Research Article / Originalarbeit

Complementary Medicine Research
Practice | Methods | Perspectives

Complement Med Res 2019;26:301–309
DOI: 10.1159/000497188

Received: March 17, 2018
Accepted: January 16, 2019
Published online: April 18, 2019

The Impact of Using Emotionally Arousing Stimuli on Muscle Response Testing Accuracy

Anne M. Jensen[a] Richard J. Stevens[b] Amanda J. Burls[c]

[a]Department of Continuing Professional Education and Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; [b]Department of Primary Care Health Sciences, Radcliffe Observatory Quarter, University of Oxford, Oxford, UK; [c]School of Health Sciences, City University London, London, UK

Keywords
Emotion · Emotional arousal · Affect · Kinesiology · Muscle weakness · Lie detection

Abstract
Introduction: Muscle response testing (MRT) is an assessment method used by 1 million practitioners worldwide, yet its usefulness remains uncertain. The aim of this study, one in a series assessing the accuracy of MRT, was to determine whether emotionally arousing stimuli influence its accuracy compared to neutral stimuli. ***Methods:*** To assess diagnostic test accuracy 20 MRT practitioners were paired with 20 test patients (TPs). Forty MRTs were performed as TPs made true and false statements about emotionally arousing and neutral pictures. Blocks of MRT alternated with blocks of intuitive guessing (IG). ***Results:*** MRT accuracy using emotionally arousing stimuli was different than when using neutral stimuli. However, MRT accuracy was found to be significantly better than IG and chance. Similar to previous studies in this series, this study failed to detect any characteristic that consistently influenced MRT accuracy. ***Conclusion:*** Using emotionally arousing stimuli had no effect on MRT accuracy compared to using neutral stimuli. This study would have been strengthened by adding personally relevant lies instead of impersonal stimuli. A limitation of this study is its lack of generalizability to other applications of MRT. This study shows that a simple yet robust methodology for assessing MRT as a diagnostic tool can be implemented effectively.

Die Auswirkung von emotional erregenden Reizen auf die Testgenauigkeit beim Muskelreaktionstest

Schlüsselwörter
Emotion · Emotionale Erregung · Affekt · Kinesiologie · Muskelschwäche · Lügendetektion

Zusammenfassung
Einleitung: Die Muskelreaktionstestung (MT) ist eine Untersuchungsmethode, die weltweit von einer Million Praktikern angewendet wird, deren Nutzen jedoch unklar ist. Mit der vorliegenden Studie (die zu einer Serie von Studien zur Bewertung der MT gehörte) sollte untersucht werden, ob emotional erregende Reize die Testgenauigkeit im Vergleich zu neutralen Reizen beeinflussen. ***Methoden:*** Zur Bestimmung der diagnostischen Testgenauigkeit wurden 20 MT-Praktiker und 20 Testpersonen (TPs) gepaart und 40 MTs durchgeführt, bei denen die TPs wahre und falsche Aussagen über emotional erregende und neutrale Abbildungen machten. Dabei wurden abwechselnd Blöcke mit MT und Blöcke mit intuitivem Raten ("intuitive guessing") verwendet. ***Ergebnisse:*** Die Genauigkeit der MT fiel bei Verwendung emotional erregender Reize anders aus als bei Verwendung neutraler Reize. Allerdings war die Genauigkeit der MT signifikant höher als bei intuitivem Raten und rein zufälliger Auswahl. Ähnlich wie in den vorherigen Studien dieser Serie fand sich auch in dieser Studie kein Merkmal mit einheitli-

KARGER

E-Mail karger@karger.com
www.karger.com/cmr

Dr. Anne M. Jensen, DC, MSc, DPhil
Wolfson College
Oxford OX2 6UD (UK)
E-Mail dranne@drannejensen.com

chen Auswirkungen auf die MT-Genauigkeit. ***Schlussfolgerung:*** Die Verwendung emotional erregender Reize hatte im Vergleich zu neutralen Reizen keinen Einfluss auf die Genauigkeit der MT. Die Aussagekraft der Studie wäre höher gewesen, wenn individuell relevante Lügen anstelle von unpersönlichen Reizen verwendet worden wären. Eine Einschränkung der Studie besteht in ihrer fehlenden Verallgemeinerbarkeit auf andere MT-Anwendungen. Die Studie zeigt, dass sich die MT als diagnostisches Instrument mit einer einfachen, jedoch robusten Methode wirksam beurteilen lässt.

