International Journal of Psychophysiology 106 (2016) 87–96

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Test-retest reliability of a single-channel, wireless EEG system

Jeffrey M. Rogers [a,*], Stuart J. Johnstone [b], Anna Aminov [a], James Donnelly [c], Peter H. Wilson [a]

[a] School of Psychology, Australian Catholic University, Australia
[b] School of Psychology, University of Wollongong, Australia
[c] School of Health and Human Sciences, Southern Cross University, Australia

ARTICLE INFO

Article history:
Received 23 March 2016
Received in revised form 14 June 2016
Accepted 15 June 2016
Available online 16 June 2016

Keywords:
EEG
Test-retest reliability
Portable
NeuroSky ThinkGear
Adults
Youth

ABSTRACT

Recording systems to acquire electroencephalogram (EEG) data are traditionally lab-based. However, there are shortcomings to this method, and the ease of use and portability of emerging wireless EEG technologies offer a promising alternative. A previous validity study demonstrated data derived from a single-channel, wireless system (NeuroSky ThinkGear, San Jose, California) is comparable to EEG recorded from conventional lab-based equipment. The current study evaluated the reliability of this portable system using test-retest and reliable change analyses. Relative power (RP) of delta, theta, alpha, and beta frequency bands was derived from EEG data obtained from a single electrode over FP1 in 19 healthy youth (10–17 years old), 21 healthy adults (18–28 years old), and 19 healthy older adults (55–79 years old), during eyes-open, eyes-closed, auditory oddball, and visual n-back conditions. Intra-class correlations (ICCs) and Coefficients of Repeatability (CRs) were calculated from RP data re-collected one-day, one-week, and one-month later. Participants' levels of mood and attention were consistent across sessions. Eyes-closed resting EEG measurements using the portable device were reproducible (ICCs 0.76–0.85) at short and longer retest intervals in all three participant age groups. While still of at least *fair* reliability (ICCs 0.57–0.85), EEG obtained during eyes-open paradigms was less stable, and any change observed over time during these testing conditions can be interpreted utilizing the CR values provided. Combined with existing validity data, these findings encourage application of the portable EEG system for the study of brain function.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The electroencephalogram (EEG) has long been used in neuroscience research and clinical practice to detect, describe, and monitor brain function in both healthy individuals and clinical populations (Andreassi, 2007). The EEG reflects the electrical currents that flow in the extracellular space, which are a product of numerous excitatory and inhibitory synaptic potentials occurring in individual neurons (Cacioppo et al., 2007). Abnormal changes in the EEG can be expressed as a disappearance of activity, appearance of abnormal patterns, or deterioration of normal background patterns (Cacioppo et al., 2007; Rijsdijk et al., 2008). Electrodes on the scalp record this electrical activity, which can then be amplified and displayed as a continuous waveform, and digitized for subsequent analysis. Abnormalities recorded by a specific electrode can generally be attributed to the underlying brain region (Andreassi, 2007).

Quantitative EEG (qEEG) refers to the mathematical processing of EEG data in order to more precisely quantify specific EEG parameters than is possible through visual inspection (Tong and Thakor, 2009). qEEG processing techniques have identified robust phase, amplitude, and power band features that reflect unique aspects of brain function. Four different frequency bands in the EEG waveform are of particular interest in characterizing brain function: delta, theta, alpha and beta (Andreassi, 2007).

The recording systems for acquiring qEEG are traditionally lab-based. The resultant lack of mobility limits the range of possible recording situations and contexts (Naunheim et al., 2010). Participants often also report discomfort during the lengthy fitting and calibration procedures (Johnstone et al., 2012) and paradigms requiring serial EEG recordings, such as monitoring change over time, can be hampered by lab-based recording systems (Johnstone et al., 2012, Gevins et al., 2012).

A number of wireless headset EEG devices have become commercially available in recent years that seek to improve usability and portability, while maintaining data quality (Gargiulo et al., 2010; Chi et al., 2010). These portable devices may provide an alternative to conventional recording systems that is better suited to evaluation over time,

* Corresponding author at: Brain and Mind Centre, University of Sydney, 94 Mallet Street, Camperdown, NSW 2050, Australia.
E-mail addresses: jeffrey.rogers@sydney.edu.au (J.M. Rogers), stuart_johnstone@uow.edu.au (S.J. Johnstone), anna.aminov@acu.edu.au (A. Aminov), jim.donnelly@scu.edu.au (J. Donnelly), peterH.wilson@acu.edu.au (P.H. Wilson).

http://dx.doi.org/10.1016/j.ijpsycho.2016.06.006
0167-8760/© 2016 Elsevier B.V. All rights reserved.

particularly when working with populations with limited tolerance for psychophysiological assessment, such as children, or individuals with neurologic, neuropsychological, or psychiatric conditions (Badcock et al., 2013; Johnstone et al., 2012).

One such system of promise is the NeuroSky ThinkGear (San Jose, California). ThinkGear consists of proprietary firmware within a single channel, dry electrode device, which claims ease of use and portability in a research-grade apparatus (Ekandem et al., 2012). An initial validity study reported EEG data derived from a ThinkGear system contained within a headset device termed MindSet (NeuroSky, San Jose, California) is comparable to EEG recorded from a conventional multi-electrode, lab-based system, and sensitive to standard variations in resting and active mental processing states (Johnstone et al., 2012). However, there exists no published data on the ThinkGear system's reliability. Until this time, the utility of the ThinkGear system as a research tool or clinical instrument cannot be fully established.

1.1. EEG test-retest reliability

The relevance and clinical utility of an instrument is largely derived from the ability to provide meaningful and stable results over time. Any observed variance in results over serial testing sessions is considered the result of two components: true change and measurement error (Näpflin et al., 2007). In this context, reliability therefore refers to the proportion of variance attributable to true change, while a *test-retest reliability* statistic reflects the degree to which a measure is consistent and free from error over time (Portney and Watkins, 2009; Sheorajpanday et al., 2009).

Visual examination of EEG traces generally produces low test-retest and inter-rater reliability (Seshia et al., 2008). In contrast, mathematically processed qEEG data derived from conventional, lab-based systems is consistently found to be highly reliable in describing brain electrical activity (Corsi-Cabrera et al., 2007; Näpflin et al., 2007; Salinsky et al., 1991). Test-retest reliabilities of resting state qEEG in healthy individuals typically exceed 0.80 over intervals ranging from one-hour (Burgess and Gruzelier, 1993; Mcevoy et al., 2000), to one-month (Cannon et al., 2012), to more than one-year (Hatz et al., 2014).

Test-retest reliability of qEEG outputs is affected by task parameters, with the highest levels of reliability typically associated with recordings during more cognitively engaging tasks (Gevins et al., 2011; Mcevoy et al., 2000). Lower test-retest reliabilities for resting state qEEG data have been attributed to difficulty controlling for transient variations in vigilance, anxiety, and thought processes during this unstructured condition (Fernández et al., 1993). As such, mental states such as mood and alertness should also be taken into account, as intra-individual differences over time in either of these state factors can significantly influence the EEG (Fernández et al., 1993; Andreassi, 2007; Gevins et al., 2011; Vuga et al., 2006).

Furthermore, research in healthy individuals has demonstrated that the synchronized neural activity of the EEG varies according to developmental stage (Koenig et al., 2002; Tong and Thakor, 2009). Delta and theta activity at rest is prominent in childhood, but these slow waves diminish with age, as higher frequency alpha and beta bands increase almost linearly across adulthood, plateauing in older adulthood (Cacioppo et al., 2007). To account for these lifespan variations, qEEG reliability metrics should be resolved in youth and adult populations separately.

1.2. Study aims and hypotheses

The aim of the current study was to address the gap in our understanding of the value of a single channel portable EEG device as a measurement tool, through an investigation of its test-retest reliability. ThinkGear system derived qEEG data was collected using the MindSet device from healthy youth, adults, and older adults serially over four time points to establish one-day, one-week, and one-month test-retest reliability coefficients and indices of reliable change during resting and active tasks. We predicted that qEEG data derived from the portable device would be stable across the three retest intervals, and comparable to reliability coefficients derived from lab-based systems. In particular qEEG data recorded during cognitively engaging conditions was expected to be more reliable than qEEG obtained during resting conditions.

2. Material and methods

2.1. Participants

This research was approved by the Human Research Ethics Committee of the Australian Catholic University, and each participant (or their guardian) provided written informed consent for voluntary participation. Participants included 19 youth (8 girls) aged 10 to 17 years ($M = 15.30$, $SD = 4.70$), 21 adults (11 female) aged 18 to 28 years ($M = 21.20$, $SD = 3.18$), and 19 older adults (10 female) aged 55 to 79 ($M = 64.84$, $SD = 8.68$). Adult participants were recruited from a university population, and youth and older adults by word of mouth and advertisement in the local community. All groups were drawn from healthy populations, with no reported history of head injury, psychiatric disorder, neurological disorder, cardiovascular disease, or substance abuse. All participants were right handed, with normal hearing and normal or corrected to normal vision.

