug;33(1):298–302. [PubMed: 32424536]
36. Fahy BG, Cibula JE, Johnson WT, Cooper LA, Lizdas D, Gravenstein N et al. An online, interactive, screen-based simulator for learning basic EEG interpretation. Neurol Sci2021 Mar;42(3):1017–22. [PubMed: 32700228]
37. Leira EC, Bertrand ME, Hogan RE, Cruz-Flores S, Wyrwich KW, Albaker OJ et al. Continuous or emergent EEG: can bedside caregivers recognize epileptiform discharges? Intensive Care Med2004 Feb;30(2):207–12. [PubMed: 14615839]
38. Nascimento FA, Jing J, Beniczky S, Benbadis SR, Gavvala JR, Yacubian EMT et al. One EEG, one read - A manifesto towards reducing interrater variability among experts. Clin Neurophysiol2022 Jan;133:68–70. [PubMed: 34814017]
39. Cavanagh TM, Kiersch C. Using commonly-available technologies to create online multimedia lessons through the application of the Cognitive Theory of Multimedia Learning. Educ Technol Res Dev2022 Dec 19:1–21.
40. Kahneman DSO, Sunstein CR. Noise: a flaw in human judgement. New York: Little Brown Spark; 2021.
41. Kural MA, Duez L, Sejer Hansen V, Larsson PG, Rampp S, Schulz R et al. Criteria for defining interictal epileptiform discharges in EEG: A clinical validation study. Neurology 2020 May 19;94(20):e2139–e47. [PubMed: 32321764]
42. Kirkpatrick JDKW. Four Levels of Training Evaluation: Association for Talent Development.; 2016.
43. Harid NM, Jing J, Hogan J, Nascimento FA, Ouyang A, Zheng WL et al. Measuring expertise in identifying interictal epileptiform discharges. Epileptic Disord2022 Jun 1;24(3):496–506. [PubMed: 35770748]

44. Kural MA, Qerama E, Johnsen B, Fuchs S, Beniczky S. The influence of the abundance and morphology of epileptiform discharges on diagnostic accuracy: How many spikes you need to spot in an EEG. Clin Neurophysiol2021 Jul;132(7):1543–49. [PubMed: 34030055]
45. Nascimento FA, Moore M, Olandoski M, Gavvala JR. Short-term effectiveness of a condensed series of standard EEG lectures for adult neurology residents. Epileptic Disord2022 Apr 1;24(2):397–403. [PubMed: 35067482]
46. Nascimento FA, Barfuss JD, Jaffe A, Brandon Westover M, Jing J. A quantitative approach to evaluating interictal epileptiform discharges based on interpretable quantitative criteria. Clin Neurophysiol2022 Nov 17;146:10–17. [PubMed: 36473334]

Bullet points

1. This novel educational program can be used to teach trainees how to identify interictal epileptiform discharges (IEDs) on EEG.
2. This educational trial showed that technology-enhanced learning and retrieval practice can effectively teach IED identification.
3. This conceptual framework may be used to teach other abilities within the skill set required to interpret EEGs.

