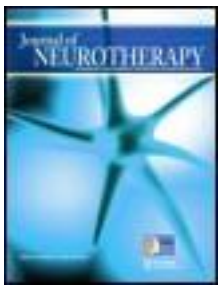


Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience

Validity and Reliability of Quantitative Electroencephalography

Robert W. Thatcher PhD [a]

[a] NeuroImaging Laboratory, Applied Neuroscience, Inc.

Published online: 18 May 2010.

To cite this article: Robert W. Thatcher PhD (2010) Validity and Reliability of Quantitative Electroencephalography, Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience, 14:2, 122-152, DOI: 10.1080/10874201003773500

To link to this article: http://dx.doi.org/10.1080/10874201003773500

Journal of Neurotherapy, 14:122–152, 2010

ISSN: 1087-4208 print/1530-017X online
DOI: 10.1080/10874201003773500

TECHNICAL CORNER

Validity and Reliability of Quantitative Electroencephalography

Robert W. Thatcher, PhD

ABSTRACT. Reliability and validity are statistical concepts that are reviewed and then applied to the field of quantitative electroencephalography (qEEG). The review of the scientific literature demonstrated high levels of split-half and test–retest reliability of qEEG and convincing content and predictive validity as well as other forms of validity. QEEG is distinguished from nonquantitative EEG (''eyeball'' examination of EEG traces), with the latter showing low reliability (e.g., 0.2–0.29) and poor interrater agreement for nonepilepsy evaluation. In contrast, qEEG is greater than 0.9 reliable with as little as 40-s epochs and remains stable with high test–retest reliability over many days and weeks. Predictive validity of qEEG is established by significant and replicable correlations with clinical measures and accurate predictions of outcome and performance on neuropsychological tests. In contrast, non-qEEG or eyeball visual examination of the EEG traces in cases of nonepilepsy has essentially zero predictive validity. Content validity of qEEG is established by correlations with independent measures such as the MRI, PET and SPECT, the Glasgow Coma Score, neuropsychological tests, and so on, where the scientific literature again demonstrates significant correlations between qEEG and independent measures known to be related to various clinical disorders. The ability to test and evaluate the concepts of reliability and validity are demonstrated by mathematical proof and simulation where one can demonstrate test–retest reliability as well as zero physiological validity of coherence and phase differences when using an average reference and Laplacian montage.

KEYWORDS. Reliability, validity, quantitative EEG

Quantitative electroencephalography (qEEG) is distinguished from visual examination of EEG traces, referred to as ''nonquantitative EEG'' by the fact that the latter is subjective and involves low sensitivity and low interrater reliability for nonepilepsy cases (Benbadis et al., 2009; Cooper, Osselton, & Shaw, 1974; Malone et al., 2009; Piccinelli et al., 2005; Seshia, Young, & Zifkin, 2008; Woody, 1966, 1968). In

Robert W. Thatcher is affiliated with NeuroImaging Laboratory, Applied Neuroscience, Inc.
Address correspondence to: Robert W. Thatcher, PhD, NeuroImaging Laboratory, Applied Neuroscience, Inc., St. Petersburg, FL 33722 (E-mail: rwthatcher@yahoo.com).

contrast, the qEEG involves the use of computers and power spectral analyses and is more objective with higher reliability and higher clinical sensitivity than visual examination of the EEG traces for most psychiatric disorders and traumatic brain injury (Hughes & John, 1999). As stated in a recent visual non-qEEG study by Malone et al. (2009),

> The interobserver agreement (Kappa) for doctors and other health care professionals was poor at 0.21 and 0.29, respectively. Agreement with the correct diagnosis was also poor at 0.09 for doctors and −0.02 for other healthcare professionals. (p. 2097)

Or in a study of non-qEEG visual examination of the EEG traces it was concluded by Benbadis et al. (2009), "For physiologic nonepileptic episodes, the agreement was low (kappa = 0.09)" (p. 843).

A recent statement by the Canadian Society of Clinical Neurophysiology further emphasizes the low reliability of visual examination of EEG traces or non-qEEG in the year 2008, where they conclude

> A high level of evidence does not exist for many aspects of testing for visual sensitivity. Evidenced-based studies are needed in several areas, including (i) reliability of LED-based stimulators, (ii) the most appropriate montages for displaying responses, (iii) testing during pregnancy, and (iv) the role of visual-sensitivity testing in the diagnosis of neurological disorders affecting the elderly and very elderly. (Sehsia et al., 2008, p. 133)

The improved sensitivity and reliability of qEEG was first recognized by Hans Berger in 1934 when he performed a qEEG analysis involving the power spectrum of the EEG with a mechanical analog computer and later by Kornmuller in 1937 and Grass and Gibbs (1938) (see Niedermeyer & Lopes Da Silva, 2005). QEEG in the year 2010 clearly surpasses conventional visual examination of EEG traces because qEEG has high resolution in the millisecond time domain and approximately 1 cm in the spatial domain, which gives qEEG the ability to measure network dynamics that are simply "invisible" to the naked eye. Over the last 40 years the accuracy, sensitivity, reliability, validity, and resolution of qEEG has steadily increased because of the efforts of hundreds of dedicated scientists and clinicians that have produced approximately 90,000 qEEG studies cited in the National Library of Medicine's database. Since approximately 1975 it is very difficult to publish a non-qEEG study in a peer-reviewed journal because of the subjective nature of different visual readers agreeing or disagreeing in their opinions about the squiggles of the "EEG" with low "interrater reliability" for nonepilepsy cases (Benbadis et al., 2009; Cooper et al., 1974; Malone et al., 2009; Piccinelli et al., 2005; Seshia et al., 2008; Woody, 1966, 1968). This article does not discuss the issue of qEEG in the detection of epilepsy. This topic is well covered by many studies (see Niedermeyer & Lopes Da Silva, 2005). Instead, this article is focused on the nonepilepsy cases, the very cases that visual non-qEEG is weakest.

It is useful to first revisit the standard concepts of "reliability" and "validity" of quantitative EEG while keeping in mind the historical background of non-qEEG visual examination of EEG traces. Although non-qEEG is insensitive and unreliable for the evaluation for the vast majority of psychiatric and psychological disorders and mild traumatic brain injury, it is used in approximately 99% of U.S. hospitals as the currently accepted standard of care. Given this background, the purpose of this article is to define the concepts of "reliability" and "validity" and evaluate these concepts as they apply to the clinical application of qEEG. Such an endeavor requires some knowledge of the methods of measurement as well as about the basic neuroanatomy and neurophysiology functions of the brain.

It is not possible to cover all clinical disorders, and therefore mild traumatic brain injury will be used as examples of qEEG validity and reliability. The same high levels of clinical validity and reliability (i.e., >0.95)

of qEEG have been published for a wide variety of psychiatric and psychological disorders—to cite only a few, for example, attention deficit disorders (Mazaheri et al., 2010; van Dongen-Boomsma et al., 2010), attention deficit hyperactivity disorder (Gevensleben et al., 2009), schizophrenia (Begić, Mahnik-Milos, & Grubisin, 2009; Siegle, Condray, Thase, Keshavan, & Steinhauer, 2010); depression (Pizzagalli et al., 2004); obsessive compulsive disorders (Velikova et al., 2010); addiction disorders (Reid et al., 2003); anxiety disorders (Hannesdóttir, Doxie, Bell, Ollendick, & Wolfe, 2010) and many other disorders. The reader is encouraged to visit the National Library of Medicine database at https://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed and use the search terms "EEG and xx" where xx = a clinical disorder. Read the Methods section to determine that a computer was used to analyze the EEG which satisfies the definition of qEEG and then read the hundreds of statistically significant qEEG studies. The search term "EEG" is necessary because the National Library of Medicine searches article titles and rarely if ever is the term "qEEG" used in the title (e.g., this author has published more than 150 peer-reviewed articles on qEEG and has never used the term "qEEG or QEEG" in the title).

VALIDITY DEFINED

Validity is defined by the extent to which any measuring instrument measures what it is intended to measure. In other words, validity concerns the relationship between what is being measured and the nature and use to which the measurement is being applied. One evaluates a measuring instrument in relation to the purpose for which it is being used. There are three different types of validity: (a) criterion-related validity, also called "predictive validity"; (b) content validity, also called "face validity"; and (c) construct validity. If a measurement is unreliable, then it can not be valid; however, if a method is reliable it can also be invalid, that is, consistently off the mark or consistently wrong. Suffice it to say that clinical correlations are fundamental to the concept of validity and are dependent on our knowledge of basic neuroanatomy and neurophysiology. These concepts are also dependent on our methods of measurement and the confidence one has in the mathematical simulations when applied in the laboratory or clinical context. Today there are a wide number of fully tested mathematical and digital signal processing methods that can be rapidly evaluated using calibrated signals and a high-speed computer to determine the mathematical validity of any method. This article does not spend time on this topic except for a brief mention of a few methods that are not valid when applied to coherence and phase measures. This is because of technical limitations, for example, the use of an average reference or the Laplacian surface transform and Independent Components Analysis (ICA) and the calculation of coherence and phase. It is shown in a later section that the average reference and the Laplacian distort the natural physiological phase relationships in the EEG and any subsequent analyses of phase and coherence are invalidated when these remontaging or reconstruction methods are used (Nunez, 1981; Rappelsberger, 1989). The average reference and Laplacian and ICA methods are valid for absolute power measures but have limitations for phase measures. This is a good example of why validity is defined as the extent to which a measuring instrument measures what it is intended to measure.

Leaving the mathematical and simulation methods aside for the moment, the most critical factor in determining the clinical validity of qEEG is knowledge about the neuroanatomy and neurophysiology and functional brain systems, because without this knowledge then it is not possible to even know if a given measurement is clinically valid in the first place. For example, neurological evaluation of space occupying lesions has been correlated with the locations and frequency changes that have been observed in the EEG traces and in qEEG analyses, for example, lesions of the visual cortex resulted in distortions of the EEG generated from the occipital scalp locations, or lesions

of the frontal lobe resulted in distortions of the EEG traces arising in frontal regions, and so on. However, early neurological and neuropsychological studies have shown that function was not located in any one part of the brain (Luria, 1973). Instead the brain is made up of complex and interconnected groupings of neurons that constitute "functional systems," such as the "digestive system" or the "respiratory system" in which cooperative sequencing and interactions give rise to an overall function at each moment of time (Luria, 1973). This widely accepted view of brain function as a complicated functional system became dominant in the 1950s and 1960s and is still the accepted view today. For example, since the 1980s new technologies such as functional MRI (fMRI), PET, SPECT, and qEEG/MEG have provided ample evidence for distributed functional systems involved in perception, memory, drives, emotions, voluntary and involuntary movements, executive functions, and various psychiatric and psychological dysfunctions (Mesulam, 2000). Modern PET, qEEG, MEG, and fMRI studies are consistent with the historical view of "functional systems" presented by Luria in the 1950s (Luria, 1973), that is, there is no absolute functional localization because a functional system of dynamically coupled subregions of the brain is operating. For example, several fMRI and MRI studies (e.g., diffusion tensor imaging, or DTI) have shown that the brain is organized by a relatively small subset of "modules" and "hubs" that represent clusters of neurons with high within-cluster connectivity and sparse long-distance connectivity (Chen, He, Rosa-Neto, Germann, & Evans, 2008; Hagmann et al., 2008; He et al., 2009). Modular organization is a common property of complex systems and "Small-World" models in which maximum efficiency is achieved when local clusters of neurons rely on a small set of long-distance connections to minimize the "expense" of wiring by shortened time delays between modules (Buzsaki, 2006; He et al., 2009). Also, recent qEEG and MEG analyses have demonstrated that important visually invisible processes such as directed coherence, phase delays, phase locking, and phase shifting of different frequencies is critical in cognitive functions and various clinical disorders (Buzsaki, 2006; Sauseng & Klimesch, 2008; Thatcher et al., 2009). Phase shift and phase synchrony has been shown to be one of the fundamental processes involved in the coordination of neural activity located in spatially distributed "modules" at each moment of time (Breakspear & Terry, 2002; Freeman, Burke, & Homes, 2003; Freeman & Rogers, 2002; Lachaux et al., 2000; Sanseug & Klemish, 2008; Thatcher, North, & Biver, 2005a, 2008a, 2008b; Thatcher et al., 2009).

VALIDITY OF COHERENCE AND PHASE

Coherence is a measure of the stability of phase differences between two time series. Coherence is not a direct measure of an attribute like "temperature" or "volts," instead it is a measure of the "reliability" of phase differences in a time series. If the phase differences are constant and unchanging over time then coherence equals 1. If, on the other hand, phase differences are changing over time and are random over time then coherence equals 0 (i.e., unreliable over time). Therefore, unlike absolute power, coherence is not a straightforward analytical measure; rather coherence depends on multiple time samples to compute a correlation coefficient in the frequency or time domains. The validity and reliability of coherence fundamentally depends on the number of time samples as well as the number of connections (N) and the strength of connections (S) in a network, or Coherence $= N \times S$. Thus, it is sensitive to the number and strength of connections and therefore as the number or strength of connections decreases then coherence decreases. This is because it is a valid network measure and as one would expect the reliability of the coherence measure declines when the number or strength of connections declines. Here is an instance where the validity of coherence is established by the fact that the reliability is low, that is, no connections means no coupling and coherence approximates zero.

To evaluate the validity of coherence, it is important to employ simulations using calibrated sine waves mixed with noise. In this manner a linear relationship between the magnitude of coherence and the magnitude of the signal-to-noise ratio can be demonstrated, which is a direct measure of the predictive validity and concurrent validity of coherence. For example, if one were to use an invalid method to compute coherence such as with an average reference, then it is irrelevant what the stability of the measure is because coherence is no longer measuring phase stability between two time series and therefore has limited physiological validity.

Figure 1 is an example of a validation test of coherence using 5 Hz sine waves and a 30° shift in phase angle with step-by-step addition of random noise. As shown in Figure 1, a simple validity test of coherence is to use a signal generator to create a calibrated 1 μV sine wave at 5 Hz as a reference signal, and then compute coherence to the same 1 μV sine wave at 5 Hz but shifted by 30° and adding 2 μV of random noise, then in the next channel add 4 μV of random noise, then 6 μV, and so on.

