Received: 27 September 2021 | Revised: 2 February 2022 | Accepted: 8 April 2022

DOI: 10.1111/psyp.14223

PSYCHOPHYSIOLOGY SPR WILEY

REGISTERED REPORT

Impact of reduced-dimensionality independent components analysis on event-related potential measurements

Victor J. Pokorny[1,2] | Scott R. Sponheim[1,2,3] | Eric Rawls[3]

[1]Minneapolis Veterans Affairs Health Care System, Minneapolis, Minnesota, USA

[2]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA

[3]Department of Psychiatry and Behavioral Science, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence
Victor J. Pokorny, Minneapolis Veterans Affairs Health Care System, 1 Veterans Drive, Minneapolis, MN 55417, USA.
Email: pokor076@umn.edu

Funding information
Health Services Research and Development, Grant/Award Number: I01CX000227; National Institute of Mental Health, Grant/Award Number: R24MH069675 and R03MH106831; National Science Foundation Graduate Research Fellowship, Grant/Award Number: #00074041

Abstract

Independent components analysis (ICA) is an effective and ubiquitous tool for cleaning EEG. To reduce computation time, many analysis pipelines decrease EEG dimensionality prior to ICA. A 2018 report by Artoni and colleagues detailed the deleterious effects of such reduced-dimensionality ICA (rdICA) on the dipolarity and reliability of independent components. Though valuable for researchers interested in directly analyzing independent components, ICA is more commonly used for cleaning EEG. Thus, a direct examination of the impact of artifact removal via rdICA on EEG data quality is needed. We conducted a registered analysis of 128 electrode recordings of 43 healthy subjects performing an active auditory oddball task. We preprocessed each subject's data under the following conditions: (1) ICA without dimension reduction, (2) ICA with only 64 electrodes included, (3) ICA preceded by PCA retaining 99% of the original data variance and (4) ICA preceded by PCA retaining 90% variance. We then quantified ERP data quality by measuring mean-amplitude, standardized measurement error (SME) of the single-trial mean-amplitudes, and split-half reliability of the N1 and P3 components. We then attempted to replicate our findings in an independent validation dataset. We observed statistically and practically significant changes in the mean amplitude of early sensory components for the 90% condition. Unexpectedly, the SME was only larger for the 64 electrode condition. Also unexpectedly, the effect of rdICA on split-half reliability was inconsistent between datasets. Based on the observed data, we argue that PCA-based rdICA is justifiable when used cautiously.

KEYWORDS
EEG, ERP, ICA

1 | INTRODUCTION

Independent components analysis (ICA) is a widely used and highly effective tool for cleaning electrophysiological recordings. A practical and theoretical limitation of ICA, in the context of electroencephalogram (EEG), is the property that the number of independent components (ICs) contained within the EEG is equal to the number of channels used for recording, which may or may not be equal to the true number of underlying signal sources. For this reason, it is common practice to reduce the dimensionality of high-density EEG data prior to ICA. A recent report (Artoni et al., 2018) detailed the deleterious effects of such dimension reduction on the dipolarity and reliability of resultant independent components (ICs). Though this report was valuable for researchers interested in directly analyzing brain ICs, many EEG researchers primarily use ICA to isolate and remove a small number of non-neural artifact components before quantifying neural responses of interest. Thus, there is a need for direct examination of the impact of reduced-dimensionality ICA (rdICA) artifact removal on EEG data quality.

To address this gap, we examined the impact of rdICA on the commonly studied N1 and P3 ERP components elicited during an active auditory oddball task. Oddball tasks consist of a series of intermixed infrequent target stimuli and frequent non-target stimuli. For active oddball tasks, participants are asked to make behavioral responses to differentiate targets and non-targets; however, oddball categorization-related neural responses are observed even in the absence of behavioral responses (Justen & Herbert, 2018). The auditory N1 or N100 component features a negative peak around 100 ms that is maximal at frontocentral electrode sites and is thought to reflect early sensory processing (Näätänen & Picton, 1987). It is weakly sensitive to oddballs, but highly sensitive to physical characteristics of a given stimulus including pitch and duration. It is also modulated by selective attention (Hillyard et al., 1973) and level of arousal (Nash & Williams, 1982). The later auditory P300 features a broad positive time course that peaks between 250 and 500 ms and is potentiated by oddballs (Picton, 1992). Unlike the N1 component, the P300 is less sensitive to physical characteristics of the stimuli and is therefore thought to reflect higher-level stimulus-categorization processes (Katayama & Polich, 1996). Previous work has estimated the internal consistency (as measured by Cronbach's alpha) of the N1 mean amplitude for standard and target stimuli as 0.94 and 0.86 respectively, and the P3 mean amplitude for target stimuli as 0.83 (Debener et al., 2002).

We chose the N1 and P300 as the primary components of interest because they are ubiquitously studied. Similarly, we chose to examine these components in the context of an active oddball task because it is a foundational and popular ERP paradigm. Moreover, it is not known whether early sensory components and late cognitive components are differentially affected by rdICA. To test whether our findings generalize to other EEG systems, task paradigms and ERP components, a validation analysis of an independent dataset was conducted in which we characterized the effect of rdICA on the visual P1 and N2 components elicited during a degraded stimulus continuous performance task (DSCPT) recorded using a different EEG recording system.

The current report focuses on two dimension reduction techniques: principal components analysis (PCA) and electrode subsetting, the latter of which approximates collecting lower-density EEG data (e.g., 64 or 32 channel montages). Other methods of dimension reduction techniques such as factor analytic and machine learning approaches exist (Tăuţan et al., 2021); however, these are not commonly used prior to ICA and thus were not investigated in the present manuscript.

Similar to ICA, PCA is ubiquitous across many fields; however, whereas ICA extracts maximally independent components, PCA identifies a sequence of orthogonal vectors that maximally account for variance in the observed data. Also similar to ICA, PCA will return as many components as there are dimensions in the data; however, it is common to set a threshold past which principal components are excluded. Methods for determining this threshold are vast, but for simplicity and alignment with Artoni et al.'s work, we focus on the percent variance thresholding procedure in which the percentage of variance accounted for by each PC is sequentially summed together (beginning with the PC that accounts for the most variance) until a threshold is reached (e.g., 90%). Once this threshold is met, all PCs that were not included in the summation are excluded.

Perhaps the simplest approach to dimension reduction of EEG is to submit a smaller subset of electrodes to ICA. We examine this approach because it is more computationally efficient than PCA: The time complexity of indexing an array is O(1), while the time complexity of creating a covariance/correlation matrix and then extracting orthogonal eigenvectors is O(min(p^3, n^3)) where p is the number of variables and n is the number of cases or observations (Johnstone & Lu, 2004). Subsetting electrodes prior to ICA is not only more computationally and conceptually parsimonious than PCA mediated rdICA, but may be a useful technique for assessing whether higher density electrode montages lead to better ICA-based artifact removal. Given the considerable debate within the EEG and ERP communities with respect to whether the costs of collecting high density recordings (e.g., longer recording preparation time, increased likelihood of bridging

electrodes) outweigh the putative benefits (e.g., higher signal to noise, better interpolation accuracy, better spatial resolution), characterizing how the quality of ICA artifact removal is impacted by the electrode density may be valuable.

To quantify the impact of rdICA, we examined subject-level mean amplitude, standardized measurement error, and split-half reliability. Mean amplitude is frequently used to quantify the magnitude of components because it is an unbiased point estimate that is robust to non-time-locked noise and unbalanced trial designs (Clayson et al., 2013). Thus we expected that subject-level mean amplitude measurements would not be meaningfully compromised by non-time-locked noise that might be added by rdICA. However, even if rdICA did not meaningfully impact mean amplitude, it might lead to different levels of single-trial noise in the ERP. To explicitly assess such single-trial level noise for a given time-window mean amplitude, we used the recently proposed standardized measurement error (SME; Luck et al., 2020). SME of a time-window mean-amplitude is computed by taking the mean amplitude of each trial, calculating the standard deviation of those single-trial mean amplitudes, and dividing by the square root of the number of trials. Thus if there is greater noise in the single-trial mean amplitude measurements (i.e., a larger standard deviation), the SME will be larger, indicating lower precision in single-trial mean amplitudes. Finally, much ERP research seeks to examine the relationship of brain potentials with individual differences (Dong et al., 2015; Marquardt et al., 2021), which requires reliable ERP measures (though see: Hedge et al., 2018). As discussed in Luck et al. (2020), classical reliability is sensitive to the degree of heterogeneity in a given sample such that greater heterogeneity leads to greater reliability estimates even when the precision of the measurement (e.g., SME) is held constant. As such, we assessed whether rdICA meaningfully impacts the split-half reliability of ERP component amplitudes.