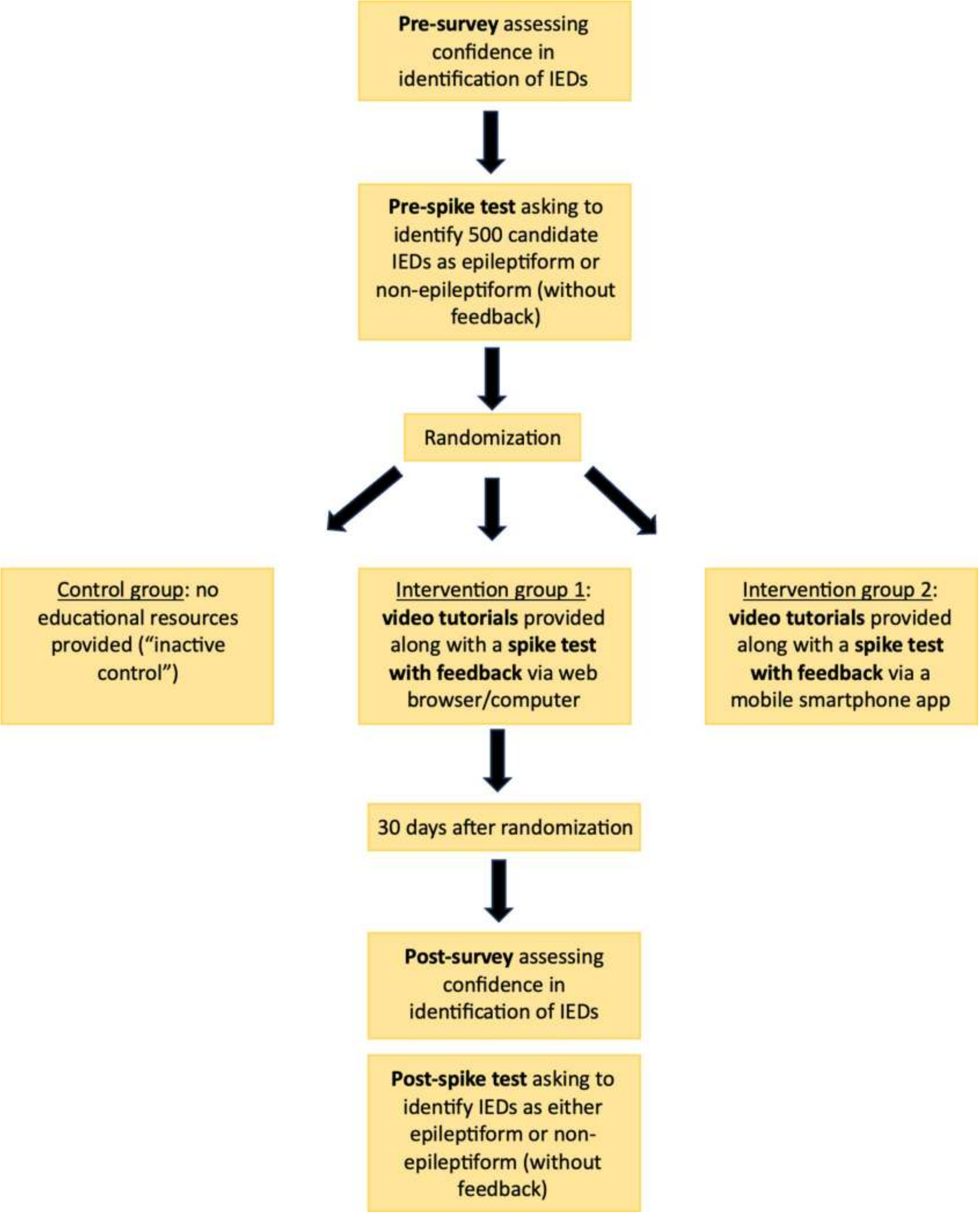

Figure 1:

Pathway of subjects during educational pilot trial including sample sizes for each study stage. IED, interictal epileptiform discharge.