Introduction

Muscle response testing (MRT) is a type of manual muscle testing used by many complementary and alternative therapists, including kinesiologists, chiropractors, naturopaths, physiotherapists, osteopaths, and psychologists, and is estimated to be used by over 1 million people worldwide [1]. Previous research suggests that at distinguishing lies from truth, it can be sufficiently accurate (65.9%); however, little is known about factors that may influence its accuracy [2].

The polygraph, while its efficacy remains controversial, can also achieve acceptable accuracy [3]. It operates on the principle that lying can cause specific physiological changes in the liar, and previous research has found that the higher the stakes of the lie, the more accurate the polygraph reading [3]. A similar theory is used to explain the effect of MRT: a stress in the body is related to a weakening of a muscle [4, 5]. Therefore, since MRT has been found to achieve a certain degree of accuracy at detecting lies using affect-neutral stimuli [1], it is hypothesized that the use of emotionally arousing stimuli may serve to improve accuracy.

Pictures and words can cause physiological and behavioral reactions, similar to the objects they represent [6–8]. The International Affective Picture System (IAPS) is a database of over 1,000 pictures that were rated by a large group of people (men and women) for the feelings they evoked [9]. Likewise, the Affective Norms for English Words is a database of English words with associated emotional ratings, which complements the IAPS [10]. Together these databases provide an emotional stimulus standardization necessary for rigorous scientific analysis.

This study used pictures and words from these databases, with specific valences known to be either neutral or emotionally arousing. The primary aim of this study was to determine whether the use of emotionally arousing stimuli had an impact on MRT accuracy, compared to affect-neutral stimuli.

Fig. 1. An example of muscle response testing: a practitioner (right) performs MRT on a patient (left) – using the patient's right deltoid muscle.

Methods

This study is a prospective study of diagnostic test accuracy. No participant was assessed prior to enrolment. This protocol received ethics committee approval by the Oxford Tropical Research Ethics Committee (OxTREC; Approval #41-10) and the Parker University Institutional Review Board for Human Subjects (Approval # R17_10). Also, this study protocol was registered with two clinical trials registries: the Australian New Zealand Clinical Trials Registry (ANZCTR; www.anzctr.org.au) and the US-based ClinicalTrials.gov. Written informed consent was obtained from all participants, and all other tenets of the Declaration of Helsinki were upheld. This paper was written in accordance with the Standards for the Reporting of Diagnostic Test Accuracy Studies (STARD) guidelines [11–13].

The methodology of this study followed closely to that of previous studies in this series [2, 14–16], with the exception that emotionally arousing pictures were mixed with the affect-neutral pictures from the database used in previous studies.

Summary of Testing Methods

Each practitioner performed a series of 40 muscle tests using MRT on volunteer test patients (TPs) after they spoke true and false statements about a picture they viewed on a computer screen. If the MRT outcome was *strong*, the practitioner interpreted this to mean that the statement was *true*, and if the MRT outcome was *weak*, the practitioner interpreted this to mean that the statement was *false*. A secondary index test was also enacted whereby practitioners had to make an intuitive guess (IG) whether the spoken statement was true or false using only visual, auditory, and kinesthetic cues (no MRT). Each pair completed 40 MRTs and 40 IGs. Blocks of 10 MRTs alternated with blocks of 10 IGs until 40 of each were performed.