Given the robustness of mean amplitude to single-trial level noise, we hypothesized that reducing dimensionality of 128 channel EEG data via channel subsetting or PCA prior to ICA would create negligible differences in subject-level mean amplitudes compared to full dimensionality ICA. We did, however, expect to see different levels of noise in the single-trial mean amplitude measurements, as measured via SME, with higher SME corresponding to electrode subsetting. Furthermore, we hypothesized that rdICA (via PCA, but not via channel subsetting) would result in negligible changes in the reliability of ERP amplitudes compared to fdICA, which would provide evidence in favor of the use of dimensionality reduction in individual differences research. By testing these hypotheses, we hoped to provide guidance for best practice with respect to removal of non-neural artifacts from high-dimensional ERP data. Ultimately, we hoped to show that the exorbitant computational costs of full dimensionality ICA (which may become a significant bottleneck in the scalability of preprocessing pipelines for large, multisite datasets) can be avoided by judicious use of dimension reduction techniques.

2 | METHODS

2.1 | Participants

Forty-three participants for the main analysis and 29 participants for the validation analysis were recruited as healthy controls as part of family studies of psychosis through the Minneapolis VA Medical Center, community mental health programs, and fliers posted throughout the community. Previous publications have reported on these samples to characterize atypical neural responses and self-report measures in patients with schizophrenia, bipolar disorder and first-degree biological relatives of patients with schizophrenia or bipolar disorder (Kang et al., 2019; Klein et al., 2020; Longenecker et al., 2020; Lynn et al., 2016; Olman et al., 2019; Pokorny et al., 2019, 2020; Schallmo et al., 2013; Van Voorhis et al., 2019). Exclusion criteria for healthy controls included intellectual disability (IQ <70), drug or alcohol dependence in past 6 months, current or past central nervous system condition, epilepsy, history of electroconvulsive therapy, history of head injury with skull fracture or loss of consciousness longer than 30 minutes, age under 18 or over 60, a history of psychotic disorder, current or past depressive or manic episodes, or family history of depression, mania, or psychotic disorder.

2.2 | Active auditory oddball task

The primary task of interest was an active directed-attention auditory oddball task administered using Neurobehavioral Systems' Presentation software on a Dell Computer running Windows XP. Tones were presented for 100 ms, with a 10 ms rise/fall time and a jittered intertrial interval of 1200–1500 ms. Participants completed a total of 800 trials over four blocks. In the first block, participants attended to the left ear and right-clicked a mouse upon hearing a high tone in their left ear. This high target tone had a pitch of 2400 Hz and was presented infrequently (10% of all trials). The low tone in the attended ear had a frequency of 1600 Hz and was presented frequently (40% of all trials). In the unattended ear, an infrequent tone (1200 Hz; 10% of all trials) and frequent tone (800 Hz;

40% of all trials) were presented. In the second block, the stimuli presentation was identical, but participants were instructed to attend to the right ear. Laterality of tones was reversed for blocks three and four (e.g., the tones presented to the right ear for the first two blocks were presented to the left ear for blocks three and four). Further information regarding the task can be found in a previous publication from our group (Force et al., 2008).

2.3 | EEG collection and processing

EEG data for the main analysis were collected using a BioSemi ActiveTwo system with a differential amplifier and a high-density 128 electrode montage and 10 auxiliary electrodes: two horizontal electrooculogram electrodes placed near the outer canthus of both eyes, two vertical electrooculogram electrodes placed above and below the right eye, two electrocardiogram electrodes placed on the left and right arms, two ear electrodes placed on the left and right ear lobe and two electromyogram electrodes placed on the left and right abductor pollicis brevis muscle just below the thumb. During acquisition, reference-free EEG were sampled at 1024 Hz with common mode sense (CMS) and driven right leg (DRL) ground electrodes.

Preprocessing was performed in EEGLAB v2021.1 (Delorme & Makeig, 2004). Data were referenced to average earlobe signal, high-pass filtered at 0.1 Hz using a 2nd order Butterworth (Tanner et al., 2015) and then resampled to 250 Hz (the MATLAB resample function implements a low-pass filter to prevent aliasing of frequencies above the Nyquist frequency during downsampling). Noisy time segments were automatically rejected using the eeglab pop_rejcont function with default values (frequency range = 20–40 Hz, epoch length = 0.5 s, upper threshold = 10db, epoch overlap = 0.25 s). Bad electrodes were automatically detected and rejected using pop_clean_rawdata with the following default parameters: FlatlineCriterion: 5, Channel Criterion: 0.8, Line Noise Criterion: 4 (note: ASR and additional removal of bad electrodes were turned off for efficiency because bad epochs should have been already detected with pop_rejcont).

Continuous EEG data were then submitted to four parallel ICA decompositions: (1) ICA without dimension reduction, (2) ICA with only odd numbered electrodes included (BioSemi electrodes are numbered A1-32, B1-32, C1-32 and D1-32 for a total of 128 electrodes; so we only included odd-numbered electrodes such as A1, A3, A5, A7, … D31 which provided full coverage of the scalp and approximated a 64 electrode montage), (3) ICA preceded by PCA retaining 99% of the original data variance and (4) ICA preceded by PCA retaining 90% of the variance. ICA was performed via the pop_runica function using the default infomax algorithm with the recommended 'extended' option enabled. We used the covariance-matrix based orthogonally-rotated PCA built into the function. The IClabel plug-in was used with default parameters to automatically identify ocular, muscular, cardiac channel noise and line-noise artifacts (Pion-Tonachini et al., 2019). Artifacts classified as ocular, muscular, cardiac, channel noise or line-noise with 80% or greater level of confidence were automatically rejected (Pernet et al., 2020). After ICA, removed channels were interpolated via the spherical spline approach.

Denoised data were epoched from −300 to 900 ms relative to stimulus onset and re-referenced to the average scalp electrode signal. The active directed-attention oddball task manipulates selective attention, stimulus rarity and relative pitch, but for simplicity we averaged across attention and pitch manipulations to focus on the more common oddball comparison of frequent vs. rare stimuli. Average ERPs were baseline corrected using the average value from −150 to 0 ms. Subjects were excluded if they had less than 25 (out of 160 total) usable infrequent trials. N1 components were defined a priori at the following frontocentral sites (F1, FZ, F2, FC1, FCz, FC2 & Cz) and P3 was defined at parieto-occipital sites (P1, Pz, P2, PO3, PO4 & POz). A priori time windows for mean amplitude extraction of the N1 were 50–150 ms (Näätänen & Picton, 1987) and P3 component was 250-500 ms (Comerchero & Polich, 1999; Polich, 2007; Simons et al., 2011). We visually inspected the grand-grand-average waveforms (averaged across participants and conditions) to determine whether these a priori time-windows provided reasonable coverage of the N1 and P3 component in our sample while avoiding biasing our results toward a particular within-subjects effect (i.e., to avoid double-dipping, Luck & Gaspelin, 2017).

2.4 | Planned statistical analyses

Six dependent variables were used to measure the impact of rdICA on ERP data quality in the main analysis: sensory N1 mean amplitude (averaged across frequent and rare conditions), frequent-rare P3 mean amplitude, standardized measurement error (SME) of the sensory N1 and frequent-rare P3 mean amplitudes, and split-half reliability of the sensory N1 and frequent-rare P3 mean amplitudes. For mean amplitude and SME measures, three one-tailed dependent samples t-tests were computed to compare: (1) full dimensionality ICA (fdICA) vs. 64 electrode ICA, (2) fdICA vs. 99% variance PCA before ICA and (3) fdICA vs. 90% variance PCA before ICA. To control for type-1 error rate, an initial alpha of 0.05 was Bonferroni corrected to 0.017 (0.05/3 = 0.017). Given the alpha and sample size, our statistical tests were able to detect a

medium Cohen's d effect size of 0.48 or greater with 80% probability (i.e., power = 0.8).

In addition to these frequentist tests, we computed Bayes factors to compare the strength of the evidence for the null and alternative hypotheses. The prior distribution of the null hypothesis was simply a *t*-distribution with *N*-1 degrees of freedom and the alternative hypothesis was a positive half of a Cauchy distribution (i.e., a *t*-distribution with one degree of freedom). This approach is recommended by (Rouder et al., 2009) and is the default for the BayesFactor R package which we used for computing the Bayes factors (Morey & Rouder, 2018).