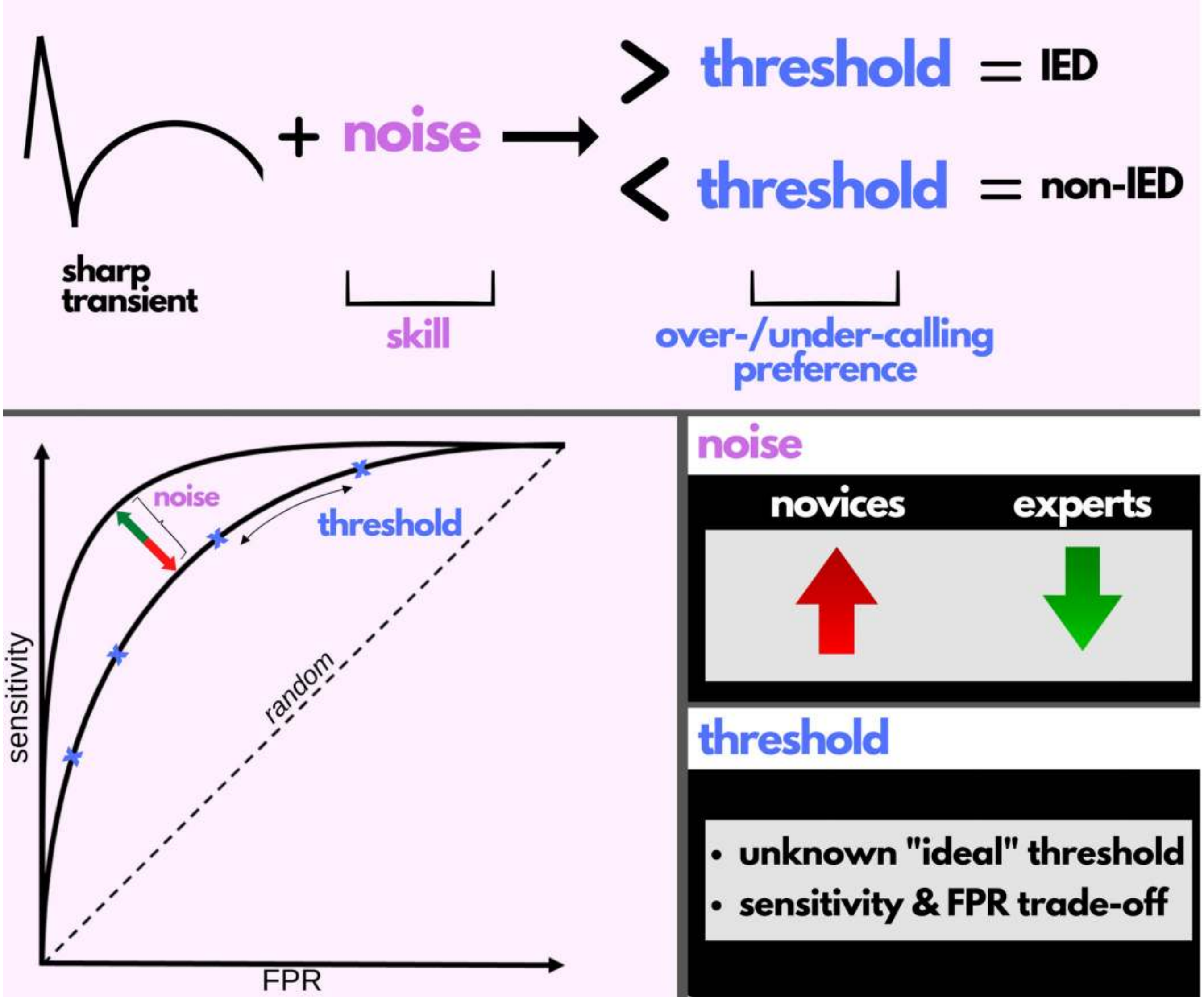

Figure 2:

The latent trail model of the psychological process related to classifying interictal epileptiform discharges (IEDs) on EEG in a binary fashion (IED vs. non-IED). This model has two parameters or “latent traits”: noise/uncertainty level, which reflects a participant’s skill in recognizing IEDs, and threshold/bias, which represents a participant’s personal preference as an over- or under-caller.