Participants and Setting

Two groups of participants were recruited: (a) health care practitioners ($n = 20$) who routinely use MRT in practice ("practitioners") and (b) test patients ($n = 20$; "TPs"). Recruitment of participants was by direct contact (via e-mail or telephone), social media, and word of mouth, in the US state of California. Any volunteer was eligible if he or she was aged 18–65 years, had fully functioning

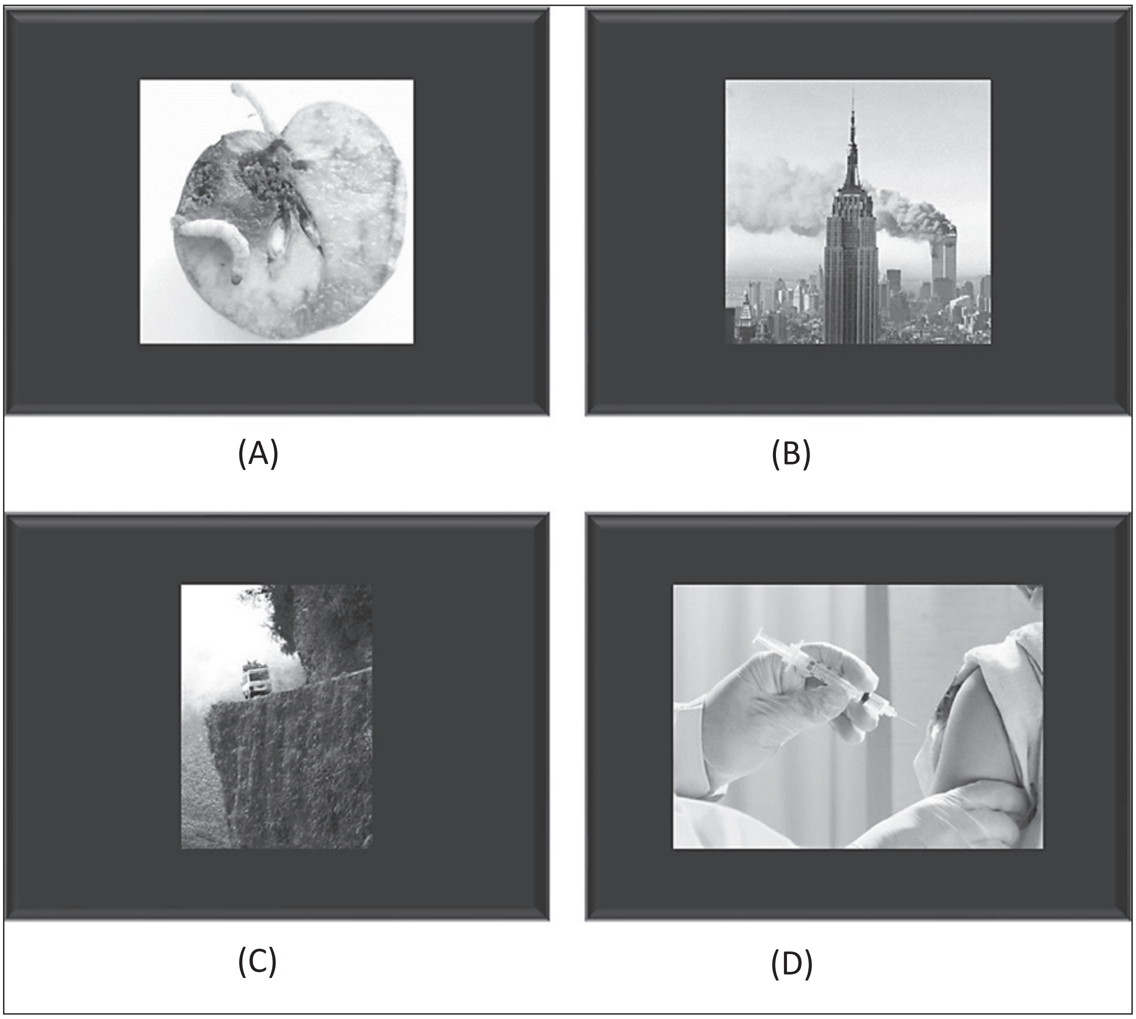

Fig. 2. Examples of emotionally arousing visual stimuli. **A–D** Examples that could have been presented to a test patient during either the MRT or the intuitive guessing blocks.

and pain-free upper extremities, and was fluent in English. Volunteers were excluded if they were blind, deaf, or mute. All practitioners who wished to participate and met the inclusion criteria were enrolled, regardless of their profession, MRT technique(s) used, or extent of MRT expertise or experience.