Split half-reliability of the components were computed as the Pearson correlation between the average of even numbered trials and odd numbered trials across participants. This reliability coefficient was then Spearman-Brown corrected to account for the halving of the number of trials using the following formula:

$$\rho_{XX'} = \frac{2\rho_{12}}{1+\rho_{12}}$$

In which $\rho_{XX'}$ is the Spearman-Brown corrected reliability coefficient and ρ_{12} is the original split-half reliability coefficient. We then normalized $\rho_{XX'}$ via Fisher's *Z* transformation:

$$Z = 0.5ln\left(\frac{1+\rho_{XX'}}{1-\rho_{XX'}}\right)$$

and computed the 95% confidence interval with alpha set at 0.05/3 as:

$$Z \pm \frac{2.39398}{\sqrt{N-3}}$$

Thus a single reliability coefficient with corresponding 95% confidence intervals was estimated for each preprocessing condition.

Finally, to characterize whether rdICA might lead to a greater amount of mixed ICs which in turn leads to poorer ERP data quality, we calculated the percent variance explained by ICs that failed to be automatically classified as eye, brain, EKG, muscle, channel noise and line noise with 80% confidence for each participant for each preprocessing condition (note: there is another IClabel category named "other", but this was not used to determine whether an IC was mixed or not). We then used the same within-subject t-test approach described above to test for significant differences between conditions. If significant differences were found between preprocessing conditions in the amount of mixed ICs and amount of noise in the ERP components then we ran a mediation analysis (using the mediation R package with non-parametric bootstrapped confidence intervals estimated via 10,000 simulation iterations; Tingley et al., 2014) to test whether increases in mixed ICs explain the relationship between rdICA and decreased data quality.

2.5 | Validation dataset

To test the generalizability of our findings, we conducted the same set of analyses on an independent dataset of 29 healthy controls that completed a degraded stimulus continuous performance task (DS-CPT) with EEG recorded on a BrainVision 128 electrode ActiCHamp system (Klein et al., 2020; Sponheim et al., 2006). The DSCPT presents a series of numbers from 0 to 9 (stimulus duration = 29 ms and SOA = 1000 ms) and requires participants to press a button when the target number "0" appears. We inspected changes in mean amplitude, SME and split-half reliability for the bilateral visual P1 and the target elicited N2. The P1 mean amplitude was measured with a time window of 80–130 ms at electrode sites PO7/PO8/P7/P8 averaged across all conditions (Mangun, 1995; Sponheim et al., 2006). Visual N2 mean amplitude was measured via target-nontarget difference waveforms averaged across electrodes Cz, FCz, and Fz with a time-window of 320–400 ms (see Klein et al., 2020) for more details regarding the dataset and paradigm.

3 | RESULTS

3.1 | Preprocessing results

After preprocessing, zero individuals in the main dataset were excluded due to low trial counts (<25 trials in the rare/target condition) while three individuals in the validation dataset were excluded for this reason. Because time segment rejection was performed prior to ICA, the number of trials per subject per condition were exactly matched (mean rare trials for main dataset: 139.9, mean frequent trials for main dataset: 556.3, mean rare trials for validation dataset: 83.6, mean frequent trials for validation dataset: 371.7). For the 99% preprocessing condition in the main dataset, one individual had zero ICs left after automatic rejection and was thus excluded. For the 90% condition, 5 individuals were excluded for this reason. For the validation dataset, no individuals were excluded in this way.

3.2 | Planned analyses: Mean amplitude

Based on visual inspection of the grand-grand average waveform, we adjusted the time-windows for the N1

from 50–150 ms to 50–180 ms while the P1, N2 and P3 time-windows were not adjusted (see Figure 1 for grand average topographies and waveforms of N1 and P3 components; Figure S1 shows the corresponding plots for P1 and N2). Figure 2 depicts the degree of change in N1 and P3 mean amplitude after preprocessing under the four conditions of interest. For the N1 component, we did not observe a significant change in mean amplitude for the full vs half and full vs. 99% comparisons. We did, however, see a significant increase in mean amplitude for the 90% variance thresholded condition. Given that the N1 is a negative-going component, this increase in amplitude is reflective of the N1 amplitudes on average shifting closer to zero (i.e., greater noise in the denominator relative to the negative-going signal in the numerator). We observed the largest Bayes factor for this full vs. 90% comparison suggesting that the data strongly supported the alternative hypothesis (i.e., that there was a difference in mean amplitude between preprocessing conditions). The P3 difference wave did not significantly differ for any of the three condition comparisons we performed. Across components, four of the six Bayes factors were below 0.33 suggesting that the data were more supportive of the null hypothesis for these four comparisons (i.e., that there was no significant difference between these preprocessing conditions). Interestingly, the Bayes factor for the full vs. half comparison of P3 amplitude was larger than 3 suggesting that the data were more consistent with the alternative hypothesis; however, the change in mean amplitude was not in the expected direction. In other words, the P3 amplitudes were *larger* in the reduced dimensionality condition which could be indicative of an increased signal to noise ratio because the P3-diff is a positive-going component.

3.3 | Exploratory analysis of mean amplitude normality and outliers

Visual inspection of the mean amplitude distributions showed that the reduced dimensionality conditions (especially the 90% condition) led to differences in shape of the mean amplitude distribution and number of outliers despite the mean of each distribution being relatively stable. To assess whether rdICA led to more dramatic departures from normality, we conducted exploratory Anderson-Darling tests for each condition. The results of

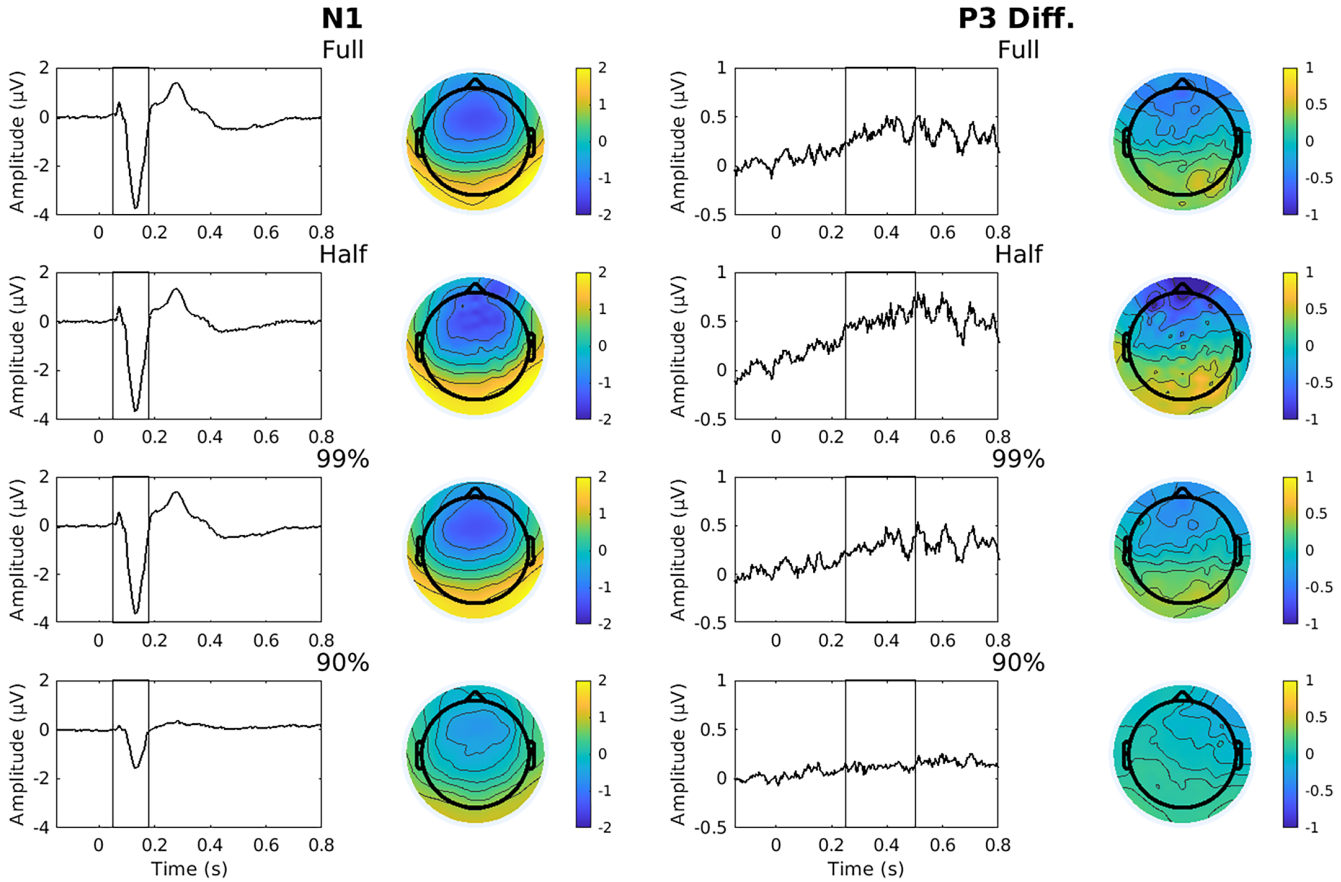

FIGURE 1 Grand average waveforms and topographies by condition. The N1 component is averaged across all trials while the P3 is a difference wave of rare minus frequent trials. To mitigate the effect of outliers, a robust biweight approach was used for grand averaging (Hoaglin et al., 1983).