2.2. Tasks

2.2.1. Control tasks

To assess mood state, adult and older adult participants completed the Depression Anxiety Stress Scale (DASS-21; Lovibond and Lovibond, 1995); youth participants completed the Child Anxiety Life Interference Scale (CALIS; Messer et al., 1995) and the Moods and Feelings Questionnaire for Children (MFQ; Lyneham et al., 2013). On all scales, higher scores reflected poorer mood/anxiety state. The Detection and Identification tasks from the computerized CogState Battery (Maruff et al., 2009) were used to measure simple and choice reaction time, respectively, with faster correct reaction times (once accuracy thresholds are met) indicating superior visuo-motor and information processing. Visual analogue scales (VAS; 0 to 100) were used to measure subjective appraisals of alertness, concentration, and attention; higher scores reflected more optimal states.

2.2.2. EEG tasks

During EEG acquisition, two resting and two active tasks were administered with the order counter-balanced within and across participants at each study session. In the resting tasks participants were asked to sit quietly for 3 min with either their eyes open or eyes closed. Active tasks included an auditory oddball task and a visual n-back task, each of 3 min duration. The auditory oddball task consisted of 500 Hz (non-target) and 1000 Hz (target) tones at 1 s intervals, with non-target tones occurring with a probability of 0.8. Tones were presented binaurally through the inbuilt headphones of the MindSet headset device (see below) at an intensity of 65 dB. For the duration of the task participants were instructed to maintain gaze on a central fixation cross on a computer screen and to react to target tones by pressing the spacebar on a computer keyboard. The visual numeric n-back task (Jaeggi et al., 2010) required participants to make a same-different comparison of numeric stimuli that occurred immediately prior to the current probe stimulus (i.e. a 1-back condition). In each trial, a number was presented for 500 ms, followed by a 2000 ms period before presenting the next number (i.e. 2500 ms inter-stimulus interval). Participants were asked to press the spacebar if the current number was identical to the number presented immediately prior (a target). Targets occurred at random intervals with a probability of 0.3.

2.3. Procedure

Participants completed four 30-min sessions. Session 2 occurred one-day after Session 1; Session 3 took place one-week after Session 2; Session 4 occurred one-month after Session 3. To control for potential moderating effects, each study session began with administration of the Control Tasks, which took approximately 15 min to complete. Setting up the MindSet device and administration of the four EEG tasks followed for another 15 min. Participants were instructed to minimize facial, head, and body movement during the EEG recording session. University students received course credit for participation in the study. Youth and older adult participants were reimbursed with a $10 gift voucher.

2.4. EEG data acquisition and analysis

EEG was obtained with the MindSet (NeuroSky, San Jose, California) portable headset device. The MindSet device consists of a ThinkGear microchip and embedded firmware, and 10 mm dry stainless steel active, material reference, and ground electrodes contained within a set of headphones. The reference and ground electrodes are housed within the left ear pad, while the EEG electrode is embedded in a flexible arm extending from the left headband, positioned at the international 10–20 system site FP1. Electrical potentials at the active and reference electrodes are subtracted through common mode rejection to derive a single EEG channel signal which is amplified 8000 times. Sampling and amplification of the raw 128 Hz data are carried out within the embedded microchip and transmitted wirelessly by Bluetooth© to a computer for recording and subsequent off-line quantitative analysis.

Scan Edit version 3 software (Neuroscan, Herndon, Virginia) was used to analyze the EEG data. The raw EEG waveform was band-pass filtered at 0.5–30 Hz, and manually inspected to identify any movement or muscle artefact. Identified sections of the trace were marked so as not to be included in further processing. Artefact-free 4.0 s EEG epochs (1/4 Hz resolution) were submitted to Fast Fourier Transforms (10% Hamming window), with the resultant power spectrum data summed to derive power in the following EEG frequency bands: delta (1.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), and beta (12.5–25 Hz). Relative power (RP) was then calculated by summing absolute power across the delta, theta, alpha and beta bands to compute total power, and then dividing the absolute power for each individual band by the total power, expressed as a percentage.

2.5. Statistical analysis

One-way repeated measures ANOVAs were conducted on the Control Tasks to detect any significant changes that could potentially influence the EEG reliability between sessions. Where assumptions of sphericity were violated, Huynh-Feldt corrections were applied. To evaluate the test-retest reliability of ThinkGear derived EEG, the RP values of the four EEG frequency bands were pair-wise compared across sessions (e.g. Session 1 vs Session 2, Session 2 vs Session 3, and Session 3 vs Session 4) to compute intra-class correlations (ICCs; Barton and Peat, 2014). This is a more appropriate measure of association than techniques such as Pearson's Product Moment Correlation as it assesses the degree of variation from each assessment (Wilk et al., 2002). The following descriptions proposed by Shrout (1998) were used to categorize the magnitude of ICC values: *virtually none* [0.00–0.10], *slight* [0.11–0.40], *fair* [0.41–0.60], *moderate* [0.61–0.80], and *substantial* [0.81–1.0].

Coefficients of Repeatability (CR; Vaz et al., 2013), also referred to as the Smallest Real Difference (SRD), were calculated from the RP values for each frequency band, by multiplying the Standard Error of Measurement (SEM) by 2.77, where $SEM = SD * \sqrt{(1 - ICC)}$ (Bland and Altman, 1999; Collado-Mateo et al., 2015). The CR provides a value computed in the same units as the measurement tool, below which 95% of retest differences (i.e. error) lie. Any retest difference measurement beyond the CR value can therefore be considered genuine and significant change (Bland and Altman, 1999).

3. Results

Means and *SDs* of the Control Tasks are presented in Table 1. For all three participant groups the CogState information processing tasks and subjective VAS self-reports were stable and within normal limits over the four sessions (Mollica et al., 2005; Maruff et al., 2009). Older adult and youth measures of mood state were also stable and within normal limits at all sessions (Messer et al., 1995; Lyneham et al., 2013). The Stress and Anxiety sub-scales of the DASS varied over time in adult participants. However, these differences were not clinically significant, with scores at all times remaining well within the "normal" range (Lovibond and Lovibond, 1995).

From the four 3-min EEG tasks, the number of valid epochs obtained at each session for each participant group is presented in Table 2. A mixed-design ANOVA, with Greenhouse-Geisser corrections, was performed on the number of valid epochs, with condition (eyes-closed, eyes-open, n-back, oddball) and session (1, 2, 3, 4), as within subjects factors, and age group (youth, adult, older adult) a between subjects factor. Where significant interactions were detected, planned contrasts exploring the source of the effect were performed, with the alpha level required for significance set at $p < 0.01$ due to the number of statistical tests conducted on the data. The main effect of condition was significant, $F(2.06, 115.26) = 60.04$, $p < 0.01$, $\eta_p^2 = 0.52$, with linear contrasts revealing a greater number of valid epochs were obtained in the eyes-closed condition in all three participant age groups. A session by group interaction effect was also significant, $F(5.25, 146.93) = 3.09$, $p < 0.01$, $\eta_p^2 = 0.10$. Repeated contrasts revealed the youth group demonstrated a reduction in the number of valid epochs at Session 4, $F(2, 56) = 6.53$, $p < 0.01$, $r = 0.32$.

Average RP values of the four conditions across the four sessions are presented in Figs. 1–3. A mixed design ANOVA found a significant main effect of frequency (delta, theta, alpha, beta) on the RP values [$F(1.70, 95.40) = 328.33$, $p < 0.01$, $\eta_p^2 = 0.85$], reflecting a general pattern of more prominent delta and theta activity relative to alpha and beta brainwaves. The frequency by group interaction was also significant [$F(6, 168) = 5.66$, $p < 0.01$, $\eta_p^2 = 0.17$], in keeping with age-related EEG differences reported in the literature.

Furthermore, consistent with earlier results (Johnstone et al., 2012), paired samples *t*-test analysis of current grand averaged RP data was sensitive to variations in task condition. Specifically, from the eyes-open to eyes-closed condition relative alpha power increased [$t(58) = 3.94$, $p < 0.01$, $d = 0.44$] while relative theta power decreased [$t(58) = -6.01$, $p < 0.01$, $d = 0.59$]. Compared to the eyes-open resting condition, relative alpha [$t(58) = 5.44$, $p < 0.01$, $d = 0.40$] and beta [$t(58) = 4.68$, $p < 0.01$, $d = 0.39$] power increased for conditions with a cognitive load (i.e. n-back and oddball tasks) while relative delta power reduced [$t(58) = -4.92$, $p < 0.01$, $d = 0.46$].

ICCs were calculated from the RP values for each frequency band during the resting and active tasks over the three retest intervals (see Figs. 4–6). The three age groups consistently demonstrated *moderate* ICC values at each retest interval, with average reliability at the one-day interval 0.71 ($SD = 0.13$) for older adults, 0.76 ($SD = 0.14$) for adults, and 0.73 ($SD = 0.15$) for youth. At the one-week interval average reliability was 0.70 ($SD = 0.10$) for older adults, 0.71 ($SD = 0.15$) for adults, and 0.74 ($SD = 0.18$) for youth. The one-month average reliability was 0.79 ($SD = 0.13$) for older adults, 0.70 ($SD = 0.20$) for adults, and 0.74 ($SD = 0.19$) for youth.