Mathematically, validity equals a linear relationship between the magnitude of coherence and the signal-to-noise ratio, that is, the greater the noise, the lower is coherence. If one fails to obtain a linear relationship, then the method of computing coherence is invalid. If one reliably produces the same set of numbers but a nonlinear relationship (i.e., no straight line) occurs, then this means that the method of computing coherence is invalid (the method reliably produces the wrong results or is reliably off the mark). Figure 2 shows the results of the coherence test in Figure 1 and demonstrates a linear relationship between coherence and the signal-to-noise ratio, thus demonstrating that a standard Fast Fourier Transform (FFT) method of calculating coherence using a single common reference (e.g., one ear, linked ears, Cz, etc.) is valid. Note that the phase difference of 30° is preserved even

FIGURE 1. An example of four 1 uV and 5 Hz sine waves with the second to the 4th sine wave shifted by 30 degrees. *Note.* Gaussian noise is added incrementally to channels 2 to 4. Channel 2 = 1 uV signal +2 uV of noise, channel 3 = 1 uV signal +4 uV of noise and channel 4 = 1 uV signal +6 uV of noise. Nineteen channels were used in the analyses of coherence in 2 uV of noise increments. The Fast Fourier Transform analysis is the mean of thirty 2-s epochs sampled at 128 Hz.

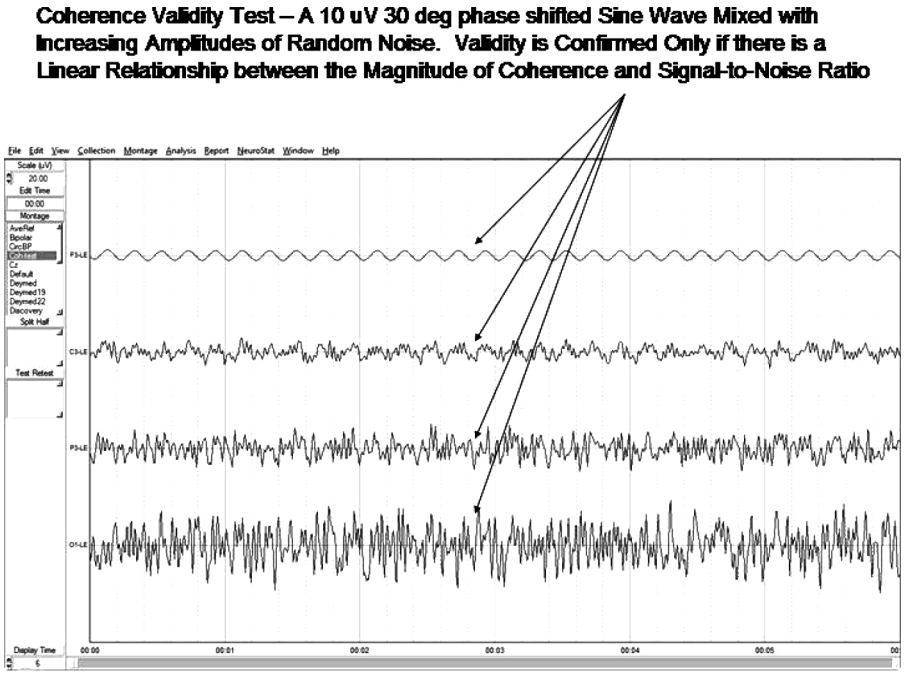
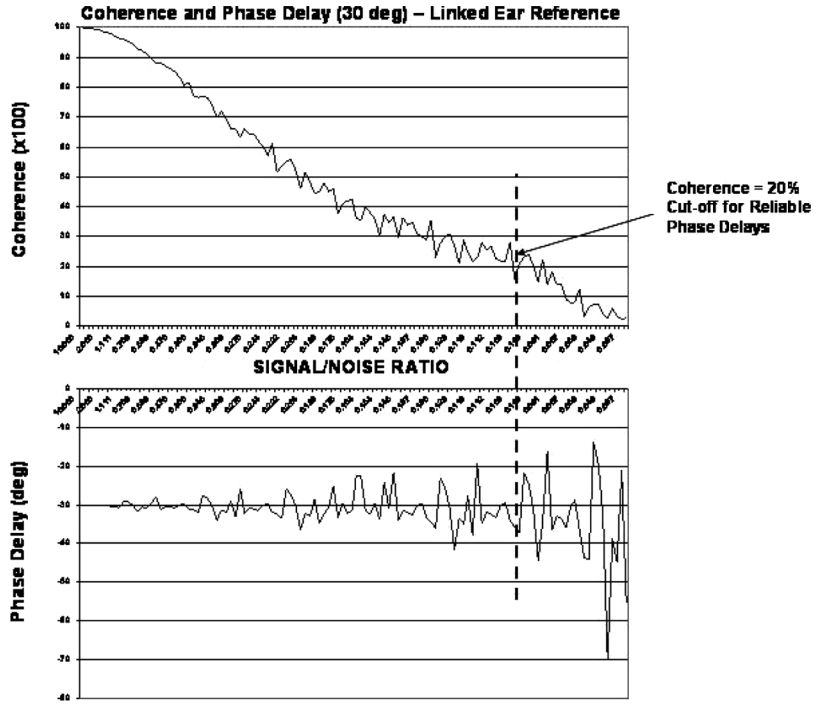


FIGURE 2. Top is coherence (*y*-axis) vs signal-to-noise ratio (*x*-axis). Bottom is phase angle on the *y*-axis and signal-to-noise ratio on the *x*-axis. *Note.* Phase locking is minimal or absent when coherence is less than approximately 0.2 or 20%. The sample size was 60 s of EEG data and smoother curves can be obtained by increasing the epoch length.



when coherence is less than 0.2. The preservation of the phase difference and the linear decrease as a function of noise is a mathematical test of the validity of coherence.

WHY THE AVERAGE REFERENCE OR LAPLACIAN MONTAGES ARE INVALID WHEN COMPUTING COHERENCE AND PHASE DIFFERENCES

An important lesson in reliability and validity is taught when examining any study that fails to use a common reference when computing coherence. For example, the average reference mathematically adds the phase differences between all combinations of scalp EEG time series, and then divides by the number of electrodes to form an average. Finally, the average is subtracted time point by time point from the original time series recorded from each individual electrode, thereby replacing the original time series with a distorted time series. This process scrambles up the physiological phase differences so that they are irretrievably lost and can never be recovered. The method of mixing phase differences precludes meaningful physiological or clinical correlations since measures such as conduction velocity or synaptic rise- or fall-times can no longer be estimated due to the average referencing.

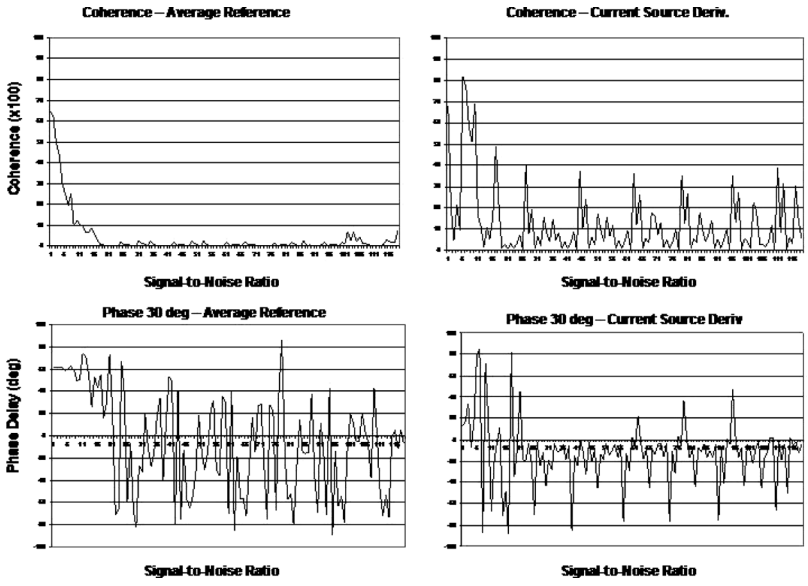
Also, coherence methods such as ''directed coherence'' cannot be computed and more sophisticated analyses such as phase reset, phase shift and phase lock are precluded when using an average reference. The mixing together of phase differences in the EEG traces is also a problem when using the Laplacian transform. Similarly, reconstruction of EEG time series using ICA also replaces the original time series with an altered one that eliminates any physiological phase relationships and therefore is an invalid method of calculating coherence as well.

One may obtain high reliability in test–retest measures of coherence; however, the reliability is irrelevant because the method of computation using an average reference or a Laplacian montage to compute coherence is invalid in the first place.

As pointed out by Nunez (1981), ''the average reference method of EEG recording requires considerable caution in the interpretation of the resulting record'' (p. 194) and that ''the phase relationship between two electrodes is also ambiguous'' (p. 195). As mentioned previously, when coherence is near unity then the oscillators are synchronized and phase and frequency locked. This means that when coherence is too low (e.g., <0.2), then the estimate of the average phase angle may not be stable and phase relationships could be nonlinear and not synchronized or phase locked.

The distortions and invalidity of the average reference and Laplacian transform are easy to demonstrate using calibrated sine waves mixed with noise just as was done in Figures 1 and 2. For example, Figure 3 shows the same simulation as shown in Figure 2, with a 30° phase shift as used for coherence with a common reference. The top row is coherence on the *y*-axis and the bottom row is the phase difference, the left column is using the average reference and the right column is the Laplacian. It can be seen here that coherence is extremely variable and does not decrease as a linear function of signal-to-noise ratio using either the average reference nor the Laplacian montage. It can also be seen in Figure 3 that EEG phase differences never approximate 30° and are extremely variable at all levels of the signal-to-noise ratio.

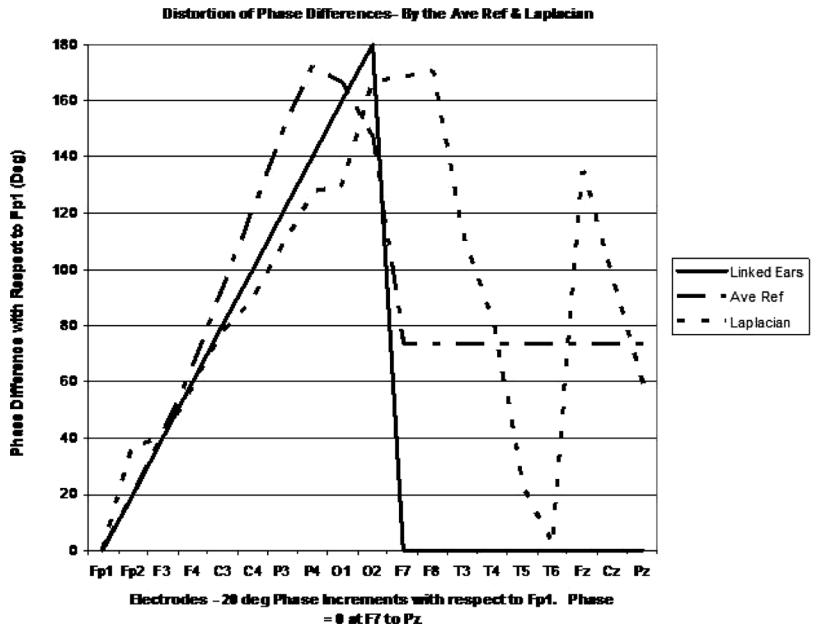
FIGURE 3. Left top is coherence (*y*-axis) versus signal-to-noise ratio (*x*-axis) with a 30° phase shift as shown in Figure 2 using the average reference. *Note.* The left bottom is phase differences in degrees in the *y*-axis and the *x*-axis is the signal-to-noise ratio using the average reference. The right top graph is coherence (*y*-axis) versus signal-to-noise ratio (*x*-axis) using the Laplacian montage. The right bottom is phase difference on the *y*-axis and signal-to-noise on the *x*-axis using the Laplacian montage. In both instances, coherence drops off rapidly and is invalid with no linear relationship between signal and noise. The bottom graphs show that both the average reference and the Laplacian montage fails to track the 30° phase shift that was present in the original time series. In fact, the phase difference is totally absent and unrepresented when using an average reference or a Laplacian montage and these simulations demonstrate that the average reference and the Laplcain montage are not physiologically valid because they do not preserve phase differences or the essential time differences on which the brain operates.



The results of these analyses are consistent with those by Rappelsberger (1989), who emphasized the value and validity of using a single reference and linked ears in estimating the magnitude of shared or coupled activity between two scalp electrodes. The use of remontage methods such as the average reference and Laplacian source derivation are useful in helping to determine the location of the sources of EEG of different amplitudes at different locations. However, the results of this study showed that coherence is invalid when using either an average reference or the Laplacian source derivation. This same conclusion was also demonstrated by Essl and Rappelsburger (1998); Kamiński and Blinowska (1991); Kamiński, Blinowska, and Szellenberger (1997); and Korzeniewska, Mańczak, Kamiński, Blinowska, and Kasicki (2003).

The average reference and the Laplacian transform also distort measures of phase differences, which is also easy to demonstrate by using calibrated sine waves. For example, a sine wave at Fp1 of 5 Hz and 100 μV with zero phase shift, Fp2 of 5 Hz and 100 μV with 20° phase shift; F3 of 5 Hz and 100 μV with 40° phase shift; F4 of 5 Hz and 100 μV with 60° phase shift; C3 of 5 Hz and 100 μV with 80° phase shift; C4 of 5 Hz and 100 μV with 100° phase shift; P3 of 5 Hz and 100 μV with 120° phase shift; P4 of 5 Hz and 100 μV with 140° phase shift; O1 of 5 Hz and 100 μV with 160° phase shift and O2 of 5 Hz and 100 μV with 180° phase shift and channels F8 to Pz = 0 μV and zero phase shift. Figure 4 compares the incremental phase shift with respect to Fp1 using linked ears common reference (solid black line), the average reference (long dashed line), and the Laplacian (short dashed line). This is another demonstration of how a noncommon reference such as the average reference and the Laplacian scramble phase differences and therefore caution should be used and only a common reference recording (any common reference and not just linked ears) is the only valid method of relating phase differences to the underlying neurophysiology, for example, conduction velocities, synaptic rise times, directed coherence, phase reset, and so on.

FIGURE 4. Demonstration of distortions in phase differences in a test using 20° increments of phase difference with respect to Fp1. *Note.* The solid black line is using a Linked Ears common reference, which accurately shows the step-by-step 20° increments in phase difference. The average reference (dash-dot line) and the Laplacian (dashed line) significantly distort the phase differences.