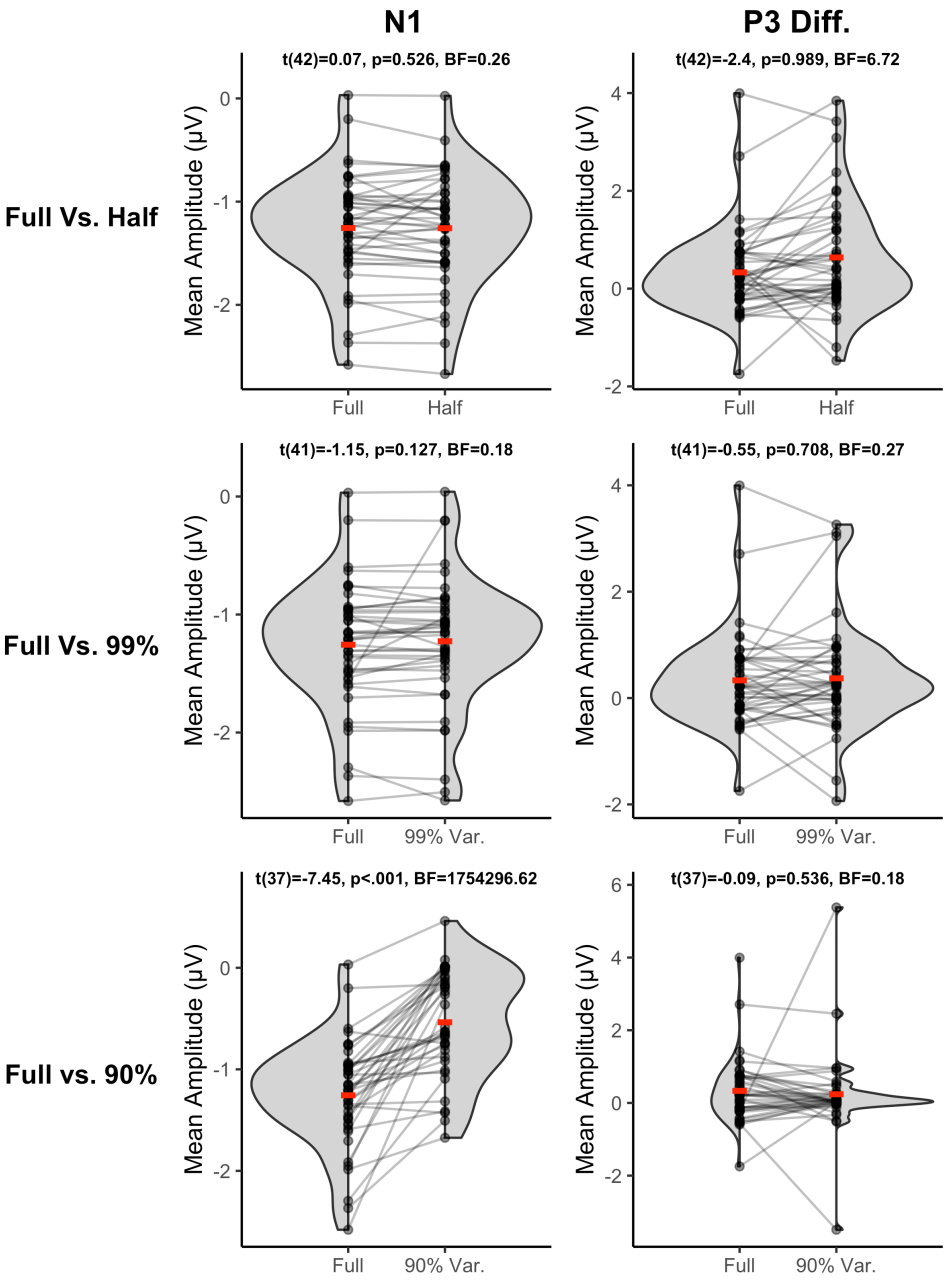

FIGURE 2 Effect of reduced dimensionality ICA on N1 and P3 mean amplitudes. Gray dots represent each individual and gray lines connect each individual across conditions. Red lines represent the mean. BF is the Bayes factor in which the alternative hypothesis is in the numerator and the null hypothesis is in the denominator. Full: Full dimensionality ICA; half: Odd electrodes only submitted to ICA; 99%: ICA preceded by PCA retaining 99% of the variance; 90%: Same as 99% except retaining only 90% of the variance.

this analysis, which can be found in Table S2, were consistent with rdICA leading to deviations from normality for the P3 component with the 90% condition leading to the largest A statistic. The N1 mean amplitude distributions, on the other hand, did not significantly deviate from normality for any condition, after correction for multiple comparisons.

To assess for outliers, we used a Generalized Extreme Studentized Deviate procedure (Rosner, 1983) using the gesdTest() function from the PMCMRplus package (Pohlert, 2022). The results of this procedure (see Table S3) generally corroborated what is evident by visual inspection of Figure 2: PCA-based rdICA led to more outliers for the P3 component, but not for the N1.

3.4 | Planned analyses: SME

Figure 3 depicts the same set of comparisons as Figure 2 except with the standardized measurement error (SME)

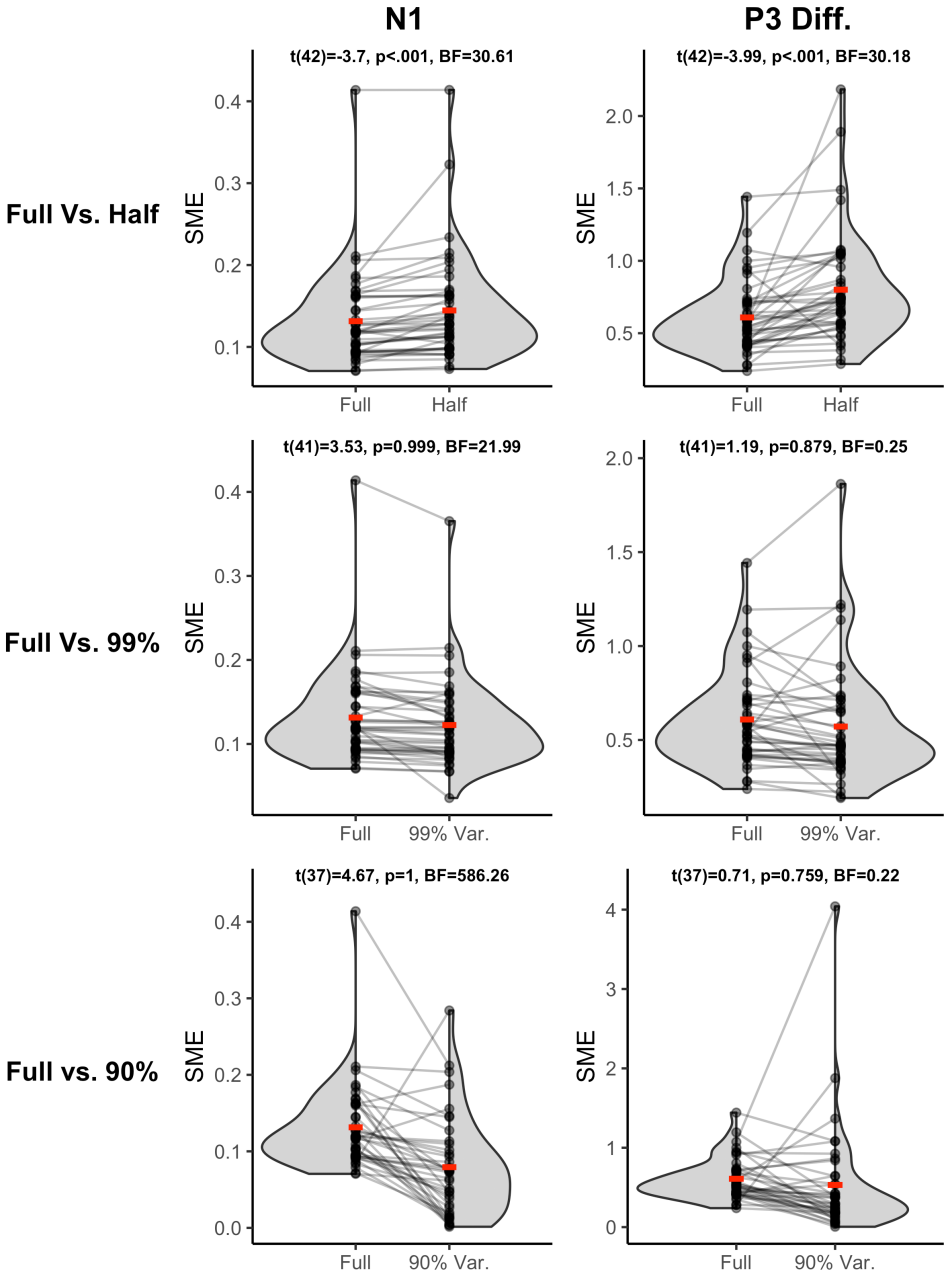

FIGURE 3 Effect of reduced dimensionality ICA on the standardized measurement error of N1 and P3. Standardized measurement error (SME) was calculated as the standard deviation of a given set of single-trial mean amplitudes divided by the square root of the number of trials. Larger SME values are interpreted as indicating more noise because a high SME value implies a large spread of single trial mean amplitudes when the number of trials are held constant. See Figure 2 for further figure description.