A mixture of MRT-naïve and non-MRT-naïve TPs were included; all TPs were naïve to the paradigm under investigation and to the interpretation of the MRT by the practitioner. That is, no TP was explicitly told that after a true statement, a strong MRT response was expected, and after a false statement, a weak MRT was expected.

In addition, there was a mix of TPs who were previously acquainted with their paired practitioner and those who were not. No practitioner's muscle testing ability was assessed in any way prior to enrolment. Practitioners and TPs were paired randomly by the principal investigator.

The Primary Index Test: MRT

During an MRT, an external force is applied to a body extremity and resisted by a particular muscle. At first the patient holds a specific joint in a fixed position, usually in partial flexion. The practitioner then applies pressure, usually into extension, as the patient resists this pressure using an isometric contraction. For example, in this study, the practitioner may have asked the TP to hold his shoulder (i.e., the glenohumeral joint) in 90-degree flexion, palm facing down, while he tests the anterior deltoid (Fig. 1). The practitioner then subjectively determines whether the muscle went "weak" or stayed "strong." Practitioners may vary in the amount of pressure applied and the location of the practitioner's hand [17]. The location is routinely on the distal forearm of the patient, just proximal to the wrist joint, but for the purposes of this study practitioners were instructed to follow their usual clinical practice in muscle testing. Also, the 40 MRTs were broken up into blocks of 10, and alternated with a secondary index test, IG.

The Secondary Index Test: Intuitive Guessing

The IG blocks proceeded in a fashion similar to the MRT blocks, except that no MRT was used to detect a false statement. Instead, the practitioner was asked to guess the verity of the statement using only visual, auditory, and kinesthetic cues. The 40 IGs were also broken up into 4 blocks of 10 and alternated with the MRT blocks. Pairs always started with an MRT block and ended with an IG block.

The Stimuli

The visual stimuli presented were a mix of affect-neutral pictures from the database used in previous studies in this series and pictures that were emotionally arousing. The emotionally arousing