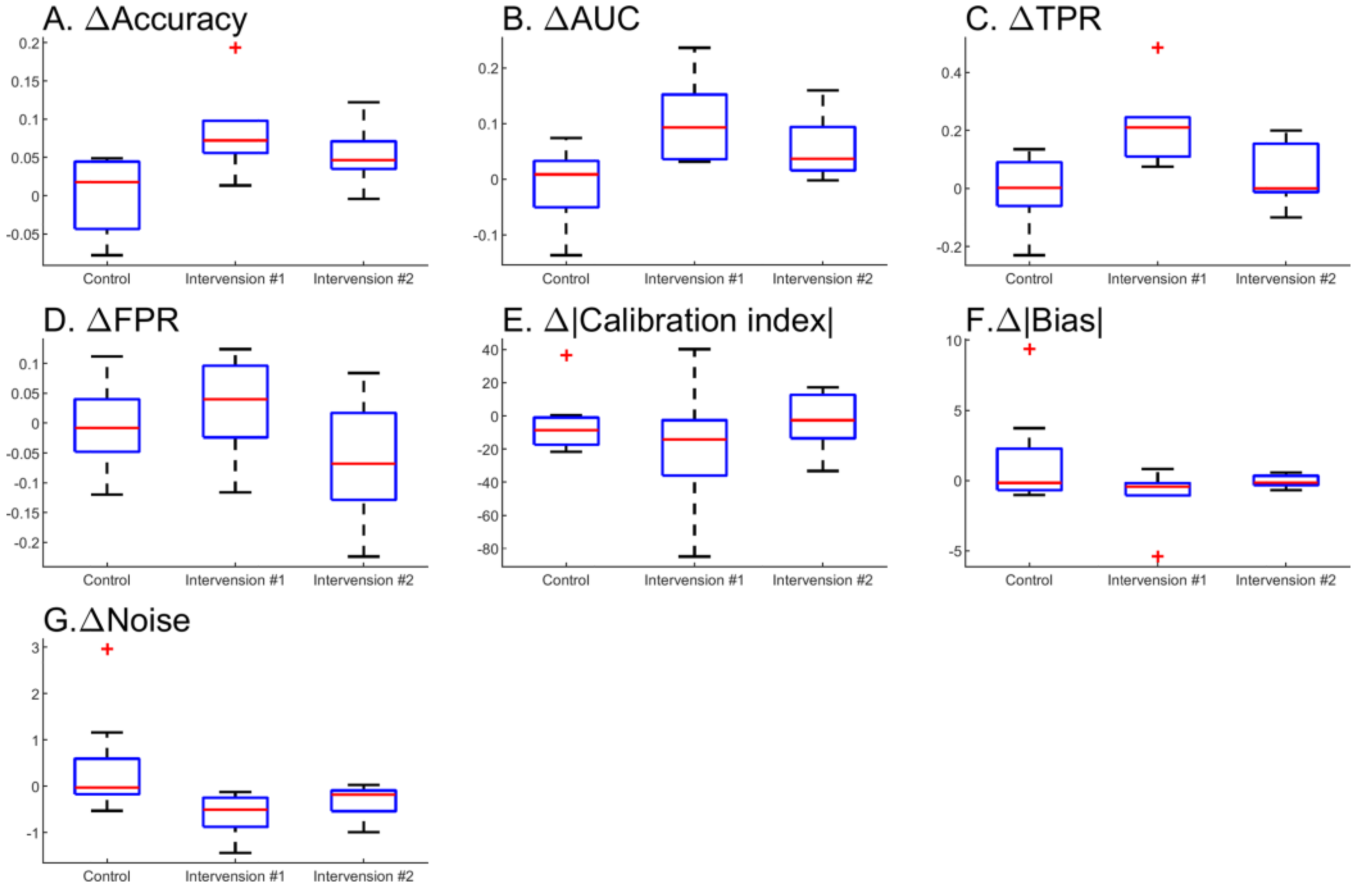

Figure 3:

Boxplots representing the difference in seven different measures (A. accuracy; B. area under the curve/AUC; C. true positive rate/sensitivity; D. false positive rate/1-specificity; E. |calibration index|; F. |threshold/bias|; and G. noise (dB)) in the pre- and post-spike tests stratified by the three different study groups: control, intervention 1, and intervention 2.

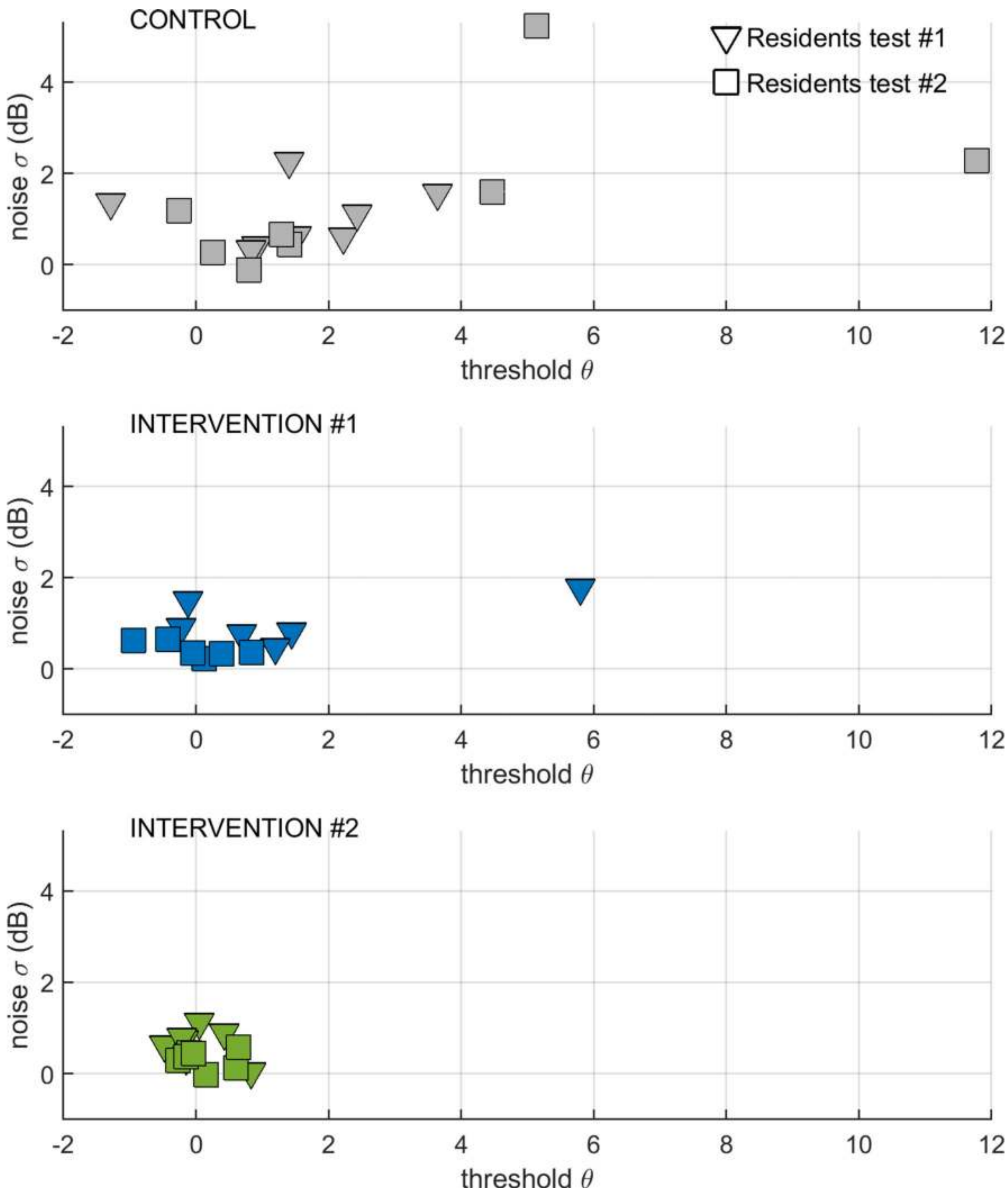

Figure 4:

Estimation of individual participant's internal parameters threshold/bias (θ) and noise (n) with latent trait model parameters from pre- and post-spike tests (test #1/triangles and test #2/squares, respectively) in the three study groups: intervention 1 (middle panel; blue), intervention 2 (lower panel; green), and control (upper panel; gray).