of the single trial mean amplitudes as the dependent variable. Only the half dimensionality ICA led to the expected increase in SME where as the 99% and 90% conditions led to an apparent *decrease* in SME (note: we could not test for significance of this decrease because we specified a priori one-tailed t-tests in the opposite direction). Bayes factors strongly supported the alternative hypothesis for the N1 SME in all condition comparisons; however, only the full vs. half difference was in the expected direction (i.e., reduced ICA led to higher SME). For the P3 component, Bayes factors for the full vs. 99% and full vs. 90% comparisons suggested that the data were most consistent with the null hypothesis: that P3 SME did not meaningfully change for the 99% or 90% conditions relative to full dimensionality.

3.5 | Planned analyses: Reliability

Figure 4 depicts the raw and Spearman-Brown corrected split-half reliabilities of the N1 and P3 components. In

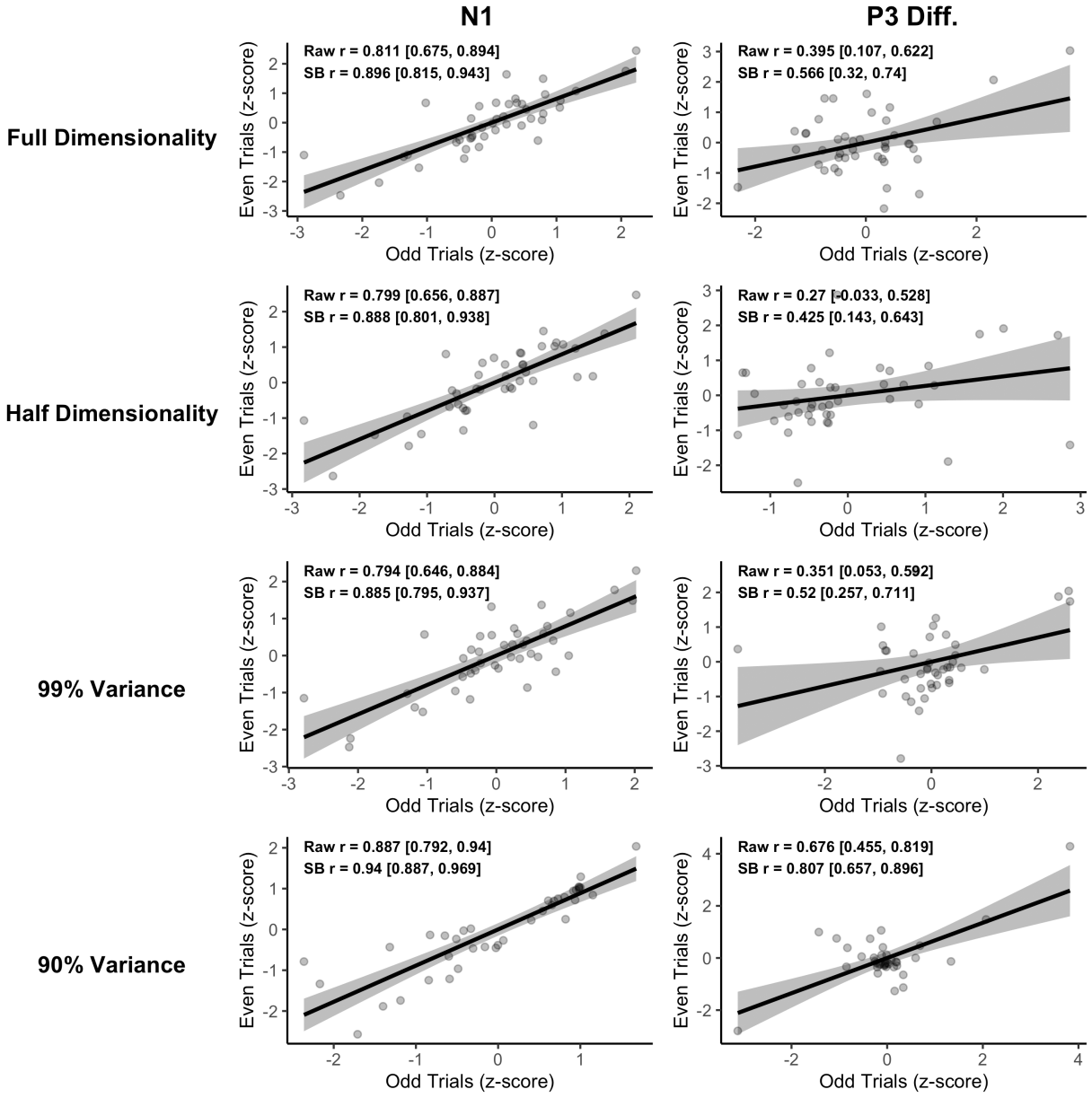

FIGURE 4 Effect of reduced dimensionality ICA on the reliability of N1 and P3. 'Raw r' is the uncorrected correlation coefficient between even and odd trials. 'SB r' is spearman-Brown corrected coefficient that estimates the reliability of the component after accounting for halving the number of trials. Numbers in brackets are 95% confidence intervals.

general, N1 reliability was greater than the P3 difference wave reliability which is consistent with the known deleterious effect of subtraction scores on reliability. Surprisingly, the 90% condition led to the most reliable N1 and P3 components. For the P3 difference score, the reliability coefficients from the full, half and 99% conditions fell outside of the 95% confidence interval of the 90% condition coefficient such that the reliability of the 90% condition was significantly higher than the other conditions. However, as can be seen in the scatterplots, the 90% condition created more heterogeneity between subjects and generated outlying data points which may have spuriously inflated reliability.

3.6 | Planned analyses: Mixed ICs

To clarify whether rdICA led to a higher proportion of mixed ICs, Figure 5 compares the percent variance accounted for (PVAF) of ICs that were labeled with low confidence by IClabel (ignoring the confidence of the "other" label classification). In general, full dimensionality ICA was associated with smaller proportions of mixed ICs as compared to the other conditions, though the difference was much greater for the 99% and 90% condition.

We then conducted follow-up mediation analyses to test whether reductions in ERP data quality associated