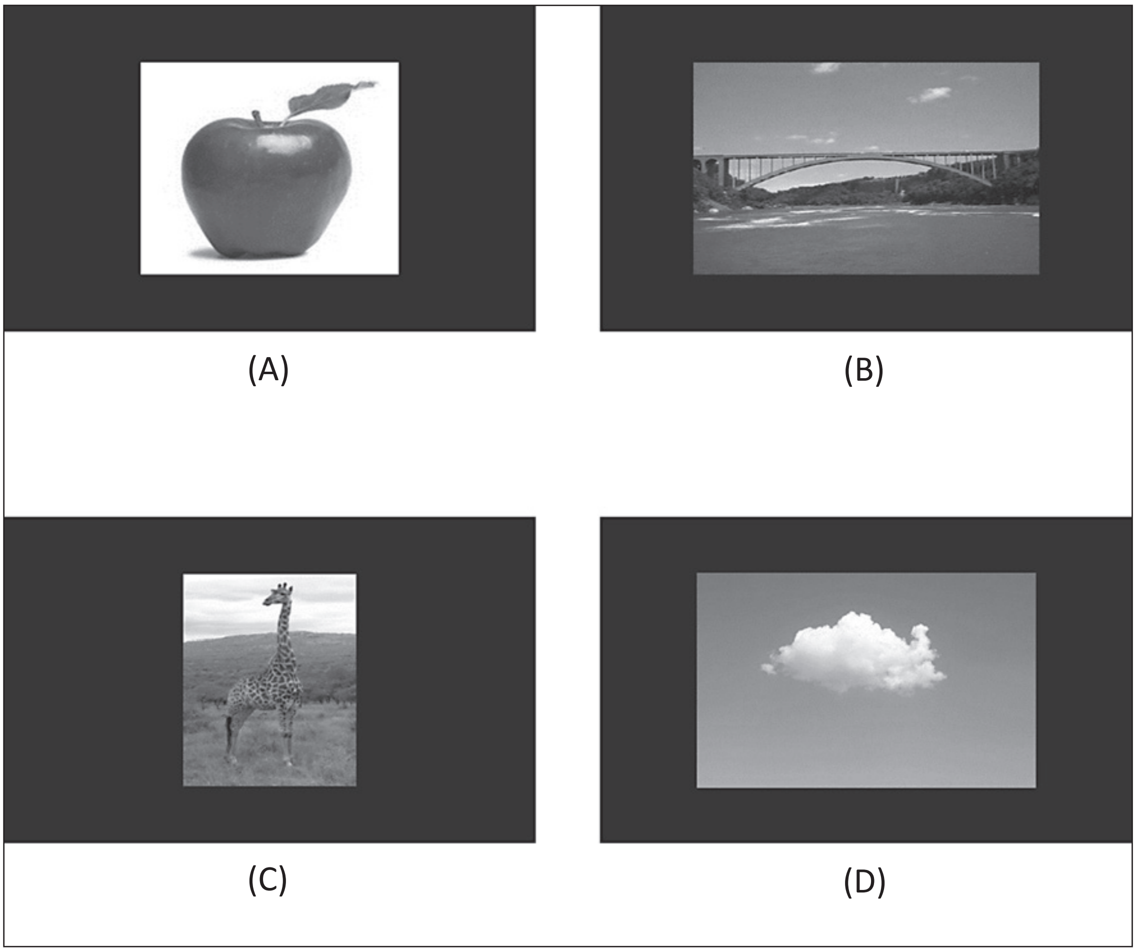

Fig. 3. Examples of neutral visual stimuli. **A–D** Examples that could have been presented to a test patient during either the MRT or the intuitive guessing blocks.

pictures were chosen for their similarity to those in the IAPS (National Institute of Mental Health Center for Emotion and Attention, University of Florida, Gainesville, FL, USA), which had mean arousal levels above 7.0, and supplemented with additional similar pictures [18]. Whenever possible, the pictures were paired with statements containing words from the Affective Norms for English Words (National Institute of Mental Health Center for Emotion and Attention, University of Florida, Gainesville, FL, USA), which had a mean arousal valence above 4.8 [10]. (For examples of emotionally arousing pictures, see Fig. 2, and for examples of neutral pictures, see Fig. 3).

For this study, 40 emotionally arousing and 40 affect-neutral pictures were placed into a database. Of the 80 pictures, half of each group was allocated to the MRT blocks and half to the IG blocks. In addition, half of these subgroups were paired with true statements and half with false statements so that the prevalence of lies was fixed at 0.50. The order of stimuli was randomly chosen and presented using DirectRT™ Research Software (Empirisoft Corporation, New York, NY, USA) so that each pair was presented with a unique sequence of stimuli.

The Testing Scenario

TPs viewed pictures on a computer screen placed out of view of the practitioners. While viewing a picture selected at random by computer, the TPs were given instructions via an earpiece, inaudible to the practitioners. Instructions took the form, "Say, 'I see a ________.'" As such, practitioners were blind to the verity of the statement and TPs were not. However, a previous study in this series found that blinding to the verity of the spoken statement made no significant difference to MRT accuracy [19, 20]. Furthermore, practitioners were intentionally not blinded to any visual, auditory, and kinesthetic cues of the TP, for two reasons: (1) because we aimed to make this assessment as clinically authentic as possible and (2) because the IG condition would establish whether this type of blinding produced any significant influence on accuracy. However, all participants were blind to study aims and were not informed of the proportions of true/false statements. Also, all participants completed the same pre- and post-testing questionnaires. For the configuration of the testing scenario see Figure 4, and for the participant flow diagram, see Figure 5.