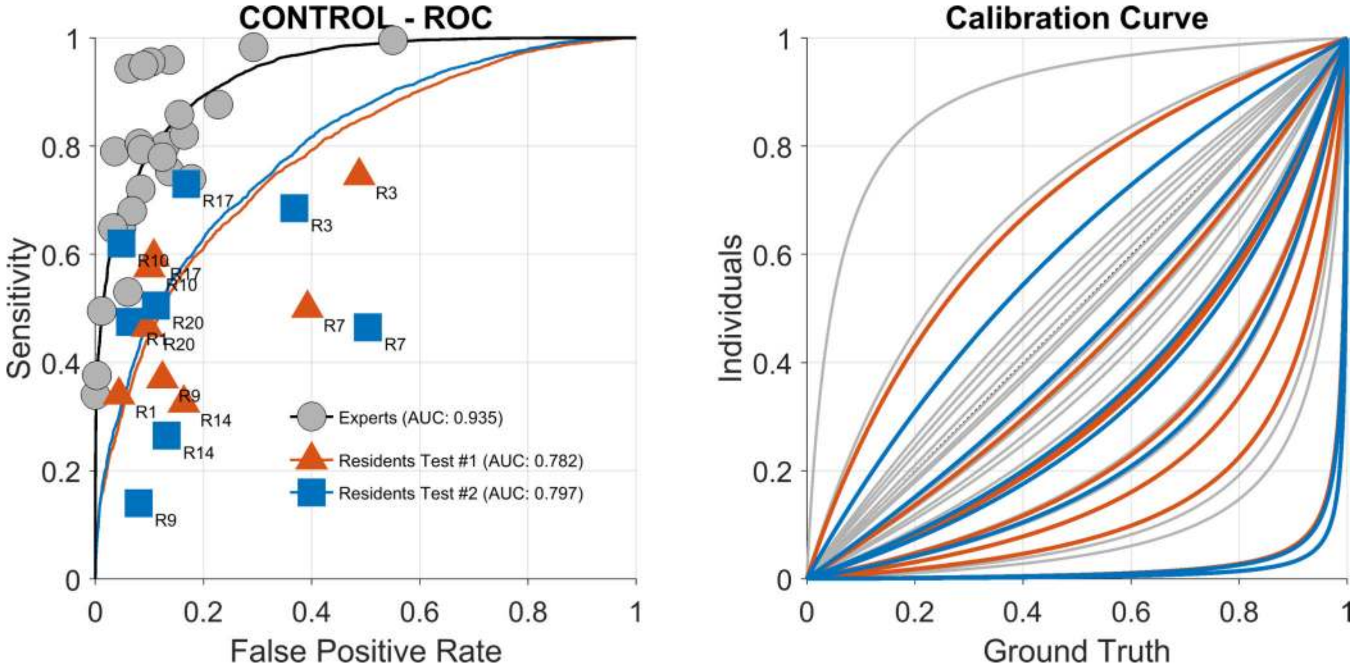

Figure 5:

Left panel: receiver operating characteristic (ROC) curve to fit performance of participants in the control group along with individual operating points based on the pre-spike test (red curve, red triangles; test #1) and post-spike test (blue curve, blue squares; test #2). ROC curve to fit 24 experts' scores (black curve) who completed the spike test with operating points represented as grey circles. Right panel: individual parametric calibration curves to fit the performance of participants in the control group based on the pre-spike test (red curves) and post-spike test (blue curves) indicating the probability of a participant to rate candidate interictal epileptiform discharges (IEDs) within a given bin as IEDs. Individual parametric calibration curves to fit 24 experts' scores (grey curves).