The ICC values for each EEG power band were also of at least *moderate* reliability. The delta power band provided *moderate* reliability, with an average ICC value of 0.71 ($SD = 0.11$) for older adults, 0.63 ($SD = 0.20$) for adults, and 0.71 ($SD = 0.19$) for youth. The theta power band also provided *moderate* reliability, with an average ICC value of 0.69 ($SD = 0.15$) for older adults, 0.75 ($SD = 0.08$) for adults, and

Table 1
Mean (Standard Deviations) results of participants' subjective ratings scales and cognitive tasks.

Older adults	Session 1	Session 2	Session 3	Session 4	F[f]	p
Detection[a]	461.3 (178.4)	467.2 (183.6)	455.1 (168.5)	495.0 (226.9)	1.33	0.3
Identification[a]	612.4 (124.7)	615.3 (144.2)	618.9 (136.4)	617.4 (98.1)	1.52	0.3
Alertness[b]	77.6 (20.9)	78.1 (22.5)	82.2 (16.1)	80.8 (15.7)	3.11	0.2
Motivation[b]	82.5 (16.5)	82.0 (19.3)	83.3 (18.6)	84.9 (15.6)	1.55	0.2
Concentration[b]	76.0 (18.6)	81.3 (19.7)	82.9 (17.2)	83.0 (14.0)	2.24	0.3
Depression[c]	1.9 (2.3)	1.2 (1.7)	1.3 (1.9)	1.7 (2.3)	3.23	0.2
Anxiety[c]	1.2 (1.8)	1 (1.7)	0.8 (1.7)	1.2 (2.0)	2.22	0.3
Stress[c]	3.1 (2.9)	2.6 (2.7)	2.2 (2.5)	2.3 (2.6)	2.34	0.3
Adults						
Detection[a]	315.5 (133.9)	291.2 (64.1)	280.7 (49.1)	273.0 (49.9)	1.72	0.3
Identification[a]	422.8 (73.2)	429.6 (41.7)	436.9 (52.5)	436.7 (35.9)	1.37	0.3
Alertness[b]	52.8 (22.6)	53.7 (24.6)	67.0 (15.3)	55.1 (22.4)	2.14	0.1
Motivation[b]	54.6 (12.5)	61.1 (17.1)	60.7 (18.4)	54.2 (23.2)	1.06	0.4
Concentration[b]	55.5 (18.4)	61.5 (15.1)	60.6 (21.5)	53.7 (19.5)	1.27	0.3
Depression[c]	3.9 (3.6)	4.6 (5.4)	2.9 (3.3)	1.9 (3.3)	3.15	0.1
Anxiety[c]	2.9 (3.3)	2.6 (2.2)	1.5 (1.8)	0.8 (1.24)	3.24	0.03
Stress[c]	6.8 (4.6)	6.7 (4.62)	3.7 (2.9)	2.9 (2.8)	4.84	0.02
Youth						
Detection[a]	299.6 (51.9)	291.2 (56.4)	280.7 (51.8)	27.3 (48.2)	0.93	0.4
Identification[a]	306.8 (56.8)	309.6 (46.3)	306.9 (42.5)	35.7 (39.1)	1.32	0.3
Alertness[b]	52.8 (22.6)	56.7 (24.7)	53.9 (20.3)	57.2 (20.6)	3.21	0.2
Motivation[b]	60.7 (17.5)	70.6 (18.3)	67.7 (19.7)	70.2 (20.2)	1.95	0.3
Concentration[b]	60.5 (18.4)	62.8 (17.1)	66.6 (18.6)	69.7 (19.5)	1.64	0.3
Depression[d]	2.9 (3.5)	2.3 (5.3)	2.3 (3.3)	1.7 (3.2)	2.43	0.2
Anxiety[e]	3.0 (3.2)	2.6 (2.2)	1.9 (2.3)	2.1 (2.5)	2.13	0.3

Notes:
[a] CogState tassk in ms.
[b] VAS range 0–100.
[c] DASS subscale range 0–21.
[d] MFQ range 0–26.
[e] CALIS range 0–36.
[f] Repeated measures ANOVAs.

0.68 ($SD = 0.21$) for youth. Similarly, the beta power band provided *moderate* reliability, with an average ICC value of 0.70 ($SD = 0.12$) for older adults, 0.77 ($SD = 0.15$) for adults, and 0.75 ($SD = 0.16$) for youth. The alpha power band provided *moderate* to *substantial* reliability, with an average ICC value of 0.83 ($SD = 0.09$) for older adults, 0.73 ($SD = 0.16$) for adults, and 0.82 ($SD = 0.08$) for youth.

The eyes-closed condition provided *moderate* to *substantial* reliability, with an average ICC of 0.76 ($SD = 0.12$) for older adults, 0.79 ($SD = 0.07$) for adults, and 0.85 ($SD = 0.08$) for youth. The *n*-back condition also provided *moderate* to *substantial* reliability, with averages of 0.75 ($SD = 0.15$) for older adults, 0.85 ($SD = 0.10$) in adults, and 0.77 ($SD = 0.09$) in youth. Reliability in the oddball condition was *moderate*, with averages of 0.71 ($SD = 0.12$) for older adults, 0.74 ($SD = 0.16$) for adults, and 0.71 ($SD = 0.22$) for youth.

Table 2
Average number of valid epochs for older adult, adult and youth groups across sessions and conditions.

Older adults	Eyes-closed	Eyes-open	N-back	Oddball
Session 1	35	28	28	25
Session 2	34	26	27	30
Session 3	34	24	25	21
Session 4	33	30	26	24
Average	34	27	27	25
Adults				
Session 1	36	20	15	18
Session 2	33	11	14	19
Session 3	32	12	14	21
Session 4	36	19	13	22
Average	34	16	14	20
Youth				
Session 1	31	20	14	17
Session 2	35	21	16	20
Session 3	32	18	16	24
Session 4	19	13	6	11
Average	29	18	13	18

The eyes-open condition had the lowest reliability, with *fair* to *moderate* reliability averages of 0.71 ($SD = 0.12$) for older adults, 0.57 ($SD = 0.25$) in adults and 0.63 ($SD = 0.19$) in youth.

As the ICC is a relative index, influenced by heterogeneity in measurement across different subjects, CR values were also analyzed to provide an absolute reliability index (Vaz et al., 2013). CR values were computed from the RP data in the four conditions across the three retest intervals for older adult, adult, and youth participants (see Table 3). The average CR value at the one-day interval was ± 0.08 in the adult group, ± 0.06 in the older adults, and ± 0.05 in the youth group. The average one-week CR was ± 0.07 in the adult group, ± 0.06 in the older adults, and ± 0.05 in the youth group. At the one-month interval the average CR was ± 0.07 for adults, $\pm$ 0.05 for youth, and ± 0.04 for older adults.

For the delta power band the average CR value was ± 0.15 in the adult group, $\pm$ 0.08 in the older adult group, and ± 0.07 in the youth group. The theta band average CR was ± 0.05 in older adults, and ± 0.04 in adults and youth. Alpha power band CRs were an average of ± 0.07 for the adults, and ± 0.04 for the youth and older adult groups. Finally, the average CR for the beta power band was ± 0.04 for all groups.

The *n*-back condition CR average was ± 0.05 for all groups. The eyes-closed condition average CR value was ± 0.09 in the adult group, $\pm$ 0.05 in the older adult group, and ± 0.03 in the youth. The average CR in the oddball condition was ± 0.06 in adults and older adults, and ± 0.05 in the youth group. Finally, the eyes-open condition had an average CR of ± 0.10 in adults, $\pm$ 0.07 in the youth group, and ± 0.06 in older adults.