Statistical Methods

Based on a previous study in this series in which the accuracy of manual MRT for lie detection had a mean of 66% and standard deviation of 13% across participants [2], we estimated that a sample size of 20 participants would have greater than 99% power to detect an overall accuracy of 66% compared to 50%.

Error-based measures of accuracy are reported in the form of *overall fraction correct* [21] with a 95% confidence interval (95% CI). The statistic *overall fraction correct* was used for accuracy, rather than the more conventional statistics of *sensitivity* and

Complement Med Res 2019;26:301–309
DOI: 10.1159/000497188

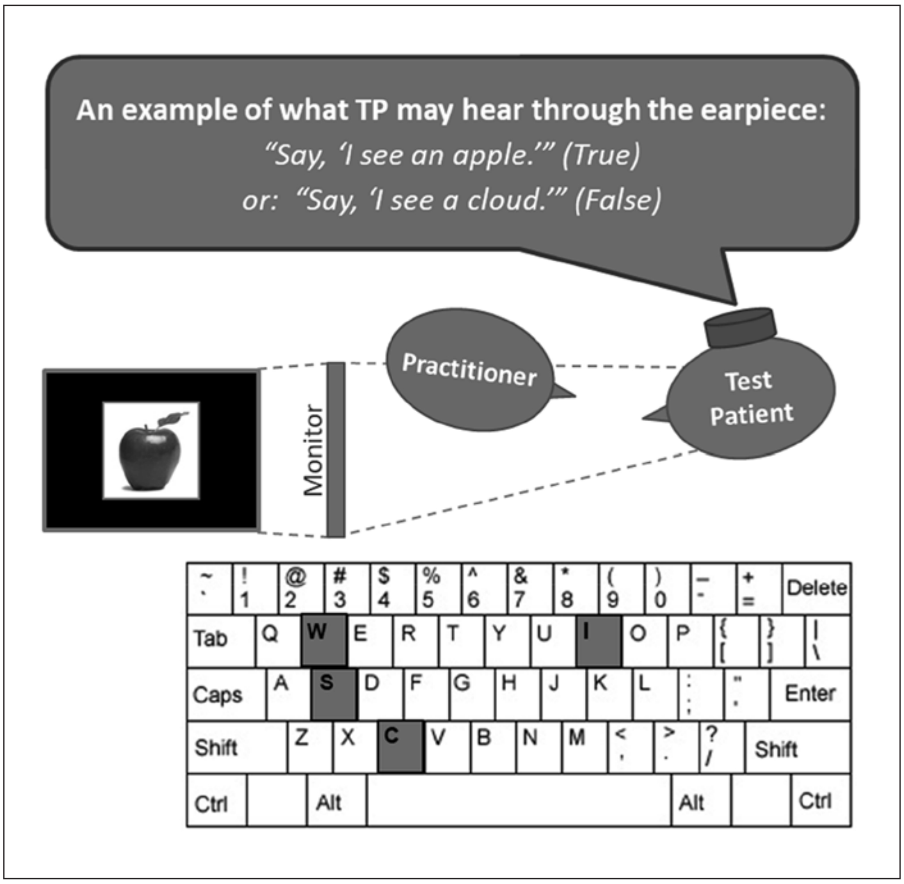

Fig. 4. Testing scenario layout: the test patient viewed a monitor which the practitioner could not see, and he had an ear piece in his ear through which he received instructions. After the muscle test, the practitioner entered his results on a keyboard.

specificity, because unlike other diagnostic tests, the outcomes for MRT (i.e., *strong* and *weak*) held equal weight or importance. All data were analyzed using STATA 17.0 (Stata Corp. LP, College Station, TX, USA), specifically the commands ttest and *pwcorr, sig*.

Results

Participants

Twenty unique practitioner-TP pairs were enrolled in October and November 2011, including 13 female and 7 male practitioners, and 12 female and 8 male TPs. Of the 20 practitioners, there were 14 chiropractors, 1 mental health professional, 1 acupuncturist, 1 naturopath, 1 other health professional and 2 other professionals (non-health).