VALIDITY BY HYPOTHESIS TESTING AND QEEG NORMATIVE DATABASES

The Gaussian or Normal distribution is an ideal bell-shaped curve that provides a probability distribution which is symmetrical about its mean. Skewness and kurtosis are measures of the symmetry and peakedness, respectively of the Gaussian distribution. In the ideal case of the Gaussian distribution skewness and kurtosis = 0. In the real world of data sampling distributions skewness and kurtosis = 0 is never achieved and, therefore, some reasonable standard of deviation from the ideal is needed to determine the approximation of a distribution to Gaussian. The primary reason to approximate ''normality'' of a distribution of EEG measures is that the sensitivity (i.e., error rate) of any normative EEG database is determined directly by the shape of the sampling distribution. In a normal distribution, for example, one would expect that approximately 5% of the samples will be equal to or greater than ± 2 *SD* and approximately $0.13\% \pm 3$ *SD* (John, 1977; John, Prichep, & Easton, 1987; Hayes, 1973; Prichep, 2005; Thatcher, Biver, & North, 2003; Thatcher, Walker, Biver, North, & Curtin, 2003).

A practical test of the sensitivity and accuracy of a database can be provided by cross-validation. There are many different ways to cross-validate the contents of a database. One is to obtain independent samples, and another is to use a leave-one-out cross-validation method to compute Z scores for each individual subject within it. The former is generally not possible because it requires sampling large numbers of additional participants who have been carefully screened for clinical normality. However, the second method is certainly possible for any database. Gaussian cross-validation of an EEG database can be accomplished by the latter method in which a participant is removed from the distribution and the Z scores

FIGURE 5. Example of Gaussian cross-validation of EEG normative database. *Source*: Thatcher, Walker, et al. (2003).

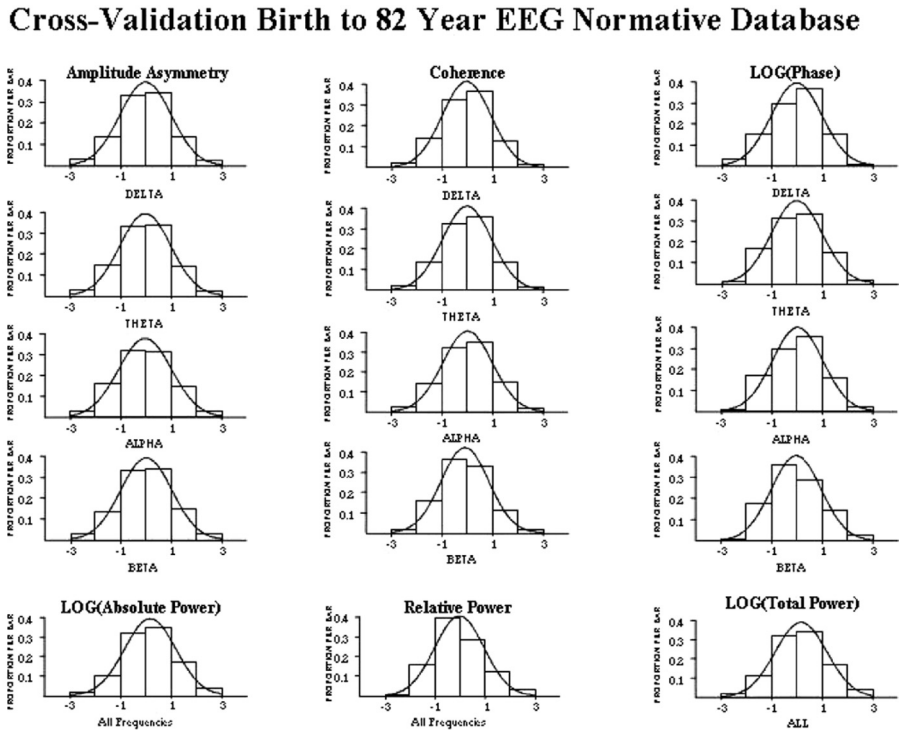


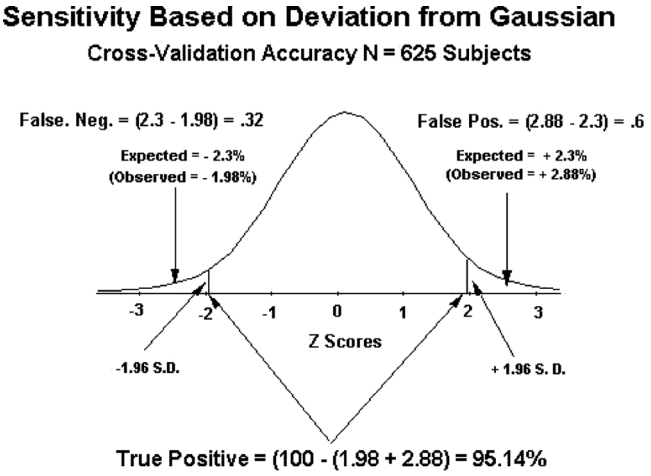
TABLE 1. Cross-validation of EEG normative database.

Measure	% > 2 *SD*	% < 2 *SD*	% > 3 *SD*	% < 3 *SD*
Delta Amplitude Asym.	2.58	3.08	0.21	0.19
Theta Amplitude Asym.	2.29	2.62	0.15	0.13
Alpha Amplitude Asym.	2.71	2.72	0.18	0.19
Beta Amplitude Asym.	2.68	2.65	0.15	0.15
Delta Coherence	1.99	2.14	0.14	0.22
Theta Coherence	2.22	1.88	0.22	0.16
Alpha Coherence	2.55	1.62	0.18	0.18
Beta Coherence	2.20	1.38	0.18	0.10
Delta Phase†	0.89	3.52	0.00	0.23
Theta Phase†	1.61	1.87	0.04	0.13
Alpha Phase†	1.61	1.66	0.04	0.24
Beta Phase†	2.83	0.72	0.27	0.03
Absolute Power†	4.15	1.67	0.23	0.12
Relative Power	4.09	0.52	0.68	0.00
Total Power†	4.23	1.60	0.08	0.04
Average	2.58	1.98	0.18	0.14

Note. Source: Thatcher, Walker, et al. (2003).
†Transformed data.

computed for all variables based on his or her respective age matched mean and standard deviation in the normative database. The participant is placed back in the distribution and then the next participant is removed and a Z score is computed, and this process is repeated for each normal participant to obtain an estimate of the false positive hit rate. A distribution of Z scores for each of the EEG variables for each participant was then tabulated. Figure 5 is an example of the Gaussian distributions of the

FIGURE 6. Illustration of method of computing error rates or sensitivity of a normative EEG database based on the cross-validation deviation from Gaussian. *Note.* Source: Thatcher, Walker, et al. (2003).



$$\text{Sensitivity} = \frac{TP}{TP + (FP + FN)} = \frac{95.14}{95.14 + 1.0} = 98.96\%$$

$$\text{Specificity} = \frac{TN}{TN + (FP + FN)} = \text{Undefined}$$

cross-validated Z scores of 625 participants from birth to 82 years of age used in a normative EEG database (Thatcher, Walker, et al., 2003).

Table 1 shows the results of a Gaussian cross-validation of the 625 participants in the normative EEG database used in the evaluation of patients (Thatcher, Walker, et al., 2003). A perfect cross-validation would be 2.3% at +2 *SD*, 2.3% at −2 *SD*, 0.13% at +3 *SD*, and 0.13% at −3 *SD*. Table 1 shows a cross-validation grand average of 2.28% ± 2 *SD* and 0.16% ± 3 *SD*. The cross-validation result shows that the EEG normative database is statistically accurate and sensitive with slight differences between variables that should be taken into account when evaluating individual Z scores.

Figure 6 is a bell-shaped curve showing the ideal Gaussian and the average cross-validation values of the EEG normative database used to evaluate patients. The error rates or the statistical sensitivity of a qEEG normative database are directly related to the deviation from a Gaussian distribution. Figure 6 also illustrates the method of estimating the statistical sensitivity of a normative EEG database in terms of the deviation from Gaussian.

Table 2 is an example of the calculated sensitivity of an EEG normative database for different age groups using the method described in Figure 6.

PREDICTIVE VALIDITY OF NORMATIVE DATABASES

Predictive (or criterion) validity has a close relationship to hypothesis testing by subjecting the measure to a discriminant or cluster analysis to some statistical analysis in order to separate a clinical subtype from a normal reference database. Nunnally (1978) gave a useful definition of predictive validity as "when the purpose is to use an instrument to estimate some important form of behavior that is external to the measuring instrument itself, the latter being referred to as criterion [predictive] validity" (p. 87). For example, science "validates" the clinical usefulness of a measure by its false positive and false negative rates and by the extent to which there are statistically significant correlations to other clinical measures and, especially, to clinical outcomes (Hughes & John, 1999).

An example of predictive validity of the Linked Ears qEEG normative database is the use of a discriminant function to evaluate the false positive/false negative classification rate using a normative database

TABLE 2. Normative EEG database sensitivities for different age groups at ±2 *SD* and ±3 *SD*.

FFT Normative Database				
2 STDEVs	CALC SENSITIVITY: FP = TP/(TP + FP) or FN = TP/(TP + FN)			
AGES	(+/−2 SD)	(> = 2 SD)	(<=−2 SD)	
0–5.99	0.95448265	0.9771774	0.97730526	+/−**2 Std. Dev.**
6–9.99	0.95440363	0.9772031	0.97720054	
10–12.99	0.9543997	0.97724346	0.97715624	
13–15.99	0.95440512	0.97723601	0.97716911	
16-ADULT	0.9543945	0.97718143	0.97721307	
ALL	0.95442375	0.97720714	0.97721661	
3 STDEVs	CALC SENSITIVITY: FP = TP/(TP + FP) or FN = TP/(TP + FN)			
AGES	(+/−3 SD)	(> = 3 SD)	(< = −3 SD)	
0–5.99	0.99743898	0.99871123	0.99872774	+/−**3 Std. Dev.**
6–9.99	0.99744112	0.99871611	0.99872501	
10–12.99	0.99744688	0.99873171	0.99871518	
13–15.99	0.99743186	0.99871951	0.99871234	
16-ADULT	0.99743835	0.99870216	0.99873619	
ALL	0.99744002	0.99871716	0.99872286	

Note. FFT = Fast Fourier Transform. Source: Thatcher, Walker, et al. (2003).

FIGURE 7. Example of predictive and content validity by clinical correlations of quantitative EEG (QEEG) with neuropsychological test scores. *Note.* TBI = traumatic brain injury; WAIS = Wechsler Adult Intelligence Scale. Source: Thatcher, North, et al. (2001), reprinted with permission.

Correlation : QEEG and Neuropsychological Test Scores in TBI Patients. J. Neuropsychiatry & Clin. Neurosci., 13: 77-87, 2001

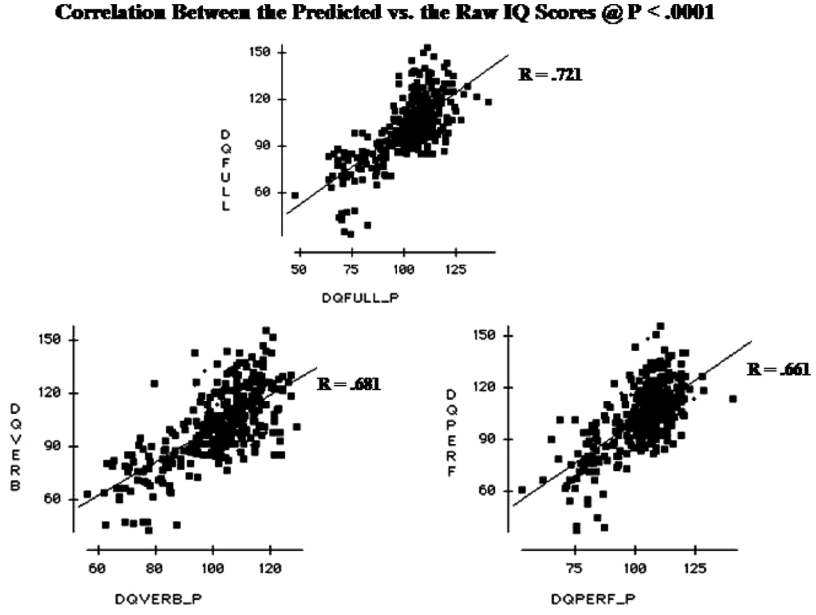
	Correlation	Probability
WAIS TEST-Scaled Scores		
Vocabulary	-0.416	0.05
Similarities	-0.640	0.001
Picture Arrangement	-0.576	0.01
Performance	-0.504	0.01
Digit Symbol	-0.524	0.01
ATTENTION TEST-Raw Scores		
Trail Making A-Response Time	0.627	0.001
Trail Making B-Response Time	0.627	0.001
Stroop-Word	-0.427	0.05
Stroop-Color	-0.618	0.001

Examples of Predictive Validity of QEEG in the Evaluation Of Traumatic Brain Injury

and traumatic brain injury (TBI) patients (Thatcher, Walker, Gerson, & Geisler, 1989). In this study the traumatic brain injured patients were distinguished from age-matched normal control participants at a classification accuracy of 96.2%. Four different cross-validations were conducted in the Thatcher et al. (1989) study and showed similar accuracies although the strength of the discrimination declined as a function of time from injury to test.

Figure 7 shows the correlation to neuropsychological test scores in an independent replication of the Thatcher et al. (1989) study. In this study a similar discriminant function produced similar sensitivities and

FIGURE 8. Example of content validity demonstrated by statistically significant correlations between Full-Scale IQ and quantitative EEG (from Thatcher, North, & Biver, 2005a, reprinted with permission).



predicted the Glasgow Coma Score with a correlation of 0.85 (Thatcher, North, et al., 2001). Another example of predictive validity is the ability of qEEG normative values to predict cognitive functioning. Figure 8 shows correlations to Full-Scale IQ as an example of predictive validity and content validity. A more complete analysis of the predictive validity of a normative EEG database is shown in Table 3 (Thatcher, Walker, et al., 2003; Thatcher, North, & Biver, 2005b, 2005c). In Table 3, the percentage of statistically significant correlations at $p < .01$ between qEEG normative EEG and WRAT School Achievement scores and measures of intelligence are shown. The relative effect size of the normative EEG correlations differs for different measures, which is valuable information when using any normative database, not just a qEEG normative database. Similar high and significant correlations between qEEG and neuropsychological test performance have been published in many studies. A search of the National Library of Medicine's database using the search terms EEG and Neuropsychological Tests produced 1,351 citations.

EXAMPLES OF CONTENT VALIDITY OF NORMATIVE DATABASES

Content validity is defined by the extent to which an empirical measurement reflects a specific domain of content. For example, a test in arithmetic operations would not be content valid if the test problems focused only on addition, thus neglecting subtraction,

TABLE 3. Examples of predictive validity by clinical correlations between quantitative EEG and intelligence (Wechsler Intelligence Scale for Children–Revised) and academic achievement tests (Wide Range Achievement Test).