4. Discussion

An efficient measure of brain function would be helpful to researchers, clinicians, and the populations they serve. The EEG reflects real-time cerebral electrical activity, providing a unique, non-invasive

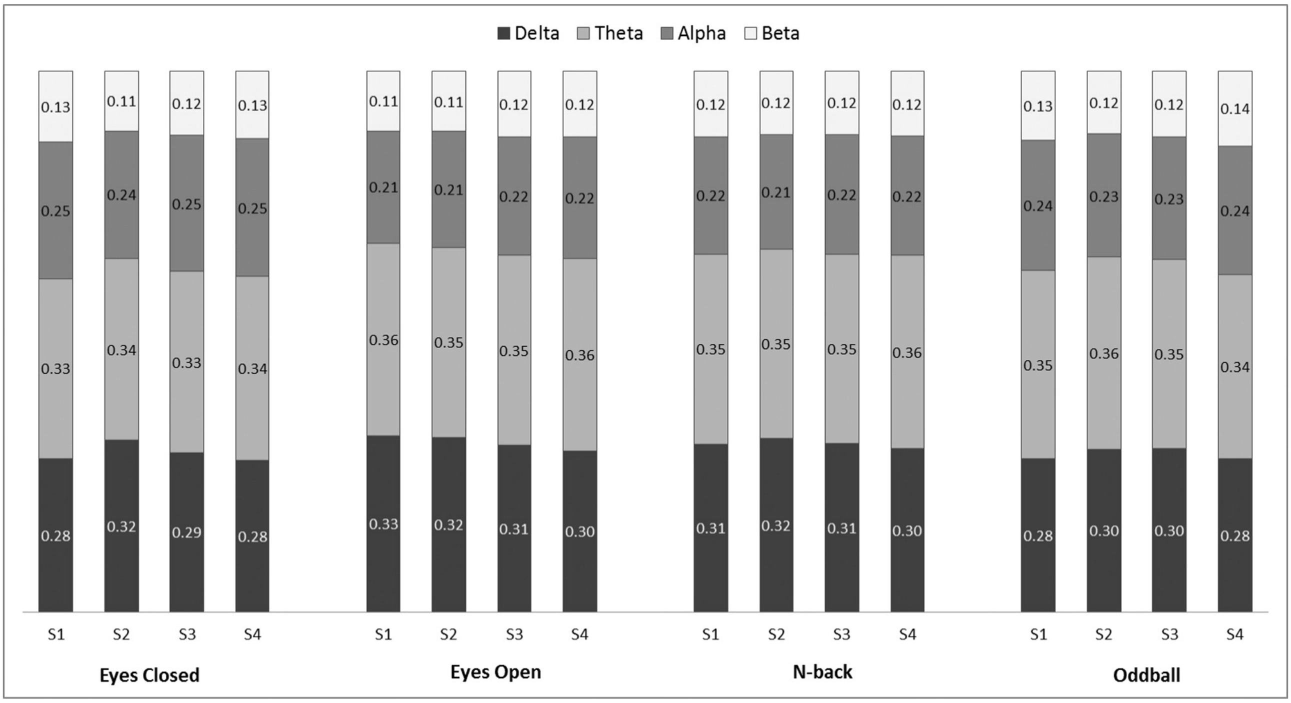

Fig. 1. Grand averaged Relative Power (in percentages) for the older adult group across sessions and conditions.

index of brain function with reasonable spatial and high temporal resolution (Kennett, 2012). qEEG data is frequently used in repeated assessment designs for evaluation of the functional status of the brain over time (Tong and Thakor, 2009). Utility in the context of serial recordings requires sufficiently high stability of qEEG parameters, but is hampered by conventional systems which are typically lab-based. Previous research (Johnstone et al., 2012) has evaluated the validity of NeuroSky ThinkGear derived EEG data, and this study extends that work by demonstrating convergent results in independent youth, adult, and older adult populations. The ThinkGear system was again sensitive to standard variations in task conditions, and produced resting state RP values compatible with previous EEG reports using conventional lab-based systems in youth (Clarke et al., 2001), adults (Barry et al., 2007), and older adults (Barry et al., 2014). As a promising alternative to lab-based systems, the current study also now establishes the test-retest reliability of single-channel EEG data collected using this portable device over one day, week and month retest intervals from healthy populations across the lifespan.

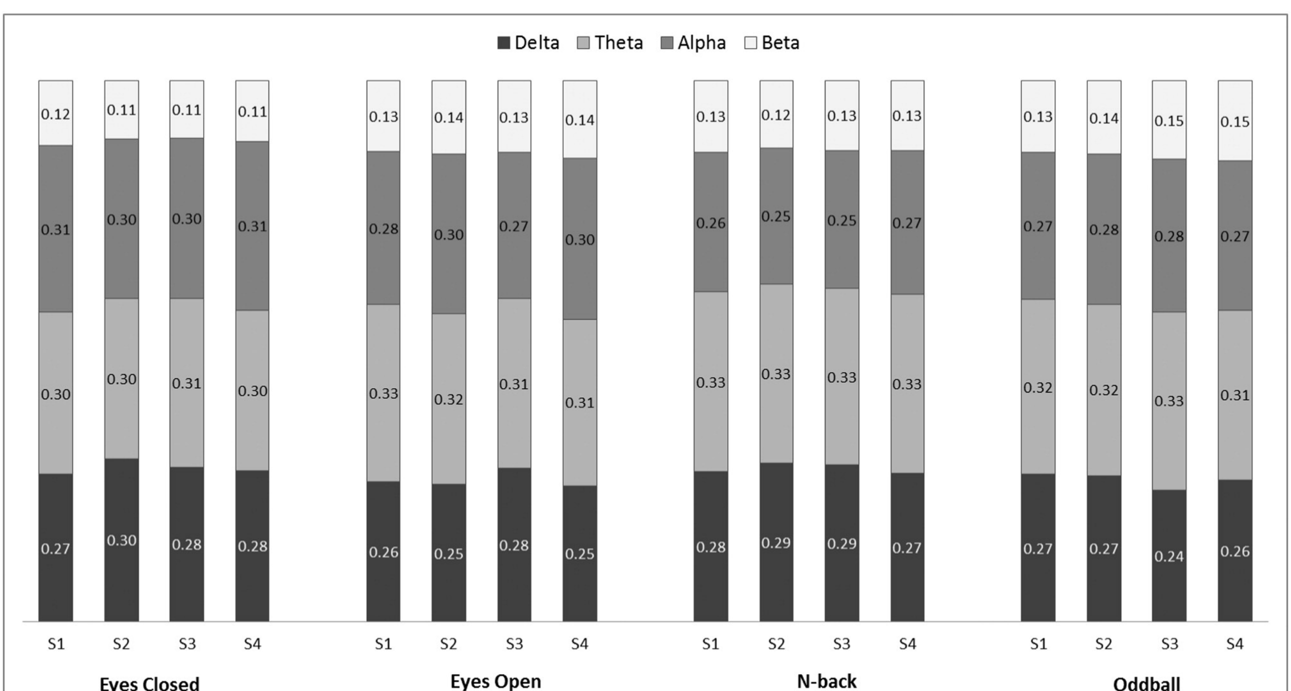

Fig. 2. Grand averaged Relative Power (in percentages) for the adult group across sessions and conditions.

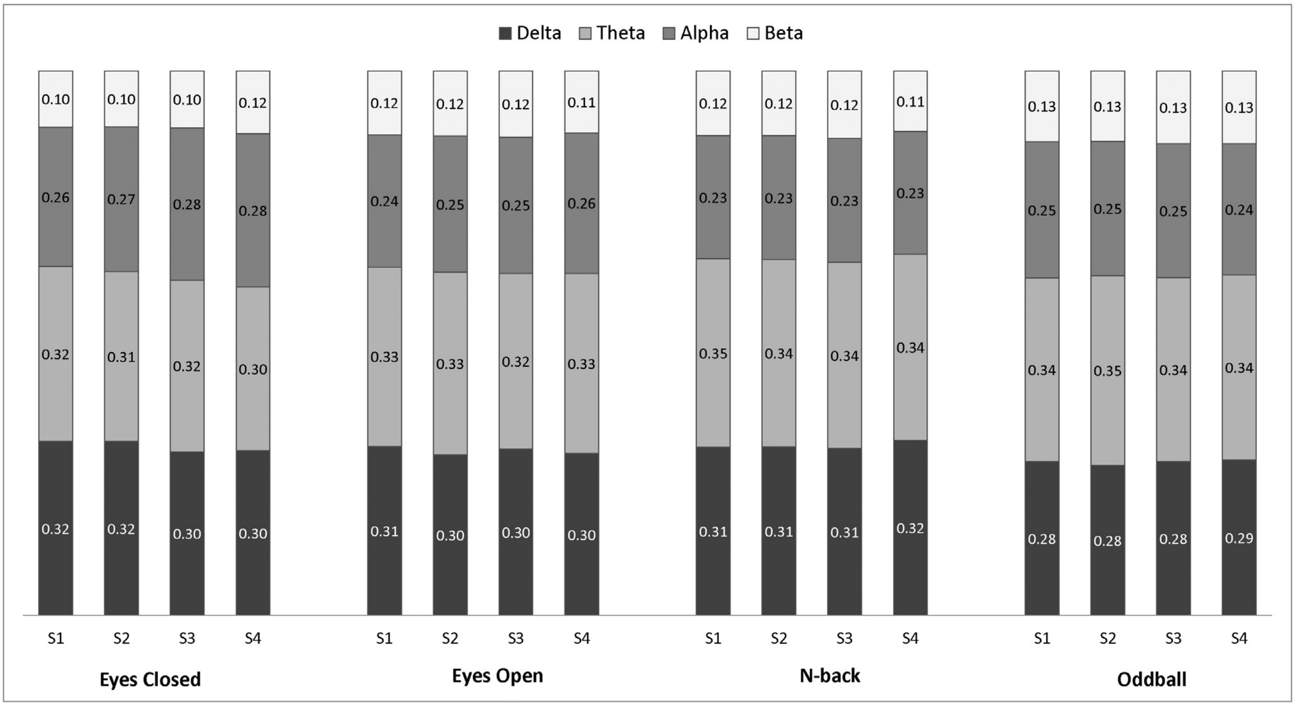

Fig. 3. Grand averaged Relative Power (in percentages) for the youth group across sessions and conditions.