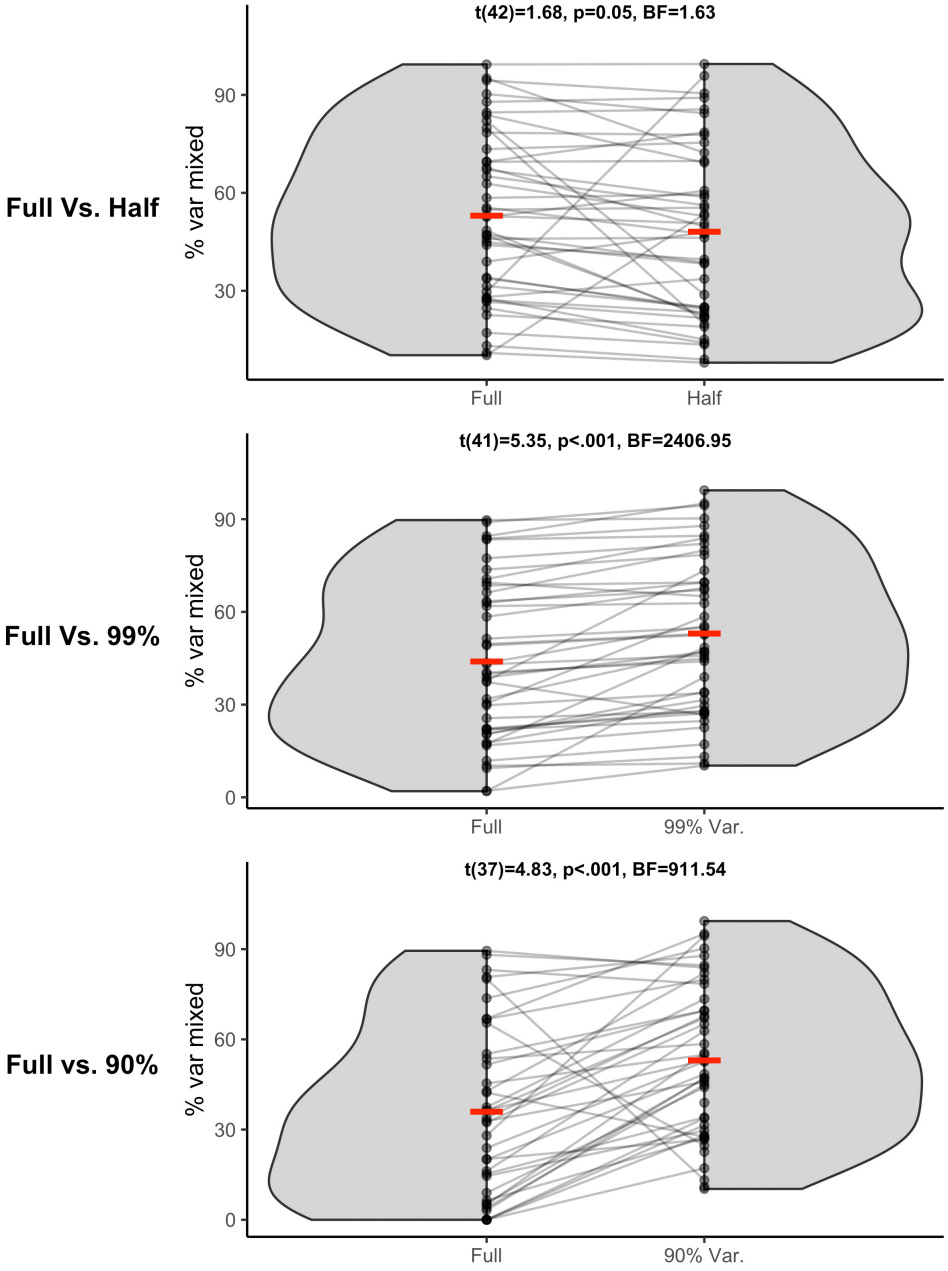

FIGURE 5 Comparison of the proportion of mixed ICs for rdICA. ICs were classified as mixed if they could not be labeled as eye, brain, EKG, muscle, channel noise or line noise with greater than 80% confidence by the automatic IC classifier IClabel. See Figure 2 for further figure description.

TABLE 1 Summary of ACME results

Component	Dependent variable	Full vs. half	Full vs. 99%	Full vs. 90%
N1	Mean Amp.	N/A	N/A	$\beta = -0.06$, 95% CI [−0.15, 0.01], $p = .13$
	SME	$\beta = 0$, 95% CI [0, 0], $p = .94$	N/A	N/A
P3 Diff.	Mean Amp.	N/A	N/A	N/A
	SME	$\beta = -0.02$, 95% CI [−0.06, 0.02], $p = .4$	N/A	N/A

Note: N/A denotes that the mediation analysis was not conducted because there was not a significant effect of condition on the dependent variable in the expected direction. Note that we only observed significant changes in data quality in the expected direction for 3 out of 12 possible comparisons.

Abbreviation: ACME, average causal mediation effect.

with reduced dimensionality ICA were due to higher proportions of mixed ICs. We did not observe strong evidence that the proportion of mixed ICs mediated the relationship between rdICA and data quality (see Table 1). This suggests that, while the proportion of mixed ICs *was* higher for reduced dimensionality ICA solutions, this increase did not compellingly account for downstream reductions in ERP data quality.

3.7 | Agreement between primary and validation results

Figures S2–S5 and Table S1 depict the same set of comparisons and statistical tests for the P1 and N2 components elicited during a visual target detection task in an independent sample of healthy controls. For mean amplitude, both the main dataset and validation dataset showed a significant attenuation of mean amplitude for the earlier, sensory components (i.e., N1 for main analysis, P1 for validation analysis) for the 90% relative to full dimensionality. Other than this decrement for the 90% condition, we did not observe any other significant differences in mean amplitude in the expected direction for either the main or validation datasets.

For SME, we observed an effect of greater SME for half dimensionality relative to full dimensionality for both P1 and N2 components in the validation dataset which is also consistent with the primary analysis. One important difference between datasets; however, is that a single large outlier (>3 SDs from mean) was evident in the 99% and 90% conditions in the validation dataset only. Interestingly, this outlier was not influential enough to create a significant difference in SME between full vs. 99% or 90%. This lack of difference is also consistent with the primary analysis.

The reliability coefficients for the validation dataset were also affected by a single large outlier for the 99% and 90% conditions which led to poor reliability for P1 component, but not the N2 component. Whereas for the primary analysis the 90% condition exhibited the highest reliability coefficients for both components, the 90% condition in the validation dataset exhibited the poorest reliability for the P1 component (though again this is likely primarily due to the highly influential outlier). The N2 component also was influenced by this outlier; however in this case the reliability coefficient was likely overestimated.

Similar to the primary analysis, the 90% and 99% conditions were associated with a substantial increase in the proportion of mixed ICs in the validation dataset. In both datasets, the half dimensionality condition was not associated with an increase in proportion of mixed ICs. Finally the lack of a mediating effect of proportion of mixed ICs observed in the primary analysis was again observed in the validation dataset.

4 | DISCUSSION

4.1 | Summary

The present study sought to examine the effect of reduced dimensionality ICA on downstream data quality for common ERP components. We observed statistically and practically significant changes in the mean amplitude of early sensory components (P1 and N1) for the most drastic dimension reduction condition (90% variance thresholded PCA), but did not observe such differences for the other conditions or later, cognitive components (N2 and P3). The standardized measurement error (a proposed universal metric of data quality), was larger for all components in the half dimensionality condition, but not in the 99% or 90% conditions. The effect of rdICA on split-half reliability was less clear: In the primary analysis, the split-half reliability was largest for the 90% condition for the N1 and P3, while, for the validation dataset, the reliability was lowest for the 90% condition at P1, but then highest again for the N2 component. Finally, the proportion of mixed ICs was generally larger for the reduced dimensionality conditions, but this increase did not explain changes in downstream data quality.

4.2 | Best practices for reducing dimensionality prior to ICA

A primary aim of this work was to establish recommendations for the judicious use of reduced dimensionality ICA. The following recommendations are based on the data we observed. Of course, the appropriateness of these recommendations will vary from researcher to researcher, lab to lab and content area to content area.

4.2.1 | Recommendation #1: The threshold for a PCA thresholded rdICA should generally be at 99% or higher

One of the more consistent effects we observed across both datasets was the shift of mean amplitudes toward zero for the 90% condition in earlier sensory components. In general, the 90% PCA seemed to underestimate the number of components in a given recording (see Figure S6). On the other hand, the 99% thresholded mean amplitudes, across both datasets, did not significantly differ from the full dimensionality mean amplitudes and the Bayes factors were

below 0.33 suggesting that the data were more consistent with the null hypothesis (i.e., no difference between conditions). Importantly, 99% variance thresholded PCA-based rdICA was not associated with increases in SME in either datasets such that single trial data quality does not appear to be meaningfully affected by the PCA-based dimension reduction procedure. Thus, the 99% percent threshold was generally acceptable for preserving data quality. However, even for this more conservative 99% condition, there were two individuals in the primary dataset and one individual in the validation dataset for which only a single component was submitted to ICA (again see Figure S6). This suggests that even the 99% condition likely underestimates the number of components in a given recording. Future work would benefit from looking at even more conservative thresholds (e.g., 99.5% or 99.9%).

4.2.2 | Recommendation #2: PCA thresholded rdICA should generally be used in situations in which scalability is a priority

One might reasonably ask, if one needs to preserve at least 99% (or more) of the variance, can we really expect ICA to run much faster? For our datasets, the answer seems to be a resounding yes! In our main analysis, the full dimensionality ICA was 4.6 times slower than 99% rdICA (see Table S4). Even more importantly, from a scalability perspective, the ratio of the maximum ICA run time for full ICA relative to 99% was 15.57 (i.e., the maximum run time for the full condition was 15x greater than the maximum for 99%). Similar ratios were observed in the validation dataset as well. Thus, for analyses in which scalability and processing speed are a priority, a PCA thresholded ICA may be valuable as long as the threshold is kept at or above 99%.