EFFECT SIZE P < .01: qEEG Correlations with School Achievement & IQ Measures						
Percent Significant Correlations @ P < .01, N = 466						
P < = .01	READING	SPELLING	ARITH	IQFULL	IQVERB	IQPERF
Amplitude Asymmetry						
DELTA	64%	61%	55%	64%	61%	61%
THETA	78%	70%	70%	70%	67%	59%
ALPHA	63%	63%	53%	64%	63%	52%
BETA	56%	56%	34%	58%	61%	47%
Coherence						
DELTA	27%	14%	41%	38%	22%	38%
THETA	27%	6%	36%	30%	27%	23%
ALPHA	9%	6%	45%	11%	14%	5%
BETA	11%	5%	38%	22%	17%	6%
Absolute Phase						
DELTA	11%	8%	8%	16%	6%	17%
THETA	9%	5%	8%	13%	9%	17%
ALPHA	9%	3%	33%	14%	19%	6%
BETA	9%	5%	30%	6%	9%	3%
Relative Power						
DELTA	13%	0%	31%	0%	6%	0%
THETA	56%	44%	94%	6%	6%	0%
ALPHA	19%	0%	75%	0%	0%	0%
BETA	13%	6%	44%	19%	13%	13%
Relative Power Ratios						
Theta/Beta	50%	44%	63%	56%	56%	50%
Theta/Alpha	13%	0%	69%	0%	0%	0%
Alpha/Beta	50%	31%	50%	38%	38%	25%
Delta/Theta	19%	25%	56%	19%	13%	25%

Note. Source: Thatcher, Walker, et al. (2003).

multiplication, and division. By the same token, a content-valid measure of cognitive decline following a stroke should include measures of memory capacity, attention and executive function, and so on.

Normative databases are distinct from small experimental control groups in their scope and their sampling restriction to clinically normal or otherwise healthy individuals for the purpose of comparison. Another distinguishing characteristic of normative databases is the ability to compare a single individual to a population of "normal" individuals in order to identify the measures that are deviant from normal and the magnitude of deviation. Normative databases themselves do not diagnose a patient's clinical problem. Rather, a trained professional first evaluates the patient's clinical history and clinical symptoms and complaints and then uses the results of normative database comparisons to aid in the development of an accurate clinical diagnosis. Most important, this is to link functional localization of deregulated brain regions (i.e., anatomical hypotheses) to a patient's symptoms and complaints.

There are many examples of the clinical content validity of qEEG and normal control groups in attention deficit disorder, attention deficit/hyperactivity disorder, schizophrenia, compulsive disorders, depression, epilepsy, TBI, and a wide number of clinical groupings of patients as reviewed by Hughes and John (1999). In most of these studies an assortment of clinical measures were correlated to a variety of brain EEG sources related to the disorder under study. One of the most consistent and relevant findings is anatomical localization related to different psychiatric and psychological disorders, for example, cingulate gyrus and depression, right parietal lobe and spatial neglect, left angular gyrus and dyslexia, and so on. QEEG anatomical correlations with clinical disorders form the foundation of modern day qEEG interpretation, another example of content validity. Since 1999, several hundred qEEG studies demonstrate anatomical and clinical validity. For example, all clinical LORETA qEEG studies demonstrate anatomical content validity in that there are no published studies showing low localization accuracy when using LORETA. The term "Low Resolution Electromagnetic Tomography" refers to a "smearing" around the spatially accurate maximum in the center of a spatial volume. This is defined by the point-spread function of the Laplacian spatial operator in LORETA, meaning that LORETA is spatially accurate but with a smeared resolution like a probability cloud. Clinical correlations consistent with PET and SPECT and fMRI are abundant in today's scientific literature (see the National Library of Medicine database at https://www.ncbi.nlm.nih.gov/sites/entrez and see the section in this article on Validity of LORETA for some specific citations).

ANATOMICAL HYPOTHESIS TESTING AND PLANNED QEEG COMPARISONS

The best use of parametric statistics is to form hypotheses prior to conducting an analysis in a procedure referred to as Planned Comparisons (Hayes, 1973). In this manner, one does not need to resort to multiple comparisons that are performed only when an experimenter isn't sure what the test is likely to yield and is unaware of possible statistically significant differences. Being unsure of possible statistically significant difference, it is not possible to form hypotheses, and one then must resort to multiple comparisons, which have may yield high Type II errors (false negatives) to reduce the Type I errors (false positives) because of possible false relationships between groups or between variables.

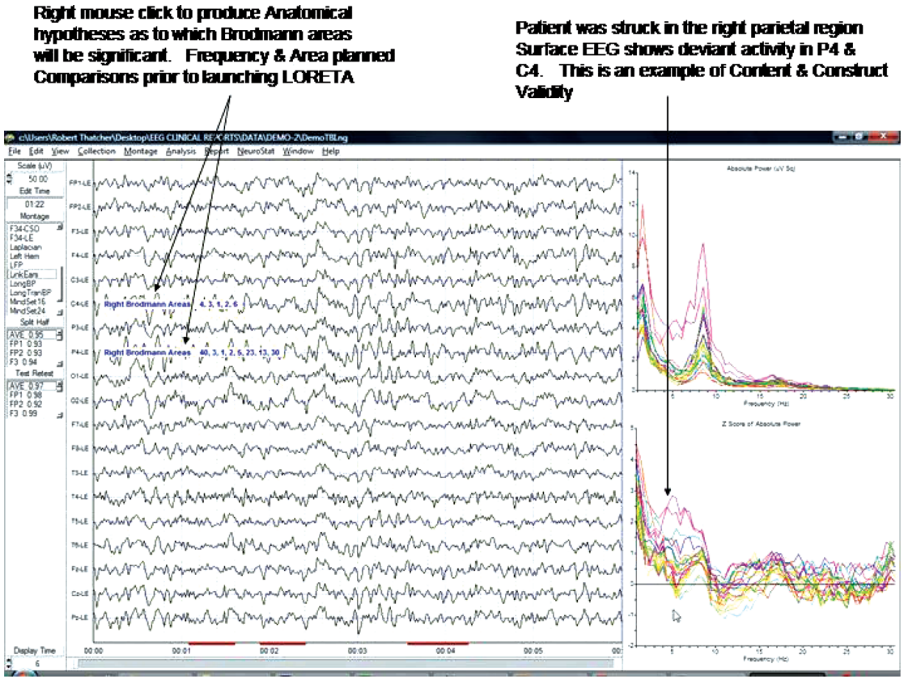
Planned comparisons are more robust and valid than multiple comparisons because specific hypotheses are generated prior to conducting statistical tests, which markedly minimizes the probability of both Type I and Type II errors. A complaint of the use of qEEG is that there are a large number of statistical tests, and one would expect 5% to be significant by chance alone. The problem with this argument is that the 5% by chance must be random in space and in qEEG features. The random chance argument is

discarded, however, when there are focal anatomical deviations that were predicted prior to analysis. Additional content validity occurs when the deviant qEEG findings are located in anatomical regions known to be linked to the patient's symptoms and clinical history. For example, the MRI uses approximately 10,000 voxels. One would expect 500 to be significant by chance at $p < .05$ if these 500 voxels are randomly distributed throughout the volume. However, if 100 voxels are statistically significant in the right parietal lobe, which is where the brain insult was, then the 5% significant multiple test argument is not valid and must be discarded. The same is true for the qEEG. For example, if one uses planned comparisons and predicts that the left parietal lobe will be deviant from normal in a dyslexic child prior to recording EEG and the qEEG shows many deviations from normal in the left parietal region, then this cannot be explained by chance alone. The use of planned comparisons is especially valuable when using LORETA source localization methods because thousands of voxels are involved. An example, of planned comparisons is in Figure 9. Here the surface qEEG analyses showed focal deviation from normal in the right hemisphere in a patient that was struck with a bat near to his right parietal lobe. The sources of the right parietal lobe deviations from normal are then predicted to appear in particular Brodmann areas prior to launching LORETA. Once LORETA is launched then the frequency and anatomical hypotheses can be tested to determine their accuracy and validity.

PREDICTIVE VALIDITY AND qEEG

Predictive validity is sometimes referred to as "criterion validity" and has a close relationship to hypothesis testing by subjecting the measure to an independent test of its ability to predict clinical measures such as

FIGURE 9. Example of "planned comparisons" using hypothesis creation prior to launching LORETA. *Note.* Content and construct validity are present because the patient was hit on the right parietal lobe and the right parietal lobe shows deviant EEG activity (e.g., >2 *SD*). Further construct validity is established by LORETA analyses that confirm anatomical hypotheses based on the surface EEG locations and frequencies of deviance.



severity of injury or intelligence, attention, executive function, and so on. Nunnally (1978) gave a useful definition of predictive validity: "When the purpose is to use an instrument to estimate some important form of behavior that is external to the measuring instrument itself, the latter being referred to as criterion-validity" (p. 88). For example, one "validates" a written driver's license test by hypothesizing that it accurately predicts how well a group of persons can operate an automobile. If the driving test fails to predict driving competence, then the test must be rejected or replaced. In the case of TBI, one "validates" the qEEG by showing that it accurately predicts severity of TBI as measured by hospital admission scores such as the Glasgow Coma Score or length of coma or in other independent tests such as neuropsychological tests and so on (Hughes & John, 1999).

FALSE POSITIVE AND FALSE NEGATIVE ERROR RATES OF qEEG: EXAMPLE OF CONTENT VALIDITY IN TRAUMATIC BRAIN INJURY

Peer-reviewed scientific publications of 608 mild TBI patients compared to 108 age-matched normal participants demonstrated, in independent cross-validations, an average false positive rate approximately 5% and an average false negative rate of approximately 10% to 15% (Thatcher et al., 1989). Similar levels of sensitivity and specificity were reported in a series of independent and replicated qEEG studies of TBI for the detection of a pattern consistent with traumatic brain injury as a causal agent (Leon-Carrion et al., 2008b; Thatcher et al., 1991; Thatcher, North, et al., 2001; Thornton, 1999; Thornton & Carmody, 2005). Obtaining a content-valid measure of any phenomenon involves at least three interrelated steps: (a) One must be able to specify the full domain of content that is relevant, (b) one must be able to identify the selection of relevant measures from the larger universe of possible measures with the understanding that over sampling is usually necessary, and (c) one must be able to test the content validity of the measuring instrument and/or be able to cite the peer reviewed literature in which the content-validity of the qEEG had been tested. As stated by Cronbach (1977), "One validates, not a test, but an *interpretation of data arising from a specified procedure*" (p. 447). This distinction is crucial because it is quite possible for a measuring instrument to be relatively valid for measuring one kind of phenomenon but entirely invalid for assessing other phenomena. The purpose of qEEG discriminant functions is not to derive a diagnosis because the diagnosis should be based on the patient's clinical history and symptoms and complaints. QEEG discriminant functions are designed to further evaluate the extent, locations, and severity of the EEG patterns that are present in individuals already diagnosed with a disorder.

QEEG involves the measurement of a relatively large number of electrical processes some of which may be affected by a TBI. For example, animal studies and imaging studies in humans have demonstrated that maximal damage to the brain following TBI occurs at the interface between the brain and the skull bone (Ommaya, 1968, 1995; Ommaya & Hirsch, 1971). Another primary and common injury to the brain due to TBI are "shear" forces in which rapid acceleration/deceleration events result in different brain parts moving at different rates; for example, the gray matter moves faster and further than the white matter, consequently stretching axonal fibers, and so on (Ommaya, 1968). Thus, a content valid qEEG measure of TBI should be capable of measuring electrical activity in frontal and temporal lobes where the brain-to-skull forces are greatest. Similarly, a content valid qEEG test of TBI must be capable of measuring EEG phase and EEG coherence, which reflect the axonal conduction velocities and long-distance cortical communication linkages (Thatcher, Biver, McAlaster, et al., 1998; Thatcher, Biver, et al., 2001; Thatcher et al., 1989). If these measures are omitted, then the test is not valid for the same reason that a test of arithmetic is invalid if it omits addition and subtraction. Over the years there is reasonable consistency of qEEG findings in TBI

across studies which can be summarized by (a) reduced power in the higher frequency bands (8–40 Hz), which is linearly related to the magnitude of injury to cortical gray matter; (b) increased slow waves in the delta frequency band (1–4 Hz) in the more severe cases of TBI, which is linearly related to the magnitude of cerebral white matter injury; (c) changes in EEG coherence and EEG phase delays, which are linearly related to the magnitude of injury to both the gray matter and the white matter, especially in frontal and temporal lobes (Thatcher, 2008).

QEEG CONSTRUCT VALIDITY

Construct validity is concerned with the validity of empirical measures and the hypothesis testing of theoretical concepts. As Carmines and Zeller (1979) stated, "Construct validity is concerned with the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts that are being measured" (p. 23) Construct validity typically involves three steps: (a) the theoretical relationship between the concepts themselves must be specified and testable hypotheses stated; (b) the empirical relationship between the measures of the concepts must be examined; and (c) the empirical evidence must be interpreted in terms of how it affirms, rejects, or clarifies the construct validity of the particular measure.

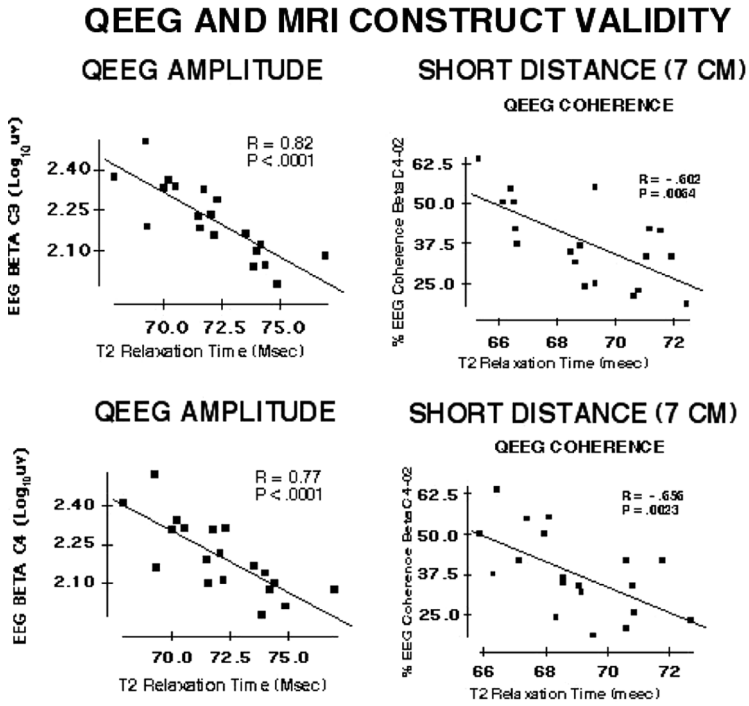
For example, in qEEG measures of TBI, one hypothesis is that rapid acceleration/deceleration contuses (bruises) brain tissue, especially where the brain sits on the bony skull vault (Ommaya, 1968, 1995), another theory is that damage to neuronal membranes will result in reduced ionic flow and reduced amplitude of the EEG and high frequencies and a shift in frequency toward the theta and delta frequencies (lower frequency ranges). These two theoretical hypotheses regarding which qEEG measures would be expected to change following TBI have been tested and confirmed in the peer reviewed scientific literature (Cao, Tutwiler, & Slobounov, 2008; Leon-Carrion et al., 2008a, 2008b; Randolph & Miller, 1998; Thatcher, Biver, Camacho, et al., 1998; Thatcher, Biver, McAlaster, et al., 1998; Thatcher, North, et al., 2001; Thatcher et al., 1991; Thatcher et al., 1989; Thornton, 1999; Thornton & Carmody, 2005).