4.1. NeuroSky ThinkGear retest reliability

ICC values obtained from the ThinkGear system deployed in a MindSet headset device were consistently *moderate* [0.61–0.80] to *substantial* [0.81–1.0] across age groups, retest intervals, and power bands. In adult participants the visual *n*-back condition had the highest overall stability over all three retest intervals. This "*substantial*" level of reliability was consistent with expectations that the more structured nature of active EEG conditions would provide the most stable results (Mcevoy et al., 2000). ICC values were also in keeping with qEEG reliability results from n-back tasks obtained using conventional lab-based recording systems (Gevins et al., 2011, Mcevoy et al., 2000). However, these findings should be interpreted with a degree of caution, as the n-back condition in adults provided less valid epochs compared

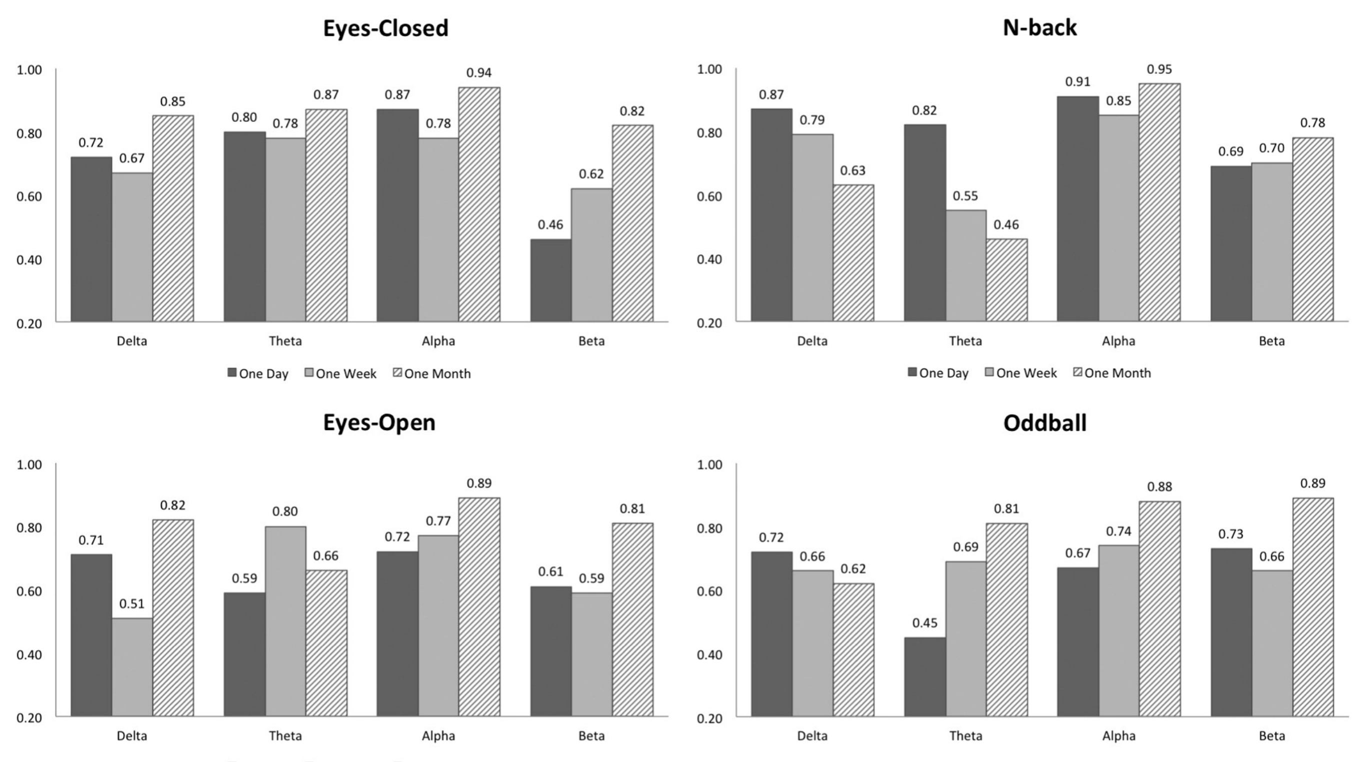

Fig. 4. Intra-Class Correlations for the older adult group across sessions, conditions, and frequency bands.

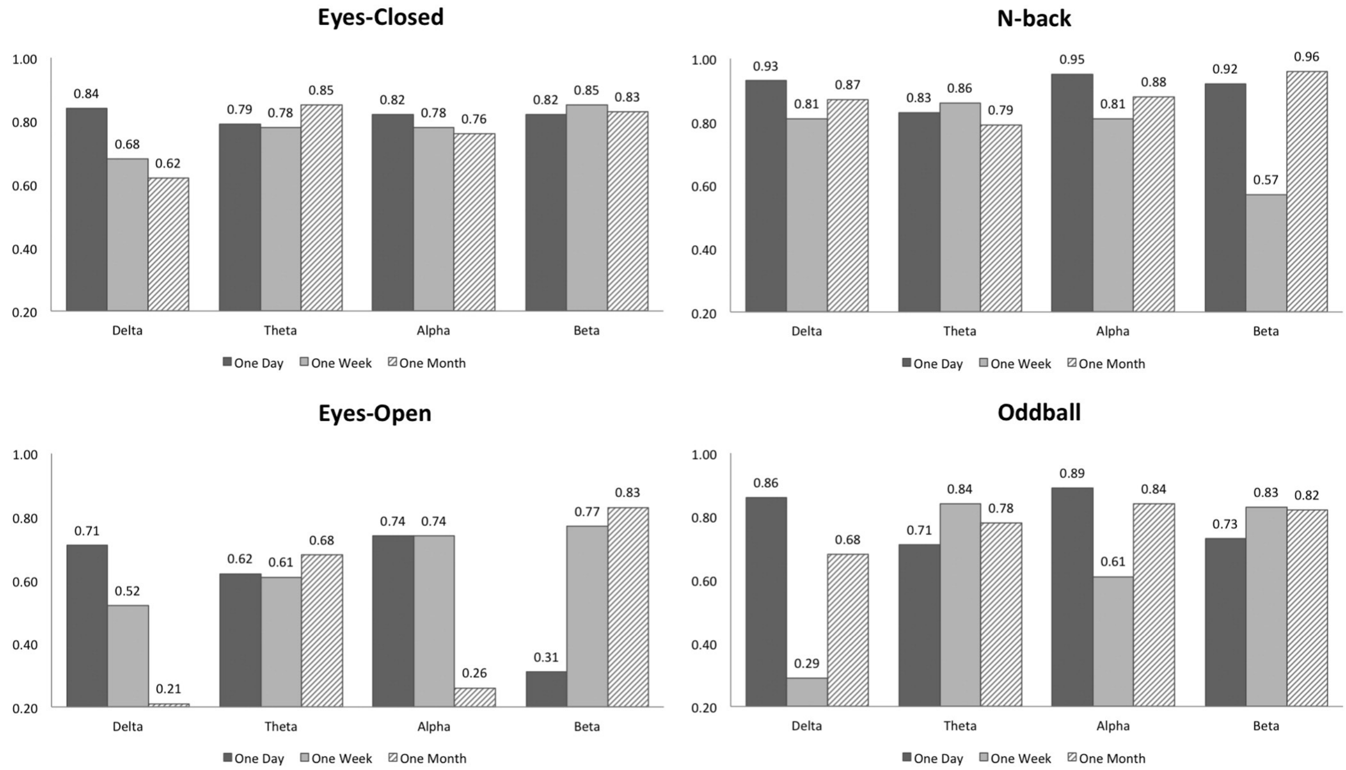

Fig. 5. Intra-Class Correlations for the adult group across sessions, conditions, and frequency bands.

with other conditions. Importantly though, the average number of valid epochs remained well above the quantity of EEG data that has been identified as necessary to achieve *high* internal consistency ($\alpha = 0.8$) and ensure small measurement error (Lund et al., 1995).