4.2.3 | Recommendation #3: Subsetting to a smaller number of electrodes leads to poorer data quality

Across both data sets, we observed a consistent increase in SME when dimensionality was decreased by subsetting electrodes prior to ICA. Thus, we cannot recommend increasing computation efficiency via electrode subsetting. This finding also speaks to the possible benefits of recording from higher density montages. The electrode subsetting condition can be thought of as an approximated 64 channel montage recording. Given that the sparser recording led to poorer data quality, an argument can be made that, everything else being equal, higher density montages lead to improved data quality of ERP components. Having said this, as the number of electrodes increases, the electrode preparation becomes more difficult and likelihood of "bridging" electrodes increases such that in a "true" 64 electrode recording, more care may have gone into abrading the scalp and gelling the electrodes. Thus, it is possible that any difference in SME between higher and lower density systems may be washed out by the improvement in SME from more careful electrode preparation.

4.2.4 | Recommendation #4: If you choose to use a PCA-based rdICA strategy, monitor for outliers

We observed a single large outlier and a few smaller mean amplitude outliers created by both the 99% and 90% rdICAs. Given the low base rate of this effect (four for 90% and one for 99%; see Table S4), it is unclear whether this is a common issue with PCA-based rdICA; however, further work may be helpful for elucidating how and why PCA thresholded rdICA might generate outlying data points. Thus, we recommend that if one chooses a PCA-based reduction strategy close monitoring for outliers is essential.

4.3 | Limitations and future directions

Although we attempted to maximize external validity by including an independent validation dataset recorded on a different EEG system with different task demands and distinct underlying cognitive processes, replication of our findings in other datasets will be crucial. Furthermore, although the 99% PCA condition appeared to best balance computational efficiency and data quality in our analyses; there may be more sophisticated methods for optimizing this trade-off between efficiency and data quality. For example, information theoretic approaches such as Bayesian Information Criterion (BIC; Schwarz, 1978) provide an alternative to selecting the number of components based on the explained variance, and have previously been applied in the context of scalp EEG recordings (Kovacevic & McIntosh, 2007). Future work might consider whether information theoretic approaches for selecting the number of components outperform component selection based on explained variance. Other methods have been developed for use with neural data other than EEG, such as targeted dimensionality reduction (TDR; Cohen & Maunsell, 2010; Cunningham & Yu, 2014; Mante et al., 2013) via multinomial logistic reduction. TDR attempts to reduce the data to a small number of dimensions that are maximally related to some task variable of interest. While this method might be useful in the context of scalp EEG

where there is a specific target variable of interest, to our knowledge this approach has not been applied to scalp EEG. Unsupervised dimension reduction methods have also been applied to scalp EEG, most notably in the case of k-means clustering for EEG microstate analysis (Khanna et al., 2015); however, the application of unsupervised clustering approaches for dimension reduction prior to ICA is unclear.

Finally, we cannot, in the present study, speak to the effect of rdICA on the split-half reliability of ERP components due the conflicting trends observed between the primary and validation datasets. Given the continued interest in understanding the relationship between individual differences in ERPs and psychological constructs, clarifying the discrepancy in reliabilities reported here may be important for further justifying the use of rdICA.

4.4 | Deviations from registered analysis plans

Great care was taken to follow the registered analysis plans as closely as possible; however, there were a handful of deviations from this plan which we believe improved the quality of the paper. First, we decided to use the pop_clean_rawdata function in place of pop_rejchan because pop_rejchan was not detecting obvious bad electrodes prior to ICA. We do not believe this biased our results toward a given condition because electrode rejection occurred prior to ICA such that all conditions were fed the exact same data. Furthermore, pop_clean_rawdata is a commonly used function for detecting bad electrodes which means this change did not sacrifice generalizability of our results. Also, we initially specified including 40 subjects in the main analysis, however, we discovered three additional subjects during our analyses and decided to include them in our sample to increase power.

4.5 | Conclusion: When is rdICA justified?

Based on the observed data, we argue that PCA-based rdICA is justifiable when used sparingly. In fact, it is often common practice to conduct a PCA prior to ICA when there is linear dependency in a given EEG data matrix. Such linear dependency means that at least one of the channels in the data matrix is not providing any independent information (this is especially common when the gel from two neighboring channels bridge together). Fortunately, the rank of a data matrix can be algebraically determined and PCA can then be used to reduce the matrix to a full rank state. Thus, reducing the dimensionality of a data matrix prior to ICA in this context is not controversial and, in fact, is a common procedure for many ICA pipelines!

Instead, it would seem the crux of the issue is: *by how much* should we reduce the dimensionality of a given data matrix? Our findings align with those of Artoni et al. that PCA thresholds below 99% tend to reduce the dimensionality too aggressively such that important information is lost and/or ICA does not have sufficient degrees of freedom to separate source signals effectively. Having said this, our results indicate that the 99% threshold PCA was less prone to "over-reducing" and the downstream data quality did not significantly suffer as compared to full dimensionality. Thus, we argue that as long as researchers are aware of possible pitfalls of reducing dimensionality (e.g., "over-reducing", outlier generation, possible departures from normality, etc.) rdICA may not only be justified, but may be crucial for enhancing the rate of progress and broadening the scope of psychophysiological research.

AUTHOR CONTRIBUTIONS

Victor Pokorny: Conceptualization; formal analysis; methodology; writing – original draft; writing – review and editing. **Scott R. Sponheim:** Funding acquisition; investigation; methodology; project administration; resources; software; supervision; writing – review and editing. **Eric Rawls:** Conceptualization; methodology; writing – review and editing.

ACKNOWLEDGMENTS

This work is supported by grants from the Veterans Health Administration (I01CX000227), National Institutes of Mental Health (R24MH069675, R03MH106831) to SRS. ER is supported by a CTSI-TL1 postdoctoral fellowship. The University of Minnesota's CTSI is supported through the NIH Clinical and Translational Science Award (CTSA) program, grants TL1TR002493 and UL1TR002494. VJP received support from the National Science Foundation (Graduate Research Fellowship Program: #00074041).

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

DATA AVAILABILITY STATEMENT

Matlab code for preprocessing EEG and R code for subsequent statistical analysis can be found at https://github.com/vpokorny123/Psychophys-rdICA. Additionally, the raw EEG files that comprise both the main and validation datasets can be found at https://osf.io/jrb3q/.

ORCID

Victor J. Pokorny https://orcid.org/0000-0002-1003-0196

REFERENCES

Artoni, F., Delorme, A., & Makeig, S. (2018). Applying dimension reduction to EEG data by principal component analysis reduces the quality of its subsequent independent component decomposition. *NeuroImage*, *175*, 176–187. https://doi.org/10.1016/j.neuroimage.2018.03.016

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, *50*(2), 174–186. https://doi.org/10.1111/psyp.12001

Cohen, M. R., & Maunsell, J. H. R. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(45), 15241–15253. https://doi.org/10.1523/JNEUROSCI.2171-10.2010

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, *110*(1), 24–30. https://doi.org/10.1016/s0168-5597(98)00033-1

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. https://doi.org/10.1038/nn.3776

Debener, S., Kranczioch, C., Herrmann, C. S., & Engel, A. K. (2002). Auditory novelty oddball allows reliable distinction of top-down and bottom-up processes of attention. *International Journal of Psychophysiology*, *46*(1), 77–84. https://doi.org/10.1016/s0167-8760(02)00072-7

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dong, S., Reder, L. M., Yao, Y., Liu, Y., & Chen, F. (2015). Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. *Brain Research*, *1616*, 146–156. https://doi.org/10.1016/j.brainres.2015.05.003

Force, R. B., Venables, N. C., & Sponheim, S. R. (2008). An auditory processing abnormality specific to liability for schizophrenia. *Schizophrenia Research*, *103*(1-3), 298–310. https://doi.org/10.1016/j.schres.2008.04.038

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). Understanding robust and exploratory data analysis. https://ui.adsabs.harvard.edu/abs/1983ured.book.....H

Johnstone, I. M., & Lu, A. Y. (2004). Sparse principal components analysis. https://imjohnstone.su.domains//WEBLIST/AsYetUnpub/sparse.pdf

Justen, C., & Herbert, C. (2018). The spatio-temporal dynamics of deviance and target detection in the passive and active auditory oddball paradigm: A sLORETA study. *BMC Neuroscience*, *19*(1), 25. https://doi.org/10.1186/s12868-018-0422-3

Kang, S. S., MacDonald, A. W., & Sponheim, S. R. (2019). Dysfunctional neural processes underlying context processing deficits in schizophrenia. Biological Psychiatry. *Cognitive Neuroscience and Neuroimaging*, *4*(7), 644–654. https://doi.org/10.1016/j.bpsc.2019.03.012