The qEEG is also used for prognoses in the neurointensive care unit. Fabregas et al. (2004) reported a cross-validation performance error of 3.06% (95% confidence interval) for predicting recovery from coma. Similar accuracy of predicting recovery of consciousness was reported by others (Buzea, 1995; Claassen, 2000; Hyllienmark & Amark, 2007; Jordan, 1993; Kane, Moss, Curry, & Butler, 1998; Scheuer, 2002; Shields et al., 2007; Thatcher et al., 1991). Jordan (1993) reported that qEEG can impact medical decision-making in 81% of the monitored patients, and Claassen, Baeumer, and Hansen (2000) reported that qEEG findings influenced therapeutic management with decisive decisions on many occasions.

Figure 10 is an example of construct validity of the qEEG in the measurement of TBI in which correlations of MRI were used to test the null hypothesis = 0, about damage to the average concentration of ionic channels in a volume of cortex that produces EEG (Thatcher, Biver, Camacho, et al., 1998; Thatcher, Biver, et al., 2001; Thatcher, Biver, McAlaster, et al., 1998).

In Figure 10, construct validity of qEEG was tested by examining the hypothesized relationship between the integrity of gray matter membranes using the MRI and the amplitude and coherence of the EEG. The hypothesis predicted reduced connectivity and a decline in amplitude of the EEG related to decreased integrity of neural membranes. The results of the construct validity tests of the qEEG in TBI were shown as valid as reported in peer-reviewed publications (Thatcher, Biver, Camacho, et al., 1998; Thatcher, Biver, et al., 2001; Thatcher, Biver, McAlaster, et al., 1998). These same studies also tested content validity by correlating the independent MRI measures with selected qEEG measures. Finally, predictive validity was also tested by correlations with neuropsychological test scores, which covaried with both the qEEG and the MRI in a

FIGURE 10. An example of construct validity of the quantitative EEG (QEEG) to correlate with the magnetic resonance imaging (MRI) in the estimate of traumatic brain injury. *Note.* Source: Adapted from Thatcher, Biver, Camacho, et al. (1998) and Thatcher, Biver, McAlaster, et al. (1998), reprinted with permissions.



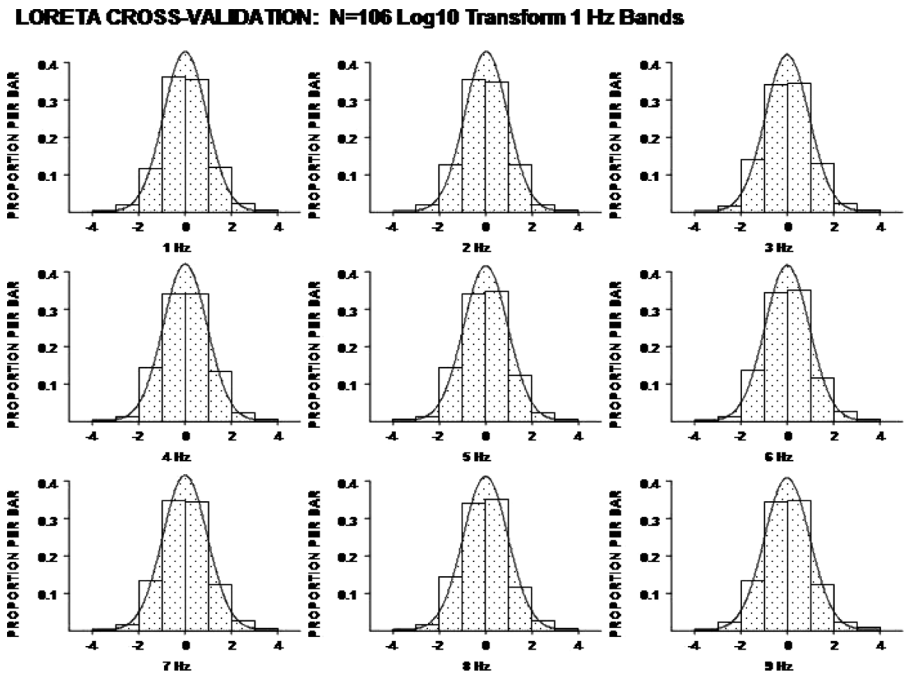
predictable manner. A similar cross-validation study was performed by Korn, Golan, Melamed, Pascual-Marqui, and Friedman (2005) showing significant correlations between LORETA current source activity and SPECT scans in TBI patients.

VALIDITY OF A LORETA qEEG NORMATIVE DATABASE

There are more than 795 peer-reviewed journal articles on the use of LORETA for the identification of the three-dimensional sources of the EEG in many different clinical populations. Because different regions in the brain are involved in different functional systems, the reliability and validity of LORETA is established by the degree to which accurate localization is demonstrated and by repeatability across participants and across experiments. It is easy to demonstrate content validity by showing that different samples of EEG yield the same localization and/or that a particular local event in the EEG corresponds to an expected source of that event. For example, alpha spindles maximum in O1 and O2 are localized to the occipital cortex by LORETA and not somewhere unexpected (e.g., right temporal lobe, etc.).

The reliability and validity of LORETA source localization can be demonstrated using mathematical simulations and standard tests in Systat and SPSS as well as by determining that the distribution of current sources is represented by a Gaussian distribution. To the extent the individual variables are Gaussian distributed, the mathematics of parametric statistics are valid and useful. Thus, the first step in evaluating the validity of a LORETA normative database is to test and establish that the current sources are Gaussian distributed. Figure 11 shows the distribution of current source densities after $\log_{10}$ transform in 1 Hz frequency bands from 1 to 9 Hz. It also shows that a reasonable approximation to a Gaussian distribution was achieved by the $\log_{10}$ transform. The distribution of current source densities with the Box-Cox transform were essentially

FIGURE 11. The distribution of the *Z* scores of the current source density LORETA values at 1 Hz resolution. The *y*-axis is the number or count and the *x*-axis is the *Z* score, defined as the mean—each value in each of the 2,394 pixels divided by the standard deviation. *Note.* Source: Thatcher, North, and Biver (2005c), reprinted with permission.



the same as for the $\log_{10}$ and therefore are not displayed.

Standard cross-validation methods can also be used to establish reliability and validity. That is, the classification of normal subjects as not being normal by a leave-one-out cross-validation procedure or by a direct cross-validation procedure provides an estimate of the false positives and false negatives of the normative database.

Table 4 shows the skewness and kurtosis of the $\log_{10}$ transformed data and the percentages of *Z* scores at ± 2 *SD* and ± 3 *SD* for each of the 1 Hz frequency bands for the eyes-closed condition for linked ears reference. The sensitivities ranged from 95.64% at 2 *SD* to 99.75% at 3 *SD*. Average skewness is 0.29 and average kurtosis is 0.68 Thus, gaussianity can be approximated at a frequency resolution of 1 Hz. The results of a leave-one-out cross-validation are published in Thatcher et al. (2005b, 2005c).

Another method of establishing content and construct validity of a LORETA normative database is to test the accuracy of the database using patients with confirmed pathologies where the location of the pathology is known by other imaging methods (e.g., CT scan or MRI or PET, etc.). Validity is estimated by the extent that there is a high correspondence between the location of the confirmed pathology and the location of the three-dimensional sources of the EEG that correspond to the location of the pathology. Here is a partial list of studies showing concordance validity with fMRI and LORETA (Brookings, Ortigue, Grafton, & Carlson, 2009; Esposito, Aragri, et al., 2009; Esposito, Mulert, et al., 2009; Mobascher, Brinkmeyer, Warbrick, Musso, Wittsack, Saleh, et al., 2009; Mobascher, Brinkmeyer, Warbrick, Musso, Wittsack, Stoermer, et al., 2009; Schulz et al., 2008; Yoshioka et al., 2008) and between PET and LORETA (Horacek et al., 2007; Hu et al., 2006; Kopeček et al., 2005; Pizzagalli et al., 2004; Tišlerová, Horáček, Brunovský, & Kopeček, 2005; Zumsteg, Wennberg, Treyer, Buck, & Wieser, 2005) and between SPECT and LORETA (Korn et al., 2005).

TABLE 4. Results of a leave-one-out cross-validation of a LORETA normative database.

LORETA CROSS-VALIDATION: NORMs_LE_EC_ADULTs_z(LOG10(LORETA))_n = 106										
			1 Standard Deviations		2 Standard Deviations		3 Standard Deviations			
Frequency	Skewness	Kurtosis	z < −1 SD	z > 1 SD	z < −2 SD	z > 2 SD	z < −3 SD	z > 3 SD	SEN 2 SD	SEN 3 SD
FREQ_1	0.08	0.84	13.76%	14.60%	2.22%	2.52%	0.51°.	0.45%	95.64%	99.75%
FREQ_2	0.19	0.39	14.47%	15.20%	1.79%	2.44%	0.21%	0.46%	95.64%	99.75%
FREQ_3	0.24	0.14	15.79%	15.62%	1.66%	2.54%	0.03%	0.35%	95.64%	99.74%
FREQ_4	0.29	0.16	15.73%	15.83%	1.21%	2.63%	0.02%	0.42%	95.64%	99.74%
FREQ_5	0.27	0.17	15.79%	15.13%	1.44%	2.85%	0.00%	0.41%	95.64%	99.74%
FREQ_6	0.33	0.32	15.39%	14.77%	1.62%	3.11%	0.01%	0.54%	95.64%	99.75%
FREQ_7	0.33	0.30	15.25%	15.44%	1.79%	2.99%	0.00%	0.50%	95.64%	99.75%
FREQ_8	0.38	0.47	15.90%	14.69%	1.72%	3.19%	0.00%	0.44%	95.65%	99.75%
FREQ_9	0.21	0.29	15.66%	15.22%	2.20%	2.88%	0.00%	0.32%	95.65%	99.74%
FREQ_10	0.17	0.23	14.52%	16.43%	2.18%	2.38%	0.00%	0.30%	95.64%	99.74%
FREQ_11	0.64	1.51	12.62%	15.23%	1.65%	3.21%	0.00%	0.87%	95.64%	99.75%
FREQ_12	0.29	0.80	13.91%	15.06%	1.88%	3.17%	0.23%	0.44%	95.65%	99.75%
FREQ_13	0.22	0.82	13.97%	15.20%	1.83%	2.93%	0.44%	0.45%	95.64%	99.75%
FREQ_14	0.25	0.82	14.08%	14.75%	1.79%	3.05%	0.34%	0.46%	95.64%	99.75%
FREQ_15	0.20	0.78	13.97%	15.10%	2.03%	2.86%	0.41%	0.46%	95.65%	99.75%
FREQ_16	0.21	0.82	14.05%	14.93%	2.02%	2.88%	0.37%	0.43%	95.65%	99.75%
FREQ_17	0.27	0.97	14.02%	14.41%	1.84%	2.92%	0.31%	0.61%	95.64%	99.75%
FREQ_18	0.27	0.94	13.85%	13.86%	1.75%	3.56%	0.38%	0.52%	95.65%	99.75%
FREQ_19	0.30	0.90	13.20%	14.21%	1.79%	3.54%	0.31%	0.54%	95.65%	99.75%
FREQ_20	0.27	0.75	13.26%	15.16%	1.74%	3.26%	0.43%	0.33%	95.65%	99.75%
FREQ_21	0.34	0.73	13.38%	15.47%	1.71%	3.14%	0.19%	0.54%	95.64%	99.75%
FREQ_22	0.33	0.73	13.93%	14.90%	1.62%	3.33%	0.24%	0.57%	95.65%	99.75%
FREQ_23	0.25	0.68	13.34%	15.09%	1.98%	3.34%	0.31%	0.52%	95.65%	99.75%
FREQ_24	0.30	0.75	13.71%	14.62%	1.81%	3.40%	0.34%	0.59%	95.65%	99.75%
FREQ_25	0.43	0.93	13.11%	14.77%	1.78%	3.39%	0.08%	0.77%	95.65%	99.75%
FREQ_26	0.42	0.94	12.78%	14.70%	1.74%	3.32%	0.14%	0.80%	95.65%	99.75%
FREQ_27	0.47	1.04	12.29%	14.51%	1.76%	3.30%	0.15%	0.83%	95.65%	99.75%
OVERALL	0.29	0.68	14.14%	15.00%	1.80%	3.04%	0.20%	0.51%	95.64%	99.75%

Note. Source: Thatcher, North, and Biver (2005c), reprinted with permission.

Figure 12 shows an example of the EEG from an epilepsy patient in which maximal epileptic discharges are present in the left temporal, left parietal, and left occipital regions. Content validity of LORETA is established by the fact that the maximum amplitude of epileptic activity was in the left temporal lobe lead (T5) at 3 Hz as measured by the FFT and the *Z* scores from the scalp surface. The sources were localized to Brodmann area 22 or left superior temporal gyrus and Brodmann area 13 of the left insular cortex.

LORETA is low resolution electromagnetic tomography (est. 2–4 cm resolution) so precise millimeter localization of epileptic foci is beyond the resolution of it. Nonetheless, verification of the surface EEG with three-dimensional source currents allows one to hypothesize the expected brain regions based on the surface EEG. In this case, the hypothesis, based upon the surface EEG, that there is a source in the left temporal regions (Brodmann areas were predicted before-hand) was confirmed.