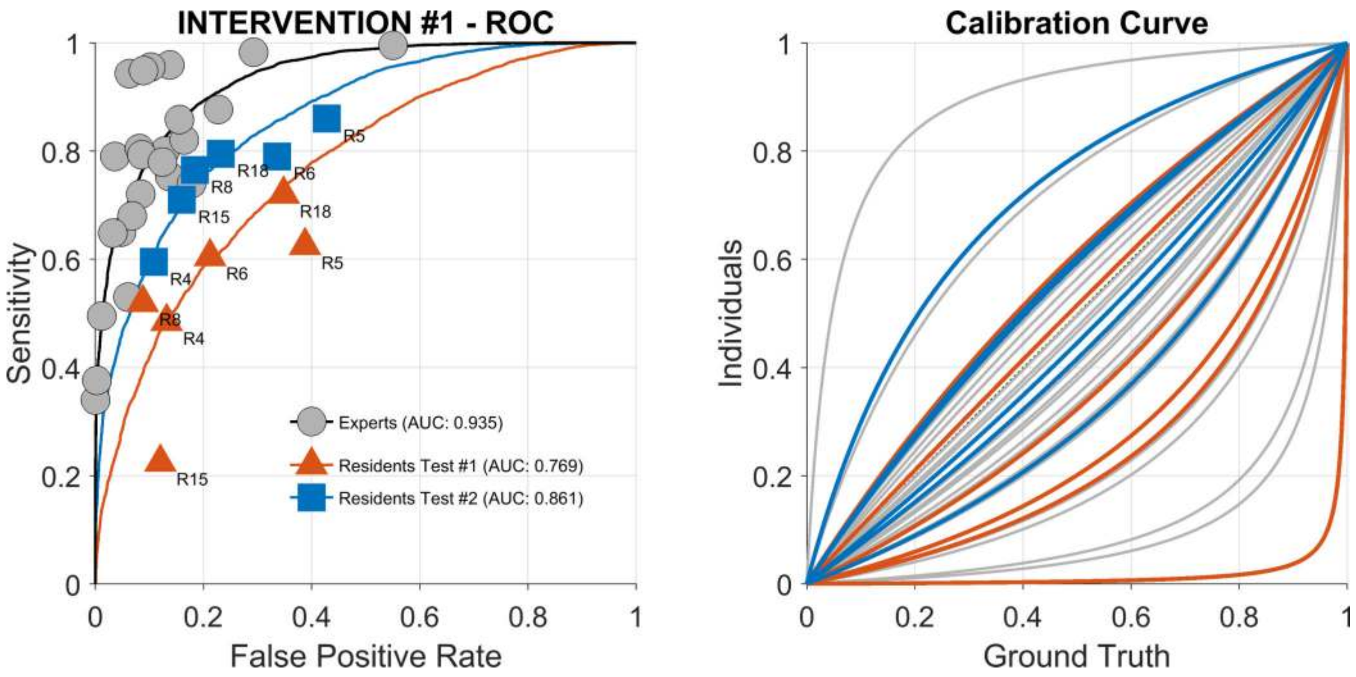

Figure 6:

Left panel: receiver operating characteristic (ROC) curve to fit performance of participants in the intervention 1 group along with individual operating points based on the pre-spike test (red curve, red triangles; test #1) and post-spike test (blue curve, blue squares; test #2). ROC curve to fit 24 experts' scores (black curve) who completed the spike test with operating points represented as grey circles. Right panel: individual parametric calibration curves to fit the performance of participants in the intervention 1 group based on the pre-spike test (red curves) and post-spike test (blue curves) indicating the probability of a participant to rate candidate interictal epileptiform discharges (IEDs) within a given bin as IEDs. Individual parametric calibration curves to fit 24 experts' scores (grey curves).