In adult participants, the eyes-closed condition provided the highest number of valid epochs, while also demonstrating *moderate* levels of reliability over the three retest intervals. Contrary to expectations, the active oddball condition was not more reliable than the resting eyes-

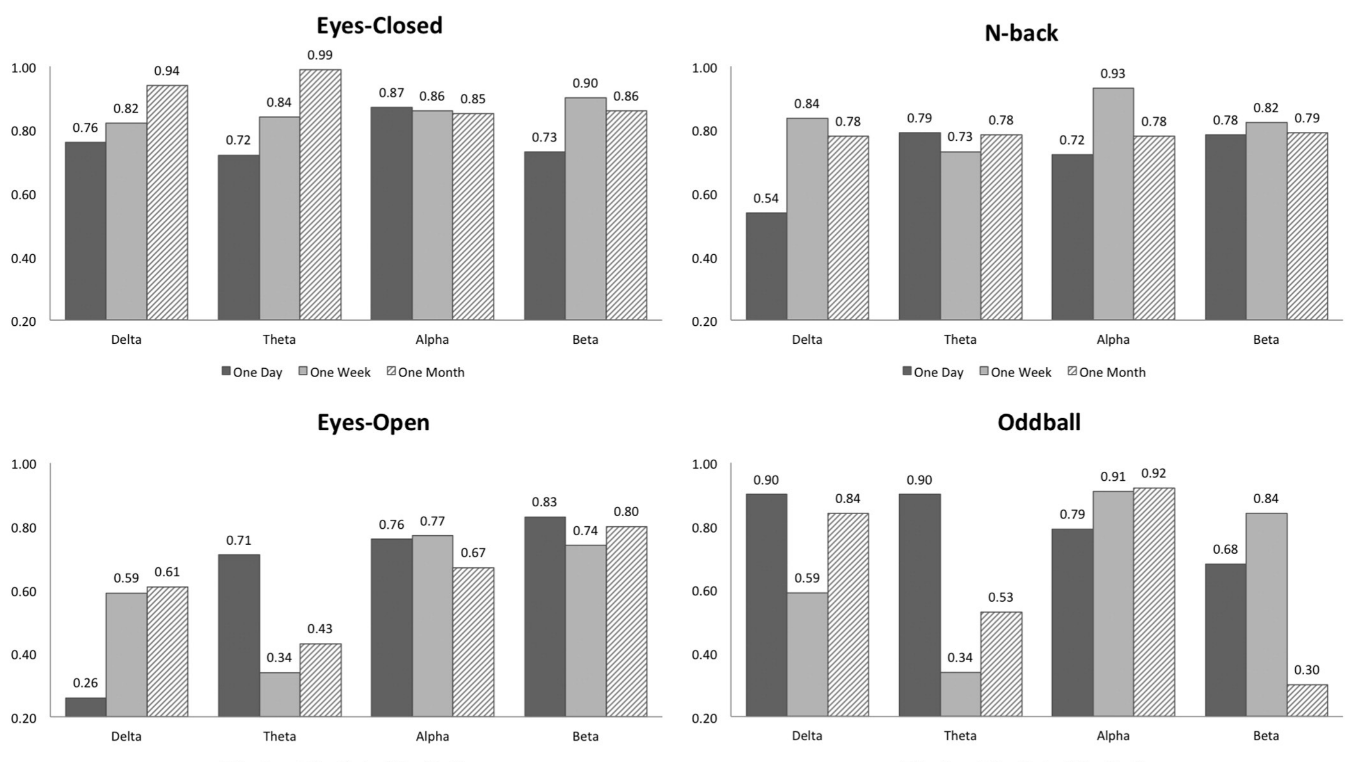

Fig. 6. Intra-Class Correlations for the youth group across sessions, conditions, and frequency bands.

Table 3
Coefficients of Repeatability (CR) for the older adult, adult, and youth groups across sessions, conditions, and frequency bands.

Older adults		Delta	Theta	Alpha	Beta
Eyes-closed	1 day	±0.07	±0.04	±0.04	±0.07
	1 week	±0.09	±0.04	±0.07	±0.04
	1 month	±0.04	±0.03	±0.02	±0.03
Eyes-open	1 day	±0.08	±0.08	±0.07	±0.05
	1 week	±0.13	±0.04	±0.05	±0.07
	1 month	±0.05	±0.06	±0.03	±0.03
N-back	1 day	±0.04	±0.03	±0.02	±0.05
	1 week	±0.06	±0.05	±0.03	±0.04
	1 month	±0.09	±0.09	±0.01	±0.03
Oddball	1 day	±0.07	±0.08	±0.07	±0.03
	1 week	±0.08	±0.06	±0.06	±0.05
	1 month	±0.09	±0.04	±0.02	±0.02
Adults					
Eyes-closed	1 day	±0.50	±0.03	±0.04	±0.03
	1 week	±0.10	±0.03	±0.06	±0.02
	1 month	±0.11	±0.02	±0.07	±0.02
Eyes-open	1 day	±0.12	±0.05	±0.08	±0.11
	1 week	±0.17	±0.06	±0.07	±0.05
	1 month	±0.22	±0.05	±0.15	±0.04
N-back	1 day	±0.06	±0.04	±0.04	±0.05
	1 week	±0.09	±0.04	±0.05	±0.02
	1 month	±0.08	±0.03	±0.04	±0.02
Oddball	1 day	±0.04	±0.04	±0.03	±0.05
	1 week	±0.20	±0.03	±0.10	±0.04
	1 month	±0.10	±0.03	±0.05	±0.05
Youth					
Eyes-closed	1 day	±0.07	±0.04	±0.03	±0.05
	1 week	±0.05	±0.02	±0.04	±0.01
	1 month	±0.03	±0.00	±0.04	±0.02
Eyes-open	1 day	±0.14	±0.03	±0.05	±0.03
	1 week	±0.10	±0.10	±0.05	±0.04
	1 month	±0.09	±0.09	±0.06	±0.03
N-back	1 day	±0.11	±0.03	±0.06	±0.03
	1 week	±0.05	±0.04	±0.02	±0.03
	1 month	±0.08	±0.02	±0.06	±0.03
Oddball	1 day	±0.04	±0.02	±0.05	±0.04
	1 week	±0.10	±0.07	±0.03	±0.03
	1 month	±0.04	±0.08	±0.03	±0.10

closed condition, again providing a *moderate* level of reliability. Finally, in adult participants, the eyes-open condition was the least reliable. However, the *fair* level of reliability in this condition may still be sufficient for research conditions (Vaz et al., 2013). Furthermore, the superior test-retest reliability for eyes-closed compared to eyes-open resting-state conditions was consistent with previous findings (Corsi-Cabrera et al., 2007; Fernández et al., 1993; Cannon et al., 2012), suggestive of a higher reactivity to environmental influences with eyes-open than with eyes-closed (Barry et al., 2007; Corsi-Cabrera et al., 2007).

Youth and older adult participants showed a similar pattern of test-retest reliability. The eyes-closed and n-back conditions were the most reliable, ranging from *substantial* to *moderate*, respectively. Similar to the adult group, the eyes-closed condition again provided the highest number of valid epochs, while the n-back condition again had more invalid epochs. Like adults, slightly lower reliability was observed in the oddball and eyes-open conditions with youth and older adult participants. While of sufficient quantity to permit analysis, Youth participants provided less valid epochs at Session 4 compared with other participants. No abnormalities were detected in the corresponding control measures. Although set-up of the ThinkGear system can be achieved in as little as 30 s (Ekandem et al., 2012), strategies to address successful completion (Paasch et al., 2012) can still be encouraged, to facilitate the tolerance of youth populations for repeated neurophysiological assessment (Johnstone et al., 2012).

Contrary to predictions, cognitively engaging tasks were not consistently more reliable than resting conditions (Fernández et al., 1993; Gevins et al., 2011, Mcevoy et al., 2000). These findings cannot easily be explained by variations in mood or attention as all control measures were consistently within normal limits across the four recording sessions. However, participants were only asked to make general session ratings, while their mood or attention may have varied from task to task. As such, the current study design was not sensitive to intra-session changes in state characteristics, for instance in participants who may have found the active EEG tasks more stressful.

An alternative hypothesis is that the demands of the n-back and oddball tasks were not sufficient to be considered "cognitively engaging." Comparison with 2- or 3-back versions of the n-back task and varying the discriminability of tones in the auditory oddball task would enable this hypothesis to be tested systematically. Such changes in the task protocols may also lead to greater similarity in test-retest statistics for the oddball and n-back tasks if both required increased vigilance on each trial and similar cognitive load. Notably, as opposed to the n-back task where every stimulus required a same-different judgement, only 20% of stimuli in the oddball condition were targets. As such, there were periods where participants could have recognized that no response was required, which might make the current oddball task more susceptible to variations in thought processes.

Furthermore, the n-back, oddball and eyes-open conditions were all performed with eyes open. The absence of ocular channels in a single electrode portable device leaves it open to eye movement and blink artifacts. This was apparent in the lower number of valid epochs for eyes-open compared with eyes-closed conditions across all three participant groups. Use of adaptive filters and blind source separation techniques to remove eye movement artefact from single-channel qEEG are techniques that warrant future investigation (Shao et al., 2009).

In addition, options for eyes-closed performance of EEG tasks are likely to increase the number of valid epochs available for analysis, and also enhance the test-retest reliability values obtained when recording from the ThinkGear system. For instance, the oddball condition of the current task could easily be modified for eyes-closed performance. Auditory n-back tasks are not widely used, but are available (Forn et al., 2007) and studies using similar tasks such as the Paced Auditory Serial Addition Test (The Psychological Corporation, 1998) also have demonstrated utility in EEG paradigms (Rogers et al., 2015; Rogers and Fox, 2012).