Figure 13 (top) shows an example of the EEG from a TBI patient with a right hemisphere hematoma. The maximum amplitude of slow waves (1–6 Hz) was in the right prefrontal (C4), right parietal (P4), and right occipital (O2) regions as measured by the FFT and the *Z* scores from the scalp surface.

Figure 13 (bottom) shows the *Z* scores in LORETA slices in the right hemisphere hematoma patient as being consistent with the surface EEG deviation from normal because they were in the right hemisphere and near to the area of maximal damage. The maximum *Z* scores were present in the right postcentral gyrus at 5 Hz and were localized to Brodmann area 43 in the right postcentral gyrus as well as Brodmann areas 13 (right insular cortex) and 41 (right transverse temporal gyrus).

Figure 14 (top) shows an example of the EEG from a right hemisphere stroke patient. The maximum *Z* scores from the scalp EEG were in the right anterior frontal regions (F4 & Fp2) at 23 Hz. It can be seen that the maximum *Z* scores were present in the right

FIGURE 12. Top is the EEG from a patient with Left Temporal Lobe epilepsy where the maximum spike and waves are present in T5, O1, P3 and T3. The FFT power spectrum and the corresponding surface EEG *Z* scores are shown in the top right side. Bottom, are the left and right hemisphere displays of the maximal Z scores using LORETA. It can be seen that only the left temporal lobe has statistically significant *Z* values. Planned comparisons and hypothesis testing based on the frequency and location of maximal deviation from normal on the surface EEG are confirmed by the LORETA *Z* score normative analysis. *Source*: Thatcher, North, and Biver (2005c), reprinted with permission.

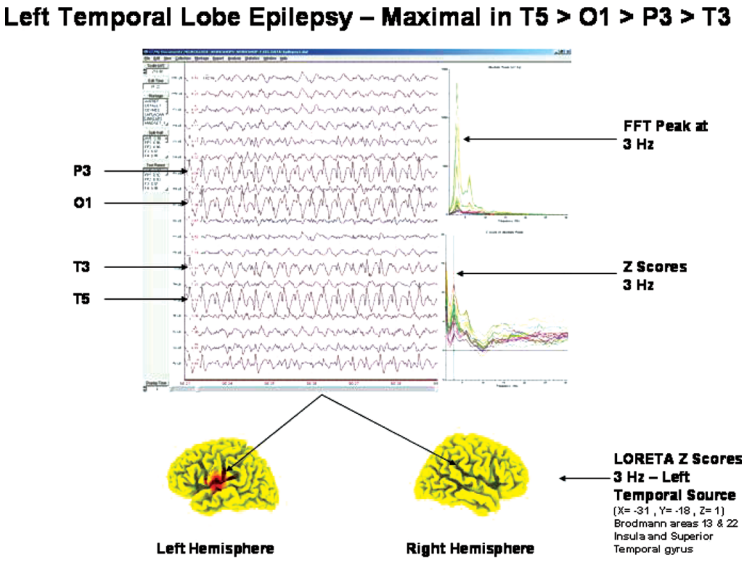
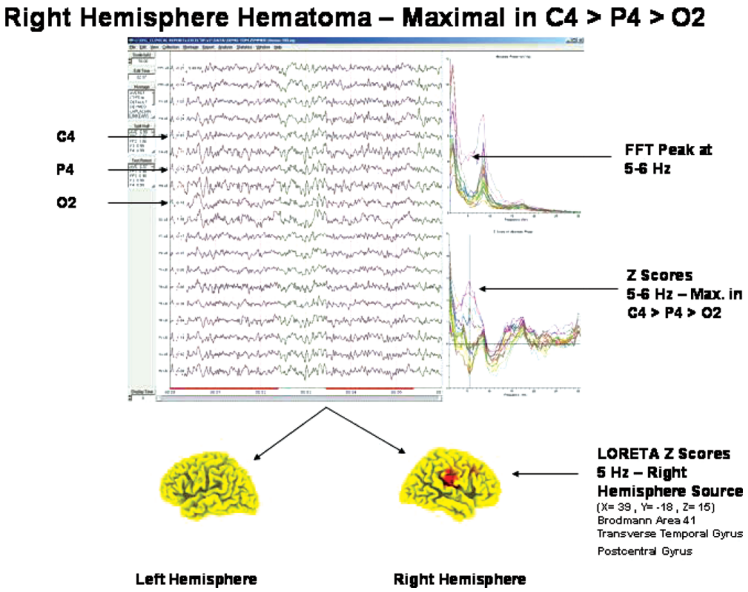


FIGURE 13. Top is the EEG from a patient with a right hemisphere hematoma where the maximum slows waves are present in C4, P4 and O2. The Fast Fourier Transform (FFT) power spectrum from 1 to 30 Hz and the corresponding *Z* scores of the surface EEG are shown in the right side of the EEG display. Bottom are the left and right hemisphere displays of the maximal *Z* scores using LORETA. It can be seen that only the right hemisphere has statistically significant *Z* values. Planned comparisons and hypothesis testing based on the frequency and location of maximal deviation from normal on the surface EEG are confirmed by the LORETA *Z* score normative analysis. Source: Thatcher, North, and Biver (2005c), reprinted with permission.



frontal regions at 23 Hz and the Key Institute Talairach Atlas were maximally localized to Brodmann area 9 (right inferior frontal gyrus) as well as Brodmann area 6 (right frontal precentral gyrus). This is another example of validation of a LORETA *Z* score normative database in which three-dimensional hypotheses are formed (and thus planned comparisons) based on the surface EEG and the hypothesis is then tested using LORETA.

CONSTRUCT VALIDITY OF A LORETA NORMATIVE DATABASE BASED ON THE SMOOTHNESS AT 1 HZ RESOLUTION AND REGIONS OF INTEREST

Figure 15 is a graph of the rank order of *Z* scores for different 1 Hz frequency bands from 1 to 10 Hz for the 2,394 current source values in the right hemisphere hematoma patient. A smooth distribution of *Z* scores with maxima near to the location of the confirmed injury is expected if parametric statistics using LORETA are valid, an example of construct validity. It can be seen that the rank ordering of the *Z* scores is smooth and well behaved at each 1 Hz frequency analysis with maximum *Z* score deviation at 2–6 Hz which is the same frequency band in which the surface EEG was most deviant from normal (see Figure 13). A smooth rank ordering of *Z* scores is expected if parametric statistical analysis is valid.

RELIABILITY DEFINED

Reliability is the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials. Researchers and clinicians would be

FIGURE 14. Top is the EEG from a patient with a right frontal lobe stroke where the maximum slows waves are present in F4 and Fp2. The FFT power spectrum from 1 to 30 Hz and the corresponding *Z* scores of the surface EEG are shown in the right side of the EEG display. Bottom are the left and right hemisphere displays of the maximal *Z* scores using LORETA. It can be seen that only the right hemisphere has statistically significant *Z* values. Planned comparisons and hypothesis testing based on the frequency and location of maximal deviation from normal on the surface EEG are confirmed by the LORETA Z score normative analysis. *Source*: Thatcher, North, and Biver (2005c), reprinted with permission.

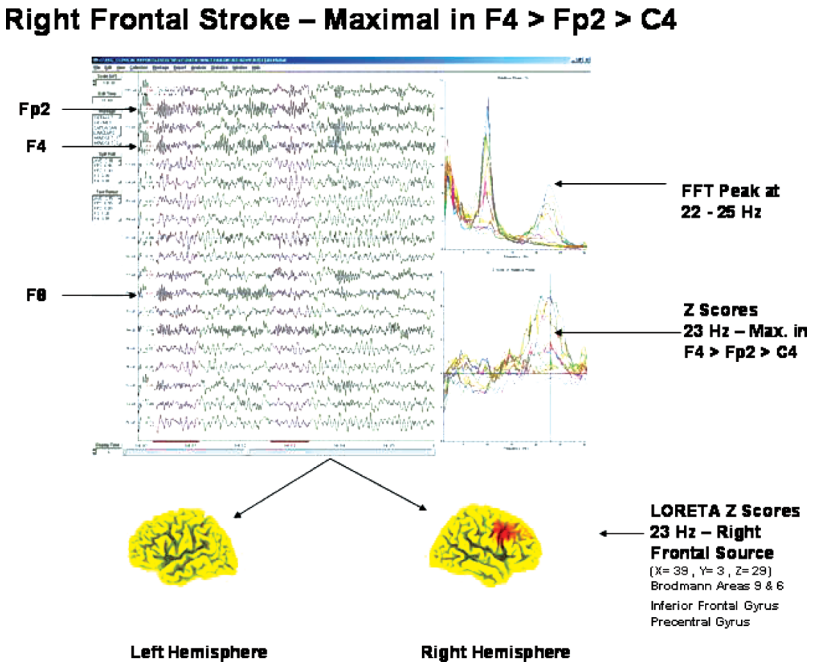
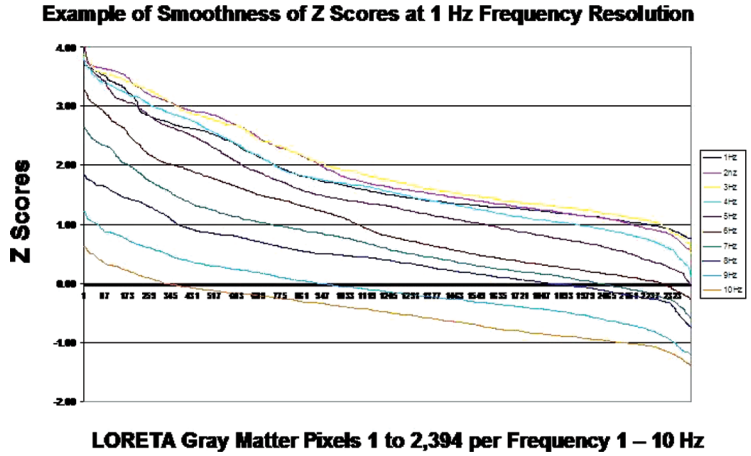


FIGURE 15. Evaluation of the smoothness of the *Z* scores in Figure 13 for frequencies 1 to 10 Hz. The LORETA current source values were rank-ordered for each single hertz frequency. The *y*-axis is *Z* scores and the *x*-axis is the number of gray matter pixels from 1 to 2,394. *Source*: Thatcher, North, and Biver (2005c), reprinted with permission.



unable to satisfactorily draw conclusions, formulate theories, or make claims about the generalizability of their research without the agreement of independent and replicable observations nor to be able to replicate research procedures, or use research tools and procedures that yield consistent measurements. The measurement of any phenomenon always contains a certain amount of chance error. The null hypothesis in any test of reliability is where reliability is 0, that is, repeated measurements of the same phenomenon never duplicate each other and they are not consistent from measurement to measurement. The Type I and Type II errors inherent in the reliability of a sample of digital EEG and/or qEEG can be measured in different ways. An acceptable level of reliability depends on the intended application of the method and on the tolerance of error.

There are various ways to measure reliability such as (a) the retest method (stability over time), (b) alternative-form method, (c) internal consistency, and (d) split-halves method (Carmines & Zeller, 1979). The particular method of computing reliability depends on the circumstances and/or personal choice. It is possible to have a measure that has high reliability but low validity, that is, one that is consistent in getting wrong information or is consistent in missing the mark. It is also possible for low reliability and low validity, that is, inconsistent and never on target. Test–retest reliability,'' also called ''stability reliability,'' is a commonly used method of reliability testing in qEEG and is generally defined as the agreement of measuring instruments over time. Alternative-form reliability is when different measures provide similar results; for example, EEG coherence and EEG phase lock duration or coherences versus comodulation, and so on. To determine stability, a measure or test is repeated on the same participants at different points in time. Results are compared and correlated with the initial test to give a measure of stability and to detect changes. The test–retest reliability statistic is a good method to detect drowsiness when comparing the beginning of the EEG recording to the end of a lengthy recording with eyes closed. For example, if there is no dramatic change in state between the beginning and end of the recording, then one would expect high test–retest reliability (e.g., >0.9). On the other hand, if a patient is drowsy or sleeping near the end of the recording, then one would expect the test–retest reliability between the beginning of the record to be low (e.g., <0.9).

RELIABILITY OF EEG AUTOPOWER SPECTRUM

The autopower spectrum is the part of the power spectrum that measures the amount of energy in a complex wave form at each frequency. The units are in microvolts squared per cycle per second or $\mu V^2/Hz$. Amplitude or magnitude is simply the square root of power and the same reliability measures are used for both power and amplitude. The scientific literature demonstrating high reliability (e.g., >0.9) of quantitative EEG is diverse and quite large and can be read by visiting the National Library of Medicine's database at https://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed; use the search terms ''EEG and Reliability'' and there are 368 citations, and a quick review of the abstracts shows that the vast majority, if not all, of these studies are qEEG studies and demonstrate high test–retest reliability of the qEEG. Next is a small but representative sample of some of the studies demonstrating high reliability with sample lengths as short as 20 s (Arruda et al., 1996; Burgess & Gruzelier, 1993; Chabot, Merkin, Wood, Davenport, & Serfontein, 1996; Corsi-Cabrera, Solis-Ortiz, & Guevara, 1997; Duffy, Hughes, Miranda, Bernad, & Cook, 1994; Fernández et al., 1993; Gasser, Bacher, & Steinberg, 1985; Gasser et al., 1987; Hamilton-Bruce, Boundy, & Purdie, 1991; Harmony et al., 1993; John, Prichep, & Easton, 1987; John, Prichep, Fridman, & Easton, 1988; Lund, Sponheim, Iacono, & Clementz, 1995; McEvoy, Smith, & Gevins, 2000; Näpflin, Wildi, & Sarnthein, 2007, 2008; Pollock, Schneider, & Lyness, 1991; Salinsky, Oken, & Morehead, 1991; Towers

& Allen, 2009; Van Albada, Rennie, & Robinson, 2007).

Gasser et al. (1985) concluded that "20 sec of activity are sufficient to reduce adequately the variability inherent in the EEG" (p. 312).

Salinski et al. (1991)) concluded, "Correlation coefficients for broad band features averaged 0.92 over the 5 min retest interval and 0.84 over the 12–16" and "coefficients based on 60 sec records were marginally higher than those of 40 or 20 sec records" (p. 382).

Corsi-Cabrera et al. (1997) concluded, "The within-subject stability was assessed calculating multiple correlation coefficients between all EEG features of the eleven sessions of each subject: R-values ranged from 0.85 to 0.97" (p. 382).

Pollock et al. (1991) concluded, "The generally higher reliabilities of absolute, as opposed to relative, amplitude measures render them preferable in clinical research" (p. 20).