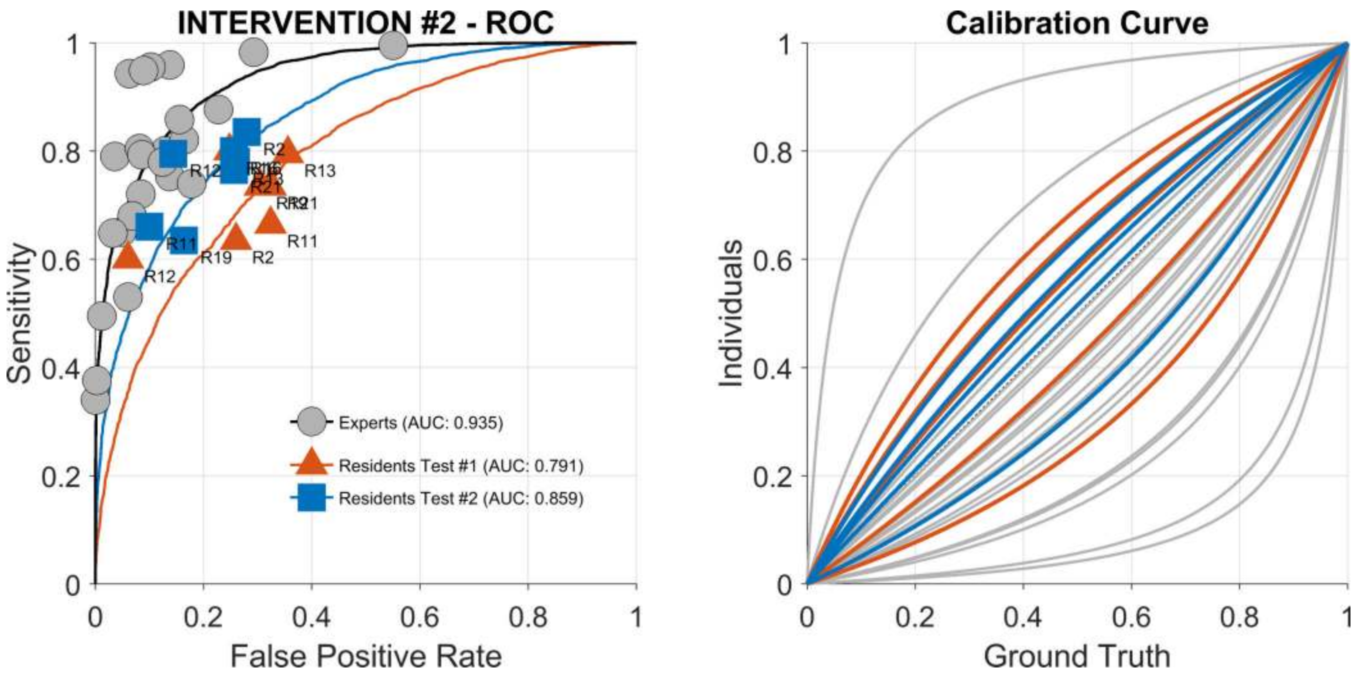

Figure 7:

Left panel: receiver operating characteristic (ROC) curve to fit performance of participants in the intervention 2 group along with individual operating points based on the pre-spike test (red curve, red triangles; test #1) and post-spike test (blue curve, blue squares; test #2). ROC curve to fit 24 experts' scores (black curve) who completed the spike test with operating points represented as grey circles. Right panel: individual parametric calibration curves to fit the performance of participants in the intervention 2 group based on the pre-spike test (red curves) and post-spike test (blue curves) indicating the probability of a participant to rate candidate interictal epileptiform discharges (IEDs) within a given bin as IEDs. Individual parametric calibration curves to fit 24 experts' scores (grey curves).

Table 1:

Summary of study participants including study group, neurology residency demographics, prior EEG experience, and time between finalizing the pre- and post- spike tests.

ID	Study group	Adult vs. pediatric	PGY	Institution	Prior EEG experience	Time (days) between pre- and post-spike tests
1	Control	Adult	2	MGB	0	24
2	Control	Adult	2	MGB	0	25
3	Control	Adult	1	MGB	0	18
4	Control	Adult	2	MGB	0	25
5	Control	Adult	2	MGB	0	34
6	Control	Adult	2	MGB	0	40
7	Control	Adult	1	MGB	0	26
8	Control	Adult	2	MGB	0	26
9	Intervention 1	Adult	1	MGB	0	39
10	Intervention 1	Adult	1	Yale	0	62
11	Intervention 1	Adult	1	Yale	0	27
12	Intervention 1	Adult	1	Yale	0	38
13	Intervention 1	Adult	1	MGB	0	33
14	Intervention 1	Adult	1	Yale	0	68
15	Intervention 2	Adult	1	MGB	0	51
16	Intervention 2	Pediatric	4*	MGB	0	35
17	Intervention 2	Adult	2	MGB	2 weeks	33
18	Intervention 2	Adult	2	Yale	4 weeks	17
19	Intervention 2	Adult	2	Yale	0	48
20	Intervention 2	Pediatric	1	MGB	0	65
21	Intervention 2	Adult	1	Yale	0	104

ID, participant identifier; PGY, postgraduate year

*, 1st year in neurology residency training.

Table 2:

Descriptive analysis of study participants including demographics, neurology residency characteristics, and study completion time between finalizing the pre- and post-spike tests. Group comparison performed using one-way ANOVA after logarithmic transformation (*p**).

	Control (n=8)	**Intervention #1 (n=6)**	**Intervention #2 (n=7)**	***p****
Time (days) between pre- and post-spike tests	27.3±2.4	44.5±6.8	50.4±10.6	0.052
Institution				
MGB	8 (100%)	2 (33.3%)	4 (57.1%)	
Yale	0 (0%)	4 (66.7%)	3 (42.9%)	-
Adult vs. pediatric				
Adult	8 (100%)	6 (100%)	5 (71.4%)	
Pediatric	0 (0%)	0 (0%)	2 (28.6%)	-
PGY				
1	2 (25%)	6 (100%)	3 (42.9%)	
2	6 (75%)	0 (0%)	3 (42.9%)	
4‡	0 (0%)	0 (0%)	1 (14.3%)	-

‡, 1st year in neurology residency training; PGY, postgraduate year.