Katayama, J., & Polich, J. (1996). P300, probability, and the three-tone paradigm. *Electroencephalography and Clinical Neurophysiology*, *100*(6), 555–562. https://doi.org/10.1016/s0168-5597(96)95171-0

Khanna, A., Pascual-Leone, A., Michel, C. M., & Farzan, F. (2015). Microstates in resting-state EEG: Current status and future directions. *Neuroscience and Biobehavioral Reviews*, *49*, 105–113. https://doi.org/10.1016/j.neubiorev.2014.12.010

Klein, S. D., Shekels, L. L., McGuire, K. A., & Sponheim, S. R. (2020). Neural anomalies during vigilance in schizophrenia: Diagnostic specificity and genetic associations. *NeuroImage: Clinical*, *28*, 102414. https://doi.org/10.1016/j.nicl.2020.102414

Kovacevic, N., & McIntosh, A. R. (2007). Groupwise independent component decomposition of EEG data and partial least square analysis. *NeuroImage*, *35*(3), 1103–1112. https://doi.org/10.1016/j.neuroimage.2007.01.016

Longenecker, J. M., Krueger, R. F., & Sponheim, S. R. (2020). Personality traits across the psychosis spectrum: A hierarchical taxonomy of psychopathology conceptualization of clinical symptomatology. *Personality and Mental Health*, *14*(1), 88–105. https://doi.org/10.1002/pmh.1448

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luck, S. J., Stewart, A. X., Simmons, A. M., & Rhemtulla, M. (2020). Standardized measurement error: A universal measure of data quality for averaged event-related potentials (v_r1). https://doi.org/10.31234/osf.io/dwm64

Lynn, P. A., Kang, S. S., & Sponheim, S. R. (2016). Impaired retrieval processes evident during visual working memory in schizophrenia. *Schizophrenia Research: Cognition*, *5*, 47. https://doi.org/10.1016/j.scog.2016.07.002

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Marquardt, C. A., Pokorny, V. J., Kang, S. S., Cuthbert, B. N., & Sponheim, S. R. (2021). Posttraumatic stress symptom dimensions and brain responses to startling auditory stimuli in combat veterans. *Journal of Abnormal Psychology*, *130*(5), 455–467. https://doi.org/10.1037/abn0000552

Morey, R. D., & Rouder, J. N. (2018). Bayes factor: Computation of Bayes factors for common designs. https://CRAN.R-project.org/package=BayesFactor

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Nash, A. J., & Williams, C. S. (1982). Effects of preparatory set and task demands on auditory event-related potentials. *Biological Psychology*, *15*(1-2), 15–31. https://doi.org/10.1016/0301-0511(82)90028-x

Olman, C. A., Espensen-Sturges, T., Muscanto, I., Longenecker, J. M., Burton, P. C., Grant, A. N., & Sponheim, S. R. (2019). Fragmented ambiguous objects: Stimuli with stable low-level features for object recognition tasks. *PLoS One*, *14*(4), e0215306. https://doi.org/10.1371/journal.pone.0215306

Pernet, C. R., Martinez-Cancino, R., Truong, D., Makeig, S., & Delorme, A. (2020). From BIDS-formatted EEG data to sensor-space group results: A fully reproducible workflow with EEGLAB and LIMO EEG. *Frontiers in Neuroscience*, *14*, 610388. https://doi.org/10.3389/fnins.2020.610388

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*(4), 456–479. https://doi.org/10.1097/00004691-199210000-00002

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Pohlert, T. (2022). PMCMRplus: Calculate pairwise multiple comparisons of mean rank sums extended. https://CRAN.R-project.org/package=PMCMRplus

Pokorny, V. J., Espensen-Sturges, T. D., Burton, P. C., Sponheim, S. R., & Olman, C. A. (2020). Aberrant cortical connectivity during ambiguous object recognition is associated with schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*, 1193–1201. https://doi.org/10.1016/j.bpsc.2020.09.018

Pokorny, V. J., Lano, T. J., Schallmo, M.-P., Olman, C. A., & Sponheim, S. R. (2019). Reduced influence of perceptual context in schizophrenia: Behavioral and neurophysiological evidence. *Psychological Medicine*, *51*(786), 794. https://doi.org/10.1017/S0033291719003751

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, *25*(2), 165–172. https://doi.org/10.1080/00401706.1983.10487848

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Schallmo, M.-P., Sponheim, S. R., & Olman, C. A. (2013). Abnormal contextual modulation of visual contour detection in patients with schizophrenia. *PLoS One*, *8*(6), e68090. https://doi.org/10.1371/journal.pone.0068090

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464 http://www.jstor.org/stable/2958889

Simons, C. J. P., Sambeth, A., Krabbendam, L., Pfeifer, S., van Os, J., & Riedel, W. J. (2011). Auditory P300 and N100 components as intermediate phenotypes for psychotic disorder: Familial liability and reliability. *Clinical Neurophysiology*, *122*(10), 1984–1990. https://doi.org/10.1016/j.clinph.2011.02.033

Sponheim, S. R., McGuire, K. A., & Stanwyck, J. J. (2006). Neural anomalies during sustained attention in first-degree biological relatives of schizophrenia patients. *Biological Psychiatry*, *60*(3), 242–252. https://doi.org/10.1016/j.biopsych.2005.11.017

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Tăuţan, A.-M., Rossi, A. C., de Francisco, R., & Ionescu, B. (2021). Dimensionality reduction for EEG-based sleep stage detection: Comparison of autoencoders, principal component analysis and factor analysis. *Biomedizinische Technik/Biomedical Engineering*, *66*(2), 125–136. https://doi.org/10.1515/bmt-2020-0139

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38 http://www.jstatsoft.org/v59/i05/

Van Voorhis, A. C., Kent, J. S., Kang, S. S., Goghari, V. M., MacDonald, A. W., 3rd, & Sponheim, S. R. (2019). Abnormal neural functions associated with motor inhibition deficits in schizophrenia and bipolar disorder. *Human Brain Mapping*, *40*(18), 5397–5411. https://doi.org/10.1002/hbm.24780

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1 Grand average waveforms and topographies by condition in validation dataset. The P1 component is averaged across all trials while the N2 is a difference wave of the target minus nontarget trials. To mitigate the effect of outliers, a robust biweight approach was used for grand averaging (Hoaglin et al., 1983). Note that the N2 difference component was not clearly recognizable in the grand average waveforms; however, we report on the data quality of this component in accordance with the registered analysis plans. It is possible that the number of trials or number of individuals in this sample was not sufficiently large to successfully characterize the component.

Figure S2 Effect of reduced dimensionality ICA on P1 and N2 mean amplitudes in validation dataset. Gray dots represent each individual and gray lines connect each individual across conditions. Red lines represent the mean. BF is the Bayes factor in which the alternative hypothesis is in the numerator and the null hypothesis is in the denominator. Full: full dimensionality ICA; Half: odd electrodes only submitted to ICA; 99%: ICA preceded by PCA retaining 99% of the variance; 90%: same as 99% except retaining only 90% of the variance.

Figure S3 Effect of reduced dimensionality ICA on the standardized measurement error of P1 and N2 in validation dataset. Standardized measurement error (SME) was calculated as the standard deviation of a given set of single-trial mean amplitudes divided by the square root of the number of trials. Larger SME values are interpreted as indicating more noise because a high SME value implies a large spread of single trial mean amplitudes when the number of trials are held constant. See Figure 2 for further description.

Figure S4 Effect of reduced dimensionality ICA on the reliability of P1 and N2 in validation dataset. 'Raw r' is the

uncorrected correlation coefficient between even and odd trials. 'SB r' is Spearman-Brown corrected coefficient that estimates the reliability of the component after accounting for halving the number of trials. Numbers in brackets are 95% confidence intervals.

Figure S5 Comparison of the proportion of mixed ICs for rdICA. ICs were classified as mixed if they could not be labeled as eye, brain, EKG, muscle, channel noise or line noise with greater than 80% confidence by the automatic IC classifier IClabel. See Figure 2 for further figure description.

Figure S6 Number of principal components after thresholding. Dashed line represents the mean for each condition. Numbers next to dashed lines indicate mean and standard deviation.

Table S1 Summary of ACME results in validation dataset

Table S2 Summary of Anderson-Darling tests of normality

Table S3 Generalized extreme studentized

Table S4 ICA run times relative to full ICA

How to cite this article: Pokorny, V. J., Sponheim, S. R., & Rawls, E. (2023). Impact of reduced-dimensionality independent components analysis on event-related potential measurements. *Psychophysiology*, *60*, e14223. https://doi.org/10.1111/psyp.14223