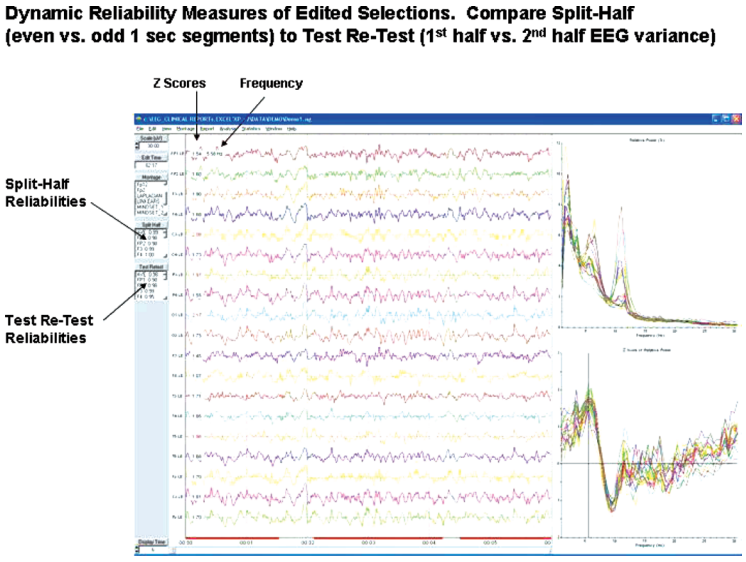
EEG spectral stability over a 1-year period was recently studied by Näpflin and colleagues with test–retest reliability greater than 0.9, and they concluded that qEEG intraindividual reliability is very high:

> Out of all 2400 pairwise comparisons 99.3% were correct, with sensitivity 87.5% and specificity 99.5%. The intra-individual stability is high compared to the inter-individual variation. Thus, interleaved EEG-fMRI measurements are valid. Furthermore, longitudinal effects on cognitive EEG can be judged against the intra-individual variability in subjects. (Näpflin et al., 2008, p. 2519)

A recent study by Van Albada et al. (2007) evaluated the variable contributions of "state" and "trait" by conducting test–retest reliability measures of the qEEG recorded from participant each week for 6 weeks and some participants for as long as 1 year and concluded, "About 95% of the maximum change in spectral parameters was reached within minutes of recording time, implying that repeat recordings are not necessary to capture the bulk of the variability in EEG spectra" (p. 279).

In general, the test–retest reliability of qEEG is an exponential function of sample length in which 20-s epochs are approximately 0.8 reliable, 40 s approx. 0.9 reliable, and 60-s asymptotes at approximately 0.95 reliability. It is easy to test the reliability of qEEG for one's self as shown in Figure 16.

FIGURE 16. An example of visual EEG traces, quantitative EEG, Split-Half reliabilities and test–retest reliabilities on the same screen at the same. *Note.* Panel to the left are the EEG traces, top right panel is the FFT power spectrum from 1 to 30 Hz and bottom right panel are *Z* scores from 1 to 30 Hz.



RELIABILITY OF EEG COHERENCE

As mentioned previously, coherence is itself a statistical measure of reliability because it is a measure of the stability of phase differences between two EEG time series. If the phase difference is unreliable, that is, phase differences are randomly changing from time sample to time sample, then coherence is 0. If the phase differences are unchanging, then coherence is 1. High test–retest reliability of EEG coherence has been reported over the years when coherence is correctly computed even though more statistical samples are often required to obtain statistical sufficiency. If regions of the brain are weakly coupled or disconnected, then coherence has low values within a subject as well as low test–reretest reliability across experiments and participants as expected. If regions of the brain are strongly coupled and coherence exhibits statistically significant values then coherence typically also exhibits high test–retest reliability (the greater the coherence then the more within-session and between-session reliability by definition). Adey, Walter, and Hendrix (1961) were among the first to measure the test–retest reliability of EEG coherence with values greater than 0.8. Subsequently, high retest reliability of EEG coherence (0.8–0.95) was reported by Chabot et al. (1996); Corsi-Cabrera, Galindo-Vilchis del-Río-Portilla, Arce, and Ramos-Loyo (2007); Gasser et al. (1987); Harmony et al. (1993); John (1977); John et al. (1987); Thatcher, Krause, and Hrybyk (1986), and Thatcher, Beaver, et al. (2003). Gudmondsson, Runarsson, Sigurdsson, Eiriksdottir, and Johnsen (2007) reported low test–retest reliability of coherence because of an invalid computation due to the use of an average reference. If the authors used a common reference and coherence was low (e.g., <0.2), then this means that two brain regions are reliably disconnected. If the reader finds any study that claims that coherence has low reliability, examine the Methods section and see if the authors used an average reference, a Laplacian reference, or ICA to create a new time series and if so, then dismiss the study because they used an invalid method of measuring coherence in the first place. Remember, reliability is irrelevant if the measure is not valid to begin with.

SUMMARY

The fact that qEEG meets high standards of reliability and validity is demonstrated by hundreds of peer-reviewed journal articles, a few of which are cited in this review. The critics of qEEG are those that rely solely on eyeball examination of the EEG traces and are biased against and opposed to the use of computers to improve the accuracy, validity, and reliability of the electroencephalogram (Nuwer, 1997). The American Academy of Neurology position paper (Nuwer, 1997) categorized qEEG as ''experimental'' for a wide range of clinical disorders with the blanket assertion that qEEG is ''unreliable.'' However, they did not cite any studies to refute the scientific literature that demonstrates high reliability and validity. It is the responsibility of those that use qEEG technology to respond to false claims by citing facts and citing the scientific literature when ever possible.

REFERENCES

Adey, W. R., Walter, D. O., & Hendrix, C. E. (1961). Computer techniques in correlation and spectral analyses of cerebral slow waves during discriminative behavior. *Experimental Neurology*, *3*, 501–524.

Arruda, J. E., Weiler, M. D., Valentino, D., Willis, W. G., Rossi, J. S., Stern, R. A., et al. (1996). A guide for applying principal-components analysis and confirmatory factor analysis to quantitative electroencephalogram data. *International Journal of Psychophysiology*, *23*(1–2), 63–81.

Begić, D., Mahnik-Milos, M., & Grubisin, J. (2009). EEG characteristics in depression, ''negative'' and ''positive'' schizophrena. *Psychiatria Danubina*, *21*, 579–584.

Benbadis, S. R., LaFrance, W. C., Jr., Papandonatos, G. D., Korabathina, K., Lin, K., & Kraemer, H. C. (2009). Interrater reliability of EEG-video monitoring. *Neurology*, *15*, 843–846.

Breakspear, M., & Terry, J. R. (2002). Nonlinear interdependence in neural systems: motivation, theory and relevance. *Int. J. Neurosci.*, *112*(10), 1263–1284.

Brookings, T., Ortigue, S., Grafton, S., & Carlson, J. (2009). Using ICA and realistic BOLD models to obtain joint EEG/fMRI solutions to the problem of source localization. *NeuroImage*, *44*, 411–420.

Burgess, A., & Gruzelier, J. (1993). Individual reliability of amplitude distribution in topographical mapping of EEG. *Electroencephalography and Clinical Neurophysiology*, *86*, 219–223.

Buzea, C. E. (1995). Understanding computerized EEG monitoring in the intensive care unit. *Journal of Neuroscience Nursing*, *27*, 292–297.

Buzsaki, G. (2006). *Rhythms of the brain*. Cambridge, MA: Oxford University Press.

Cao, C., Tutwiler, R. L., & Slobounov, S. (2008). Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal support vector machine. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*, 327–350.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Chabot, R., Merkin, H., Wood, L., Davenport, T., & Serfontein, G. (1996). Sensitivity and specificity of QEEG in children with attention deficit or specific developmental learning disorders. *Clinical Electroencephalography*, *27*, 34–36.

Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., & Evans, A. C. (2008). Revealing Modular architecture of human brain structural networks by using cortical thickness from MRI. *Cerebral Cortex*, *18*, 2374–2381.

Claassen, J., Baeumer, T., & Hansen, H. C. (2000). Continuous EEG for monitoring on the neurological intensive care unit. New applications and uses for therapeutic decision making. *Nevenarzt*, *71*, 813–821.

Cooper, R., Osselton, J. W., & Shaw, J. G. (1974). *EEG technology*. London: Butterworth.

Corsi-Cabrera, M., Galindo-Vilchis, L., del-Río-Portilla, Y., Arce, C., & Ramos-Loyo, J. (2007). Within-subject reliability and inter-session stability of EEG power and coherent activity in women evaluated monthly over nine months. *Clinical Neurophysiology*, *11*, 9–21.

Corsi-Cabrera, M., Solis-Ortiz, S., & Guevara, M. A. (1997). Stability of EEG inter- and intrahemispheric correlation in women. *Electroencephalography and Clinical Neurophysiology*, *102*, 248–255.

Cronbach, L. J. (1977). Test validation. In R. Thorndike (Ed.), *Educational measurement*, (pp. 443–507). Washington, DC: American Council on Education.

Duffy, F. H., Hughes, J. R., Miranda, F., Bernad, P., & Cook, P. (1994). Status of quantitative EEG (QEEG) in clinical practice, 1994. *Clinical Electroencephalography*, *25*, 6–22.

Esposito, F., Aragri, A., Piccoli, T., Tedeschi, T., Goebel, T., & Di Salle, F. (2009). Distributed analysis of simultaneous EEG-fMRI time-series: Modeling and interpretation issues. *Magnetic Resonance Imaging*, *27*, 1120–1130.

Esposito, F., Mulert, C., & Goebel, R. (2009). Combined distributed source and single-trial EEG-fMRI Modeling: Application to effortful decision making processes. *NeuroImage*, *47*, 112–121.

Essl, M., & Rappelsberger, P. (1998). EEG coherence and reference signals: Experimental results and mathematical explanations. *Medical and Biological Engineering and Computing*, *36*, 399–406.

Fabregas, N., Gamus, P. L., Valero, R., Carrero, E. J., Salvador, L., Zavala, E., et al. (2004). Can bispectral index monitoring predict recovery of consciousness in patients with severe brain injury? *Anesthesiology*, *101*, 43–51.

Fernández, T., Harmony, T., Rodríguez, M., Reyes, A., Marosi, E., & Bernal, J. (1993). Test–retest reliability of EEG spectral parameters during cognitive tasks: I. Absolute and relative power. *International Journal of Neuroscience*, *68*, 255–261.

Freeman, W. J., Burke, B. C., & Homes, M. D. (2003). Aperiodic phase re-setting in scalp EEG of beta-gamma oscillations by state transitions at alpha-theta rates. *Human Brain Mapping*, *19*, 248–272.

Freeman, W. J., & Rogers, L. J. (2002). Fine temporal resolution of analytic phase reveals episodic synchronization by state transitions in gamma EEGs. *Journal of Neurophysiology*, *87*, 937–945.

Gasser, T., Bacher, P., & Steinberg, H. (1985). Test–retest reliability of spectral parameters of the EEG. *Electroencephalography and Clinical Neurophysiology*, *60*, 312–319.

Gasser, T., Jennen-Steinmetz, C., & Verleger, R. (1987). EEG coherence at rest and during a visual task in two groups of children. *Electroencephalogr Clin Neurophysiol.*, *67*(2), 151–158.

Gevensleben, H., Holl, B., Albrecht, B., Schlamp, D., Kratz, O., Studer, P., et al. (2009). Distinct EEG effects related to neurofeedback training in children with ADHD: A randomized controlled trial. *International Journal of Psychophysiology*, *74*, 149–157.

Grass, A. M., & Gibbs, F. A. (1938). A Fourier transform of the electroencephalogram. *Journal of Neurophysiology*, *1*, 521–526.

Gudmundsson, S., Runarsson, T. P., Sigurdsson, S., Eiriksdottir, G., & Johnsen, K. (2007). Reliability of quantitative EEG features. *Clinical Neurophysiology*, *118*, 2162–2171.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol*, *6*, e159.

Hamilton-Bruce, M. A., Boundy, K. L., & Purdie, G. H. (1991). Interoperator variability in quantitative electroencephalography. *Clinical and Experimental Neurology*, *28*, 219–224.

Hannesdóttir, D. K., Doxie, J., Bell, M. A., Ollendick, T. H., & Wolfe, C. D. (2010). A longitudinal study of emotion regulation and anxiety in middle childhood: Associations with frontal EEG asymmetry in early childhood. *Developmental Psychobiology*, *52*, 197–204.

Harmony, T., Fernandez, T., Rodriguez, M., Reyes, A., Marosi, E., & Bernal, J. (1993). Test–retest reliability of EEG spectral parameters during cognitive tasks: II. *Coherence*, *68*, 263–271.

Hayes, W. L. (1973). *Statistics for the social sciences*. New York: Holt, Reinhart and Winston.

He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., et al. (2009). Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS ONE*, *4*, e5226.

Horacek, J., Brunovsky, M., Novak, T., Skrdlantova, L., Klirova, M., Bubenikova-Valesova, V., et al. (2007). Effect of low-frequency rTMS on electromagnetic tomography (LORETA) and regional brain metabolism (PET) in schizophrenia patients with auditory hallucinations. *Neuropsychobiology*, *55*, 132–142.

Hu, J., Tian, J., Yang, L., Pan, X., & Liu, J. (2006). Combination of PCA and LORETA for sources analysis of ERP data: An emotional processing study. Progress in biomedical optics and imaging. *Proceedings of SPIE*, 6143 II.

Hughes, J. R., & John, E. R. (1999). Conventional and quantitative electroencephalography in psychiatry. *Neuropsychiatry*, *11*, 190–208.

Hyllienmark, L., & Amark, P. (2007). Continuous EEG monitoring in a paediatric intensive care unit. *European Journal of Paediatric Neurology*, *11*, 70–75.

John, E. R. (1977). *Functional neuroscience, Vol. II: Neurometrics. Quantitative electrophysiological analyses* (E. R. John & R. W. Thatcher, Eds.). Hillsdale, NJ: Erlbaum.

John, E. R., Prichep, L. S., & Easton, P. (1987). Normative data banks and neurometrics: Basic concepts, methods and results of norm construction. In A. Remond (Ed.), *Handbook of electroencephalography and clinical neurophysiology, Vol. III. Computer analysis of the EEG and other neurophysiological signals* (pp. 449–495). Amsterdam: Elsevier.

John, E. R., Prichep, L. S., Fridman, J., & Easton, P. (1988). Neurometrics: Computer assisted differential diagnosis of brain dysfunctions. *Science*, *293*, 162–169.