In sum, a portable EEG device was reliable over short (one-day) and longer (one-week and one-month) retest-intervals. These findings are in accord with reliability data obtained for conventional lab-based EEG recording systems (Fernández et al., 1993; Mcevoy et al., 2000; Salinsky et al., 1991; Burgess and Gruzelier, 1993; Corsi-Cabrera et al., 2007; Cannon et al., 2012). Reliability values from the portable EEG device also compare well to those reported for other commonly utilized functional measures, include resting (Zuo and Xing, 2014) and active fMRI (Bennett and Miller, 2010), PET/SPECT (Egerton et al., 2010; Catafau et al., 2008), and MEG (Martín-Buro et al., 2016). The reliability of conventional EEG systems has also been demonstrated over periods in excess of one-year (Näpflin et al., 2007; Hatz et al., 2014; Vuga et al., 2006; Gevins et al., 2012). Investigation of the ThinkGear system over longer time intervals is recommended to establish its utility in longitudinal evaluations and for clinical applications where change with age or recovery must be monitored precisely.

Furthermore, consistent with previous research (Andreassi, 2007; Cacioppo et al., 2007; Handy, 2005; Mostow et al., 2011; Fein et al., 1983) the older adult, adult, and youth participants consistently demonstrated stable results across sessions, conditions, and EEG frequencies. These findings support use of single-channel qEEG data in clinical and research applications across the lifespan. However, differences in EEG patterns have been reported between adults and children aged 10 years or younger (Schlaggar et al., 2002; Schomer and Lopes De Silva, 2010); this lower age range was not included in the current study and the current portable device findings would need to be verified with younger children.

Finally, there exists some inconsistency in estimates of reliability for individual EEG frequency bands. Lower frequency delta activity is less reliable according to some authors (Pollock et al., 1991; Gudmundsson et al., 2007). In contrast, Lewis et al. (2007) found higher frequency alpha was less reliable than other EEG bands. Others have reported no significant differences between frequency bands (Salinsky et al., 1991; Sheorajpanday et al., 2009; Enoch et al., 2008), consistent with the average ICC values in the current study for EEG frequency bands within a task condition. ICC values obtained in the current study were also consistent with previous reliability data derived from conventional lab-based EEG recording systems that explored the four characteristic frequency bands in older adults and adults (Gasser et al., 1985), and youth (Fein et al., 1983).

4.2. Boundaries of a true-change

Widespread use of standardized tools places the onus on clinicians and researchers alike to understand the upper and lower boundaries of the true-change measured by the tool at retest, in order to interpret results and make significance decisions (Vaz et al., 2013; Barton and Peat, 2014). The current study is one of the first to report on reliable change for qEEG data, and the CRs provided can be of value to clinicians and researchers interested in using a portable EEG device to monitor change. CRs contain 95% of differences between repeated measurements on the same subject. These values permit calculation of significant and meaningful change limits over the three retest intervals. For example, to obtain a significant change in the RP of beta over a one-week retest interval during the eyes-closed condition for an adult participant, differences of at least ± 0.02 should be observed at re-assessment to be 95% confident that a significant change has occurred. Because the CR is quantified in the same units as the assessment tool, it lends itself to easy interpretation. A change score below the CR value may simply reflect measurement error.

In the current study, the average magnitude for clinically relevant change did not tend to exceed ± 0.1 for any retest interval, EEG frequency, or task condition. Collado-Mateo et al. (2015) also recently examined the test-retest reliability of a 20 channel dry electrode wireless EEG device in adults. They found the within-session average CR of alpha band data derived from frontal and central electrodes during balance tasks ranged from 0.02 to 0.03. This compares well to results from the current study, which found the average one-day CRs for the alpha power band in adults ranged from 0.03 to 0.04 for the eyes-closed, n-back, and oddball tasks. The narrow absolute differences between repeated measurements obtained in the current study indicate a high degree of reliability of the EEG measurements obtained from the single channel device. This normative data from across the lifespan during both resting and active EEG tasks can be used to reliably detect change in repeated measurements. The low CR values also suggest the stability of qEEG obtained from the ThinkGear system is a potentially sensitive biological marker for tracking longitudinal changes deviant from healthy individuals, but this requires further investigation (Cannon et al., 2012; Gudmundsson et al., 2007).

4.3. Conclusions

EEG is not a new technique, and developments in computing and technology are continually extending its roles and use (Kennett, 2012). Of course there are trade-offs associated with the use of a single-channel device to acquire EEG data. While the ThinkGear system is convenient and wireless, there is a minor trade-off in terms of data quality compared to conventional lab-based systems (Johnstone et al., 2012), and a major trade-off in the number of recording locations. Furthermore, a single channel of data does not permit computational methods such as coherence (Guevara and Corsi-Cabrera, 1996), source localization (Jatoi et al., 2014), or event-related potentials (Woodman, 2010). However, not all clinical populations will tolerate, and not all research approaches require such designs. In these circumstances the availability of a valid and reliable single channel device potentially enables more diverse research. To this end, eyes-closed NeuroSky ThinkGear qEEG data recorded from a left frontal scalp site is reproducible over both short and longer intervals in healthy participants drawn from across the lifespan. Eyes-open EEG conditions provide somewhat lower reliability, indicating the need for stricter controls when these conditions are applied in serial EEG recording paradigms; confidence in change over time can be evaluated using the provided CRs. Combined with existing validity data (Johnstone et al., 2012), the current results suggests a portable device may provide a viable alternative to conventional lab-based recording systems for assessing changes in electrophysiological signals, and further application to the study of brain function using the system can be encouraged.

Funding acknowledgment

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

References

Andreassi, J.L., 2007. Psychophysiology: Human Behavior and Psychological Response. Psychology Press, New York, NY.
Badcock, N.A., Mousikou, P., Mahajan, Y., De Lissa, P., Thie, J., Mcarthur, G., 2013. Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. PeerJ 1, e38.
Barry, R.J., Clarke, A.R., Johnstone, S.J., Magee, C.A., Rushby, J.A., 2007. EEG differences between eyes-closed and eyes-open resting conditions. Clin. Neurophysiol. 118, 2765–2773.
Barry, R.J., Deblasio, F.M., Cave, A.E., 2014. Eyes-closed vs. eyes-open EEG in young and older adults. Int. J. Psychophysiol. 94, 236.
Barton, B., Peat, J., 2014. Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal. John Wiley & Sons, UK.
Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? Ann. N. Y. Acad. Sci. 1191, 133–155.
Bland, J.M., Altman, D.G., 1999. Measuring agreement in method comparison studies. Stat. Methods Med. Res. 135–160.
Burgess, A., Gruzelier, J., 1993. Individual reliability of amplitude distribution in topographical mapping of EEG. Electroencephalogr. Clin. Neurophysiol. 86, 219–223.
Cacioppo, J.T., Tassinary, L.G., Berntson, G., 2007. Handbook of Psychophysiology. Cambridge University Press, Cambridge, England.
Cannon, R.L., Baldwin, D.R., Shaw, T.L., Diloreto, D.J., Phillips, S.M., Scruggs, A.M., Riehl, T.C., 2012. Reliability of quantitative EEG (qEEG) measures and LORETA current source density at 30 days. Neurosci. Lett. 518, 27–31.
Catafau, A.M., Bullich, S., Danús, M., Penengo, M.M., Cot, A., Abanades, S., Farré, M., Pavía, J., Ros, D., 2008. Test-retest variability and reliability of 123I-IBZM SPECT measurement of striatal dopamine D2 receptor availability in healthy volunteers and influence of iterative reconstruction algorithms. Synapse 62, 62–69.
Chi, Y.M., Jung, T.P., Cauwenberghs, G., 2010. Contact and noncontact biopotential electrodes: methodological review. IEEE Rev. Biomed. Eng. 3, 106–119.
Clarke, A.R., Barry, R.J., Mccarthy, R., Selikowitz, M., 2001. Age and sex effects in the EEG: development of the normal child. Clin. Neurophysiol. 112, 806–814.
Collado-Mateo, D., Adsuar, J.C., Olivares, P.R., Cano-Plasencia, R., Gusi, N., 2015. Using a dry electrode EEG device during balance tasks in healthy young-adult males: test-retest reliability analysis. Somatosens. Mot. Res. 32, 219–226.
Corsi-Cabrera, M., Galindo-Vilchis, L., Del-Rio-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. Clin. Neurophysiol. 118, 9–21.
Egerton, A., Demjaha, A., Mcguire, P., Mehta, M.A., Howes, O.D., 2010. The test-retest reliability of 18F-DOPA PET in assessing striatal and extrastriatal presynaptic dopaminergic function. NeuroImage 50, 524–531.
Ekandem, J.I., Davis, T.A., Alvarez, I., James, M.T., Gilbert, J.E., 2012. Evaluating the ergonomics of BCI devices for research and experimentation. Ergonomics 55, 292–298.
Enoch, M.-A., Shen, P.-H., Ducci, F., Yuan, Q., Liu, J., White, K.V., Albaugh, B., Hodgkinson, C.A., Goldman, D., 2008. Common genetic origins for EEG, alcoholism and anxiety: the role of CRH-BP. PLoS One 3, e3620.
Fein, G., Galin, D., Johnstone, J., Yingling, C.D., Marcus, M., Kiersch, M.E., 1983. EEG power spectra in normal and dyslexic children. I. Reliability during passive conditions. Electroencephalogr. Clin. Neurophysiol. 55, 399–405.
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. Int. J. Neurosci. 68, 255–261.
Forn, C., Barros-Loscertales, A., Escudero, J., Benlloch, V., Campos, S., Antònia Parcet, M., Ávila, C., 2007. Compensatory activations in patients with multiple sclerosis during preserved performance on the auditory N-back task. Hum. Brain Mapp. 28, 424–430.