Table 3:

Summary of educational trial results including performance metrics (accuracy, sensitivity, false positive rate/FPR, calibration, threshold/bias, area under the curve/AUC, and noise) related to the pre- and post-spike tests stratified by study group. Intra-group comparisons (pre-test metric vs. post-test metric) were calculated using non-parametric Wilcoxon test (*p**). Inter-group comparisons (pre-post change in each metric between the three study groups) were calculated using non-parametric Kruskal-Wallis test (*p***).

	Control (n=8)	**Intervention #1 (n=6)**	**Intervention #2 (n=7)**	***p*****
Accuracy (mean±SEM)				
Pre-test	0.67±0.03	0.67±0.02	0.72±0.02	
Post-test	0.67±0.04	0.76±0.02	0.77±0.01	
Post-test – pre-test	0.00	0.08	0.05	0.019[†]
*p**	0.833	0.028[†]	0.028[†]	
Sensitivity (mean±SEM)				
Pre-test	0.49±0.05	0.53±0.07	0.71±0.03	
Post-test	0.49±0.07	0.75±0.04	0.75±0.03	
Post-test – pre-test	0.00	0.22	0.04	0.021[†]
*p**	1.000	0.028[†]	0.463	
FPR (mean±SEM)				
Pre-test	0.19±0.06	0.21±0.05	0.27±0.04	
Post-test	0.18±0.06	0.24±0.05	0.21±0.03	
Post-test – pre-test	−0.01	0.03	−0.06	0.209
*p**	0.779	0.344	0.237	
Calibration (mean±SEM)				
Pre-test	−40±14	−33±16	3.8±8.0	
Post-test	−39±14	3.3±10	0.37±6.6	
Post-test – pre-test	0.87	36	−3.4	0.030[†]
*p**	0.575	0.046[†]	0.735	
Threshold, θ (mean±SEM)				
Pre-test	1.5±0.51	1.5±0.91	0.04±0.17	
Post-test	3.1±1.4	−0.01±0.25	0.12±0.14	
Post-test – pre-test	1.7	−1.5	0.08	0.027[†]
*p**	0.263	0.046[†]	0.612	
AUC (mean±SEM)				
Pre-test	0.75±0.04	0.74±0.03	0.81±0.02	
Post-test	0.74±0.05	0.85±0.01	0.86±0.01	
Post-test – pre-test	−0.01	0.11	0.06	0.027[†]
*p**	0.889	0.028[†]	0.028[†]	
Noise, n (mean±SEM)				
Pre-test	1.0±0.24	1.0±0.20	0.65±0.14	
Post-test	1.4±0.61	0.42±0.07	0.32±0.08	

	Control (n=8)	**Intervention #1 (n=6)**	**Intervention #2 (n=7)**	***p*****
Post-test – pre-test	0.40	−0.62	−0.33	0.037†
*p**	1.000	0.028†	0.028†	

SEM, standard error of the mean.

†, statistically significant ($p<0.05$).

Table 4:

Post-hoc inter-group pairwise comparisons (pre-post change in each metric) were calculated using Dunn's test followed by Bonferroni correction (p^*).

Groups compared (p^*)	**Difference accuracy**	**Difference sensitivity**	**Difference calibration**	**Difference threshold/ θ**	**Difference AUC**	**Difference noise/ n**
I1 x I2	0.773	0.106	0.033†	0.120	0.869	0.812
I1 x C	0.017†	0.022†	0.132	0.029†	0.025†	0.034†
I2 x C	0.284	1	1	1	0.320	0.426

I1, intervention 1 group; I2, intervention 2 group; C, control. AUC, area under the curve;

†, statistically significant ($p<0.05$).

Table 5:

Summary of educational trial results including level of confidence related to the pre- and post-surveys stratified by study group. Intra-group comparisons (pre-survey metric vs. post-survey metric) were calculated using non-parametric Wilcoxon test (p*). Inter-group comparisons (pre-post change in each metric between the three study groups) were calculated using non-parametric Kruskal-Wallis test (p**).

	Control (n=8)	**Intervention #1 (n=6)**	**Intervention #2 (n=7)**	***p*****
Epileptiform discharge (mean)				
Pre-survey	1.50	2.33	2.00	
Post-survey	2.13	3.33	3.14	
Post-survey – pre-survey	0.63	1.00	1.14	0.449
p*	0.059	0.109	0.023†	
Spike-wave discharge (mean)				
Pre-survey	1.63	2.50	2.00	
Post-survey	2.25	3.50	3.14	
Post-survey – pre-survey	0.63	1.00	1.14	0.306
p*	0.059	0.063	0.011†	
Epileptiform discharge and committing patient to an ASD (mean)	1.25	1.33	1.43	
Pre-survey	1.63	2.33	2.00	
Post-survey	0.38	1.00	0.57	0.405
Post-survey – pre-survey	0.317	0.034†	0.102	
p*				

†, statistically significant ($p<0.05$). ASD, antiseizure drug.