Jordan, K. G. (1993). Continuous EEG and evoked potential monitoring in the neuroscience intensive care unit. *Journal of Clinical Neurophysiology*, *10*, 445–475.

Kamiński, M., & Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, *65*, 203–210.

Kamiński, M., Blinowska, K. J., & Szelenberger, W. (1997). Topographic analysis of coherence and propagation of EEG activity during sleep wakefulness. *EEG and Clinical Neurophysiology*, *102*, 216–227.

Kane, N. M., Moss, T. H., Curry, S. H., & Butler, S. R. (1998). Quantitative electroencephalographic evaluation of non-fatal and fatal traumatic coma. *Electroencephalography and Clinical Neurophysiology*, *106*, 244–250.

Kopeček, M., Brunovský, M., Bareš, M., Španiel, F., Novák, T., Dockery, C., et al. (2005). Regional cerebral metabolic abnormalities in individual patients with nonquantitative 18FDG PET and qEEG (LORETA). *Psychiatrie*, *9*(Suppl. 3), 56–63.

Korn, A., Golan, H., Melamed, I., Pascual-Marqui, R., & Friedman, A. (2005). Focal cortical dysfunction and blood-brain barrier disruption in patients with Postconcussion syndrome. *Journal of Clinical Neurophysiology*, *22*, 1–9.

Kornmuller, A. E. (1937). *Die bioelektrischer erseheinungen der hirnrindenfelder*. Leipzig, Germany: Thieme.

Korzeniewska, A., Mańczak, M., Kamiński, M., Blinowska, K., & Kasicki, S. (2003). Determination of information flow direction between brain structures by a modified Directed Transfer Function method (dDTF). *Journal of Neuroscience Methods*, *125*, 195–207.

Lachaux, J.-P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., & Varela, F. J. (2000). Studying single-trials of phase synchronous activity in the brain. *International Journal of Bifurcation and Chaos*, *10*, 2429–2439.

Leon-Carrion, J., Martin-Rodriguez, J. F., Damas-Lopez, J., Martin, J. M. B., & Dominguez-Morales, M. R. (2008a). Brain function in the minimally conscious state: A quantitative neurophysiological study. *Clinical Neurophysiology*, *119*, 1506–1514.

Leon-Carrion, J., Martin-Rodriguez, J. F., Damas Lopez, J. Y., Martin, J. M. B., & Dominguez-Morales, M. (2008b). A QEEG index of level of functional dependence for people sustaining acquired brain injury: The Seville Independence Index (SINDI). *Brain Injury*, *22*, 61–74.

Lund, T. R., Sponheim, S. R., Iacono, W. G., & Clementz, B. A. (1995). Internal consistency

reliability of resting EEG power spectra in schizophrenic and normal subjects. *Psychophysiology*, *32*, 66–71.

Luria, A. (1973). *The working brain: An introduction to neuropsychology*. Baltimore: Penguin.

Malone, A., Ryan, C. A., Fitzgerald, A., Burgoyne, L., Connolly, S., & Boylan, G. B. (2009). Interobserver agreement in neonatal seizure identification. *Epilepsia*, *50*, 2097–2101.

Mazaheri, A., Coffey-Corina, S., Mangun, G. R., Bekker, E. M., Berry, A. S., & Corbett, B. A. (2010). Functional disconnection of frontal cortex and visual cortex in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *67*, 617–623.

McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test–retest reliability of cognitive EEG. *Clinical Neurophysiology*, *111*, 457–463.

Mesulam, M.-M. (2000). *Principles of behavioral and cognitive neurology, 2nd edition*. Cambridge, MA: Oxford University Press.

Mobascher, A., Brinkmeyer, J., Warbrick, T., Musso, F., Wittsack, H. J., Saleh, A., et al. (2009). Laser-evoked potential P2 single-trial amplitudes covary with the fMRI BOLD response in the medial pain system and interconnected subcortical structures. *NeuroImage*, *45*, 917–926.

Mobascher, A., Brinkmeyer, J., Warbrick, T., Musso, F., Wittsack, H. J., & Stoermer, R. (2009). Fluctuations in electrodermal activity reveal variations in single trial brain responses to painful laser stimuli—A fMRI/EEG study. *NeuroImage*, *44*, 1081–1092.

Näpflin, M., Wildi, M., & Sarnthein, J. (2007). Test–retest reliability of resting EEG spectra validates a statistical signature of persons. *Clinical Neurophysiology*, *118*, 2519–2524.

Näpflin, M., Wildi, M., & Sarnthein, J. (2008). Test–retest reliability of EEG spectra during a working memory task. *NeuroImage*, *43*, 687–693.

Niedermeyer, E., & Lopes Da Silva, F. (2005). *Electroencephalography: Basic principles, clinical applications and related fields* (5th ed.). Baltimore: Williams & Wilkins.

Nunez, P. (1981). *Electrical fields of the brain*. Cambridge, MA: Oxford University Press.

Nunez, P. (1995). *Neocortical dynamics and human EEG rhythms*. New York: Oxford University Press.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nuwer, M. R. (1997). Assessment of digital EEG, quantitative EEG and EEG brain mapping report of the American Academy of Neurology and the American Clinical Neurophysiology Society. *Neurology*, *49*, 277–292.

Ommaya, A. K. (1968). The mechanical properties of tissues of the nervous system. *Journal of Biomechanics*, *2*, 1–12.

Ommaya, A. K. (1995). Head injury mechanisms and the concept of preventive management: A review and critical synthesis. *Journal of Neurotrauma*, *12*, 527–546.

Ommaya, A. K., & Hirsch, A. E. (1971). Tolerances for cerebral concussion from head impact and whiplash in primates. *Journal of Biomechanics*, *4*, 13–21.

Piccinelli, P., Viri, M., Zucca, C., Borgatti, R., Romeo, A., Giordano, L., et al. (2005). Inter-rater reliability of the EEG reading in patients with childhood idiopathic epilepsy. *Epilepsy Research*, *66*, 195–198.

Prichep, L. S. (2005). Use of normative databases and statistical methods in demonstrating clinical utility of QEEG: Importance and cautions. *Clinical EEG Neuroscience*, *36*, 82–87.

Pizzagalli, D. A., Oakes, T. R., Fox, A. S., Chung, M. K., Larson, C. L., Abercrombie, H. C., et al. (2004). Functional but not structural subgenual prefrontal cortex abnormalities in melancholia. *Molecular Psychiatry*, *9*, 393–405.

Pollock, V. E., Schneider, L. S., & Lyness, S. A. (1991). Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. *Electroencephalography and Clinical Neurophysiology*, *79*, 20–26.

Randolph, C., & Miller, M. H. (1998). EEG and cognitive performance following closed head injury. *Neuropsychobiology*, *20*, 43–50.

Rappelsberger, P. (1989). The reference problem and mapping of coherence: A simulation study. *Brain Topography*, *2*, 63–72.

Reid, M. S., Prichep, L. S., Ciplet, D., O'Leary, S., Tom, M., Howard, B., et al. (2003). Quantitative electroencephalographic studies of cue-induced cocaine craving. *Clinical Electroencephalography*, *34*, 110–123.

Salinsky, M. C., Oken, B. S., & Morehead, L. (1991). Test–retest reliability in EEG frequency analysis. *Electroencephalography and Clinical Neurophysiology*, *79*, 382–392.

Sauseng, P., & Klimesch, W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience and Biobehavioral Review*, *32*, 1001–1013.

Scheurer, M. L. (2002). Continuous EEG monitoring in the intensive care unit. *Epilepsia*, *43*(Suppl. 3), 114–127.

Schulz, E., Maurer, U., van der Mark, S., Bucher, K., Brem, S., Martin, E., et al. (2008). Impaired semantic processing during sentence reading in children with dyslexia: Combined fMRI and ERP evidence. *NeuroImage*, *41*, 153–168.

Seshia, S. S., Young, G. B., & Zifkin, B. G. (2008). Guidelines for visual-sensitive EEG testing. *Canadian Journal of Neurology Science*, *35*, 133–139.

Shields, D. C., Liephart, J. W., McArthur, D. L., Vespa, P. M., LeVan Quyen, M., Martinerie, J., et al. (2007). Cortical synchrony changes detected by scalp electrode EEG as traumatic brain injury patients emerge from coma. *Surgical Neurology*, *67*, 354–359.

Siegle, G. J., Condray, R., Thase, M. E., Keshavan, M., & Steinhauer, S. R. (2010). Sustained gamma-band EEG following negative words in depression and schizophrenia. *International Journal of Psychophysiology*, *75*, 107–118.

Thatcher, R. W. (2008). EEG evaluation of traumatic brain injury and EEG biofeedback treatment. In T. Budzinsky, H. Budzinsky, J. Evans, & A. Abarbanel (Eds.), *Introduction to QEEG and neurofeedback: Advanced theory and applications*. (pp. 32–52). San Diego, CA: Academic.

Thatcher, R. W., Biver, C., Camacho, M., McAlaster, R., & Salazar, A. M. (1998). Biophysical linkage between MRI and EEG amplitude in traumatic brain injury. *NeuroImage*, *7*, 352–367.

Thatcher, R. W., Biver, C. L., Gomez-Molina, J. F., North, D., Curtin, R., Walker, R. W., et al. (2001). Estimation of the EEG power spectrum by MRI T2 relaxation time in traumatic brain injury. *Clinical Neurophysiology*, *112*, 1729–1745.

Thatcher, R. W., Biver, C., McAlaster, R., & Salazar, A. M. (1998). Biophysical linkage between MRI and EEG coherence in traumatic brain injury. *NeuroImage*, *8*, 307–326.

Thatcher, R. W., Biver, C., & North, D. (2003). Quantitative EEG and the frye and daubert standards of admissibility. *Clinical Electroencephalography*, *34*(2), 39–53.

Thatcher, R. W., Cantor, D. S., McAlaster, R., Geisler, F., & Krause, P. (1991). Comprehensive predictions of outcome in closed head injury: The development of prognostic equations. *Annals New York Academy of Sciences*, *620*, 82–104.

Thatcher, R. W., Krause, P., & Hrybyk, M. (1986). Corticocortical association fibers and EEG coherence: A two compartmental model. *Electroencephalography and Clinical Neurophysiology*, *64*, 123–143.

Thatcher, R. W., North, D., & Biver, C. (2005a). EEG and intelligence: Univariate and multivariate comparisons between EEG coherence, EEG phase delay and power. *Clinical Neurophysiology*, *116*, 2129–2141.

Thatcher, R. W., North, D., & Biver, C. (2005b). EEG inverse solutions and parametric vs. non-parametric statistics of Low Resolution Electromagnetic Tomography (LORETA). *Clinical EEG and Neuroscience*, *36*, 1–9.

Thatcher, R. W., North, D., & Biver, C. (2005c). Evaluation and validity of a LORETA normative EEG database. *Clinical EEG and Neuroscience*, *36*, 116–122.

Thatcher, R. W., North, D., & Biver, C. (2008a). Intelligence and EEG phase reset: A two-compartmental model of phase shift and lock. *NeuroImage*, *42*, 1639–1653.

Thatcher, R. W., North, D., & Biver, C. (2008b, January 24). Self-organized criticality and the development of EEG phase reset. *Human Brain Mapping.*

Thatcher, R. W., North, D., Curtin, R., Walker, R. A., Biver, C., Gomez, M., et al. (2001). An EEG severity index of traumatic brain injury. *Journal of Neuropsychiatry and Clinical Neuroscience*, *13*, 77–87.

Thatcher, R. W., North, D., Neurbrander, J., Biver, C. J., Cutler, S., & DeFina, P. (2009). Autism and EEG phase reset: Deficient GABA mediated inhibition in thalamo-cortical circuits. *Developmental Neuropsychology*, *34*, 780–800.

Thatcher, R. W., Walker, R. A., Biver, C., North, D., & Curtin, R. (2003). Quantitative EEG normative databases: Validation and clinical correlation. *Journal of Neurotherapy*, *7*, 87–122.

Thatcher, R. W., Walker, R. A., Gerson, I., & Geisler, F. (1989). EEG discriminant analyses of mild head trauma. *EEG and Clinical Neurophysiology*, *73*, 93–106.

Thornton, K. (1999). Exploratory investigation into mild brain injury and discriminant analysis with high frequency bands (32–64 Hz). *Brain Injury*, *13*, 477–488.

Thornton, K., & Carmody, D. P. (2005). Electroencephalogram biofeedback for reading disability and traumatic brain injury. *Child and Adolescent Psychiatry Clin North America*, *14*, 137–162.

Tišlerová, B., Horáček, J., Brunovský, M., & Kopeček, M. (2005). 18FDG PET and qEEG imaging of hebephrenic schizophrenia. A case study. Hebefrenní schizofrenie v obraze 18FDG PET a qEEG. *Kazuistika*, *9*, 144–149.

Towers, D. N., & Allen, J. J. (2009). A better estimate of the internal consistency reliability of frontal EEG asymmetry scores. *Psychophysiology*, *46*, 132–142.

Van Albada, S. J., Rennie, C. J., & Robinson, P. A. (2007). Variability of model-free and model-based quantitative measures of EEG. *Journal of Integrative Neuroscience*, *6*, 279–307.

van Dongen-Boomsma, M., Lansbergen, M. M., Bekker, E. M., Kooij, J. J., van der Molen, M., Kenemans, J. L., et al. (2010). Relation between resting EEG to cognitive performance and clinical symptoms in adults with attention-deficit/hyperactivity disorder. *Neuroscience Letters*, *18*, 102–106.

Velikova, S., Locatelli, M., Insacco, C., Smeraldi, E., Comi, G., & Leocani, L. (2010). Dysfunctional

brain circuitry in obsessive-compulsive disorder: Source and coherence analysis of EEG rhythms. *NeuroImage*, *49*, 977–983.

Woody, R. H. (1968). Inter-judge reliability in clinical EEG. *Journal of Clinical Psychology*, *24*, 251–261.

Woody, R. H. (1966). Intra-judge reliability in clinical EEG. *Journal of Clinical Psychology*, *22*, 150–159.

Yoshioka, T., Toyama, K., Kawato, M., Yamashita, O., Nishina, S., Yamagishi, N., et al. (2008). Evaluation of hierarchical Bayesian method through retinotopic brain activities reconstruction from fMRI and MEG signals. *NeuroImage*, *42*, 1397–1413.

Zumsteg, D., Wennberg, R. A., Treyer, V., Buck, A., & Wieser, H. G. (2005). H215O or 13NH3 PET and electromagnetic tomography (LORETA) during partial status epilepticus. *Neurology*, *65*, 1657–1660.