Gargiulo, G., Calvo, R.A., Bifulco, P., Cesarelli, M., Jin, C., Mohamed, A., Van Schaik, A., 2010. A new EEG recording system for passive dry electrodes. Clin. Neurophysiol. 121, 686–693.

Gasser, T., Bächer, P., Steinberg, H., 1985. Test-retest reliability of spectral parameters of the EEG. Electroencephalogr. Clin. Neurophysiol. 60, 312–319.

Gevins, A., Smith, M.E., Mcevoy, L.K., Ilan, A.B., Chan, C.S., Jiang, A., Sam-Vargas, L., Abraham, G., 2011. A cognitive and neurophysiological test of change from an individual's baseline. Clin. Neurophysiol. 122, 114–120.

Gevins, A., Mcevoy, L.K., Smith, M.E., Chan, C.S., Sam-Vargas, L., Baum, C., Ilan, A.B., 2012. Long-term and within-day variablity of working memory performance and EEG in individuals. Clin. Neurophysiol. 123, 1291–1299.

Gudmundsson, S., Runarsson, T.P., Sigurdsson, S., Eiriksdottir, G., Johnsen, K., 2007. Reliability of quantitative EEG features. Clin. Neurophysiol. 118, 2162–2171.

Guevara, M.A., Corsi-Cabrera, M., 1996. EEG coherence or EEG correlation? Int. J. Psychophysiol. 23, 145–153.

Handy, T.C., 2005. Event-Related Potentials: A Methods Handbook. MIT press.

Hatz, F., Hardmeier, M., Bousleiman, H., Rüegg, S., Schindler, C., Fuhr, P., 2014. Reliability of fully automated versus visually controlled pre-and postprocessing of resting-state EEG. Clin. Neurophysiol.

Jaeggi, S.M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., Perrig, W.J., 2010. The relationship between n-back performance and matrix reasoning—implications for training and transfer. Intelligence 38, 625–635.

Jatoi, M.A., Kamel, N., Malik, A.S., Faye, I., Begum, T., 2014. A survey of methods used for source localization using EEG signals. Biomed. Signal Process. Control 11, 42–52.

Johnstone, S., Blackman, R., Bruggemann, J., 2012. EEG from a single channel dry sensor recording device. Clin. Electroencephalogr. Neurosci. 43, 112–120.

Kennett, R., 2012. Modern electroencephalography. J. Neurol. 259, 783–789.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P.V., Braeker, E., Kleinlogel, H., Isenhart, R., John, E.R., 2002. Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. NeuroImage 16, 41–48.

Lewis, R.S., Weekes, N.Y., Wang, T.H., 2007. The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health. Biol. Psychol. 75, 239–247.

Lovibond, P.F., Lovibond, S.H., 1995. The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the Beck depression and anxiety inventories. Behav. Res. Ther. 33, 335–343.

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.

Lyneham, H.J., Sburlati, E.S., Abbott, M.J., Rapee, R.M., Hudson, J.L., Tolin, D.F., Carlson, S.E., 2013. Psychometric properties of the child anxiety life interference scale (CALIS). J. Anxiety Disord. 27, 711–719.

Martín-Buro, M.C., Garcés, P., Maestú, F., 2016. Test-retest reliability of resting-state magnetoencephalography power in sensor and source space. Hum. Brain Mapp. 37, 179–190.

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., Pietrzak, R.H., 2009. Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Arch. Clin. Neuropsychol. 24, 165–178.

Mcevoy, L.K., Smith, M.E., Gevins, A., 2000. Test–retest reliability of cognitive EEG. Clin. Neurophysiol. 111, 457–463.

Messer, S.C., Angold, A., Costello, E.J., Loeber, R., Van Kammen, W., Stouthamer-Loeber, M., 1995. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: factor composition and structure across development. Int. J. Methods Psychiatr. Res. 5, 251–262.

Mollica, C.M., Maruff, P., Collie, A., Vance, A., 2005. Repeated assessment of cognition in children and the measurement of performance change. Child Neuropsychol. 11, 303–310.

Mostow, J., Chang, K.-M., Nelson, J., 2011. Toward exploiting EEG input in a reading tutor. Artif. Intell. Educ. 4, 230–237.

Näpflin, M., Wildi, M., Sarnthein, J., 2007. Test–retest reliability of resting EEG spectra validates a statistical signature of persons. Clin. Neurophysiol. 118, 2519–2524.

Naunheim, R.S., Treaster, M., English, J., Casner, T., Chabot, R., 2010. Use of brain electrical activity to quantify traumatic brain injury in the emergency department. Brain Inj. 24, 1324–1329.

Paasch, V., Hoosier, T.M., Accardo, J., Ewen, J.B., Slifer, K.J., 2012. Technical tips: performing EEGs and polysomnograms on children with neurodevelopmental disabilities. Neuroradiol. J. 52, 333–348.

Pollock, V.E., Schneider, L.S., Lyness, S.A., 1991. Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. Electroencephalogr. Clin. Neurophysiol. 79, 20–26.

Portney, L.G., Watkins, M.P., 2009. Foundations of Clinical Research: Applications to Practice. Pearson/Prentice Hall, Upper Saddle River, NJ.

Rijsdijk, M., Leijten, F.S.S., Slooter, A.J.C., 2008. Continuous EEG monitoring in the intensive care unit. Neth. J. Crit. Care 12, 157–162.

Rogers, J.M., Fox, A.M., 2012. Event-related potential practice effects on the paced auditory serial addition test (PASAT). Adv. Cogn. Psychol. 8, 281–291.

Rogers, J.M., Fox, A.F., Donnelly, J., 2015. Impaired practice effects following mild traumatic brain injury: an event-related potential investigation. Brain Inj. 29, 343–351.

Salinsky, M.C., Oken, B.S., Morehead, L., 1991. Test-retest reliability in EEG frequency analysis. Electroencephalogr. Clin. Neurophysiol. 79, 382–392.

Schlaggar, B.L., Brown, T.T., Lugar, H.M., Visscher, K.M., Miezin, F.M., Petersen, S.E., 2002. Functional neuroanatomical differences between adults and school-age children in the processing of single words. Science 296, 1476–1479.

Schomer, D.L., Lopes De Silva, F., 2010. Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Wolters Kluwer Health, Philadelphia, PA.

Seshia, S.S., Young, G.B., Zifkin, B.G., 2008. Guidelines for visual-sensitive EEG testing. Can. J. Neurol. Sci. 35, 133–135.

Shao, S.-Y., Shen, K.-Q., Ong, C.J., Wilder-Smith, E., Li, X.-P., 2009. Automatic EEG artifact removal: a weighted support vector machine approach with error correction. Biomed. Eng. IEEE Trans. 56, 336–344.

Sheorajpanday, R.V.A., Nagels, G., Weeren, A.J.T.M., Van Putten, M.J.A.M., De Deyn, P.P., 2009. Reproducibility and clinical relevance of quantitative EEG parameters in cerebral ischemia: a basic approach. Clin. Neurophysiol. 120, 845–855.

Shrout, P.E., 1998. Measurement reliability and agreement in psychiatry. Stat. Methods Med. Res. 7, 301–317.

The Psychological Corporation, 1998. The Paced Auditory Serial Addition Test. Harcourt Brace and Company, San Antonio, TX.

Tong, S., Thakor, N.V., 2009. Quantitative EEG Analysis Methods and Clinical Applications. Artech House, Boston.

Vaz, S., Falkmer, T., Passmore, A.E., Parsons, R., Andreou, P., 2013. The case for using the repeatability coefficient when calculating test–retest reliability. PLoS One 8, e73990.

Vuga, M., Fox, N.A., Cohn, J.F., George, C.J., Levenstein, R.M., Kovacs, M., 2006. Long-term stability of frontal electroencephalographic asymmetry in adults with a history of depression and controls. Int. J. Psychophysiol. 59, 107–115.

Wilk, C.M., Gold, J.M., Bartko, J.J., Dickerson, F., Fenton, W.S., Knable, M., Randolph, C., Buchanan, R.W., 2002. Test-retest stability of the repeatable battery for the assessment of neuropsychological status in schizophrenia. Am. J. Psychiatr. 159, 838–844.

Woodman, G.F., 2010. A brief introduction to the use of event-related potentials in studies of perception and attention. Atten. Percept. Psychophys. 72, 2031–2046.

Zuo, X.N., Xing, X.X., 2014. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. Neurosci. Biobehav. Rev. 45, 100–118.