Twelve practitioners were in full-time practice, 8 were in part-time practice, and all enrolled were currently practicing. The practitioners' mean (SD) number of years in practice was 16.8 (12.7) years. The mean age for practitioners was 48.5 (12.0) years, and for TPs 37.9 (12.4) years. For a summary of practitioner demographics, see Table 1. Additionally, there were 13 TPs who were MRT-naïve and 7 who had prior experience with MRT, and 8 TPs knew their paired practitioner and 12 who did not.

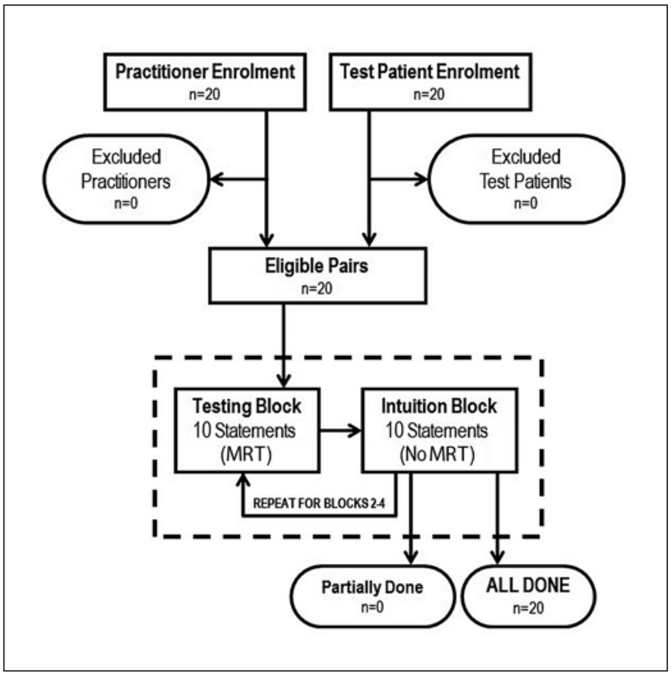

Fig. 5. Participant flow diagram. MRT, muscle response testing.

Table 1. Demographics of practitioners

	Practitioners (n = 20)
Gender (M:F)	11:9
Mean age (SD), years	48.5 (12.0)
Mean number of years in practice (SD)	16.8 (12.7)
Practitioner type, n	
Chiropractor	14
Mental health professional	1
Acupuncturist	1
Naturopath	1
Other health professional	1
Other professional	2
Practitioner practice status, n	
Full-time	12
Part-time	8
Not practicing	0
Mean years of MRT experience (SD)	17.6 (12.2)
Mean hours of MRT/day (SD)	8.6 (10.6)
Mean self-ranked MRT expertise[1] (SD)	3.5 (0.8)

MRT, muscle response testing; SD, standard deviation; M, male; F, female. [1] Self-ranked MRT expertise ranged from 0 = none to 4 = expert.

Table 2. The influence of stimulus valence on MRT accuracy

Stimulus type	MRT accuracy		p value
	mean	95% CI	
Pictures			
Affect-neutral	0.661	0.561–0.762	0.34
Emotionally arousing	0.631	0.542–0.720	
Words			
Affect-neutral	0.658	0.552–0.764	0.51
Emotionally arousing	0.638	0.555–0.721	
Pictures and words combined			
Affect-neutral	0.659	0.560–0.757	0.35
Emotionally arousing	0.632	0.544–0.720	

MRT, muscle response testing; CI, confidence interval.

Test Results

Pairs took between 10 and 45 min to complete their participation. All pairs completed all testing in full. Aside from TP arm fatigue, there were no adverse events reported from any testing. All accuracies were normally distributed, so parametric statistics were used, primarily the Student t test.

MRT and IG Accuracies

Since the primary aim of this study was to determine whether the use of emotionally arousing stimuli had an impact on MRT accuracy, compared to affect-neutral stimuli, we first report and compare their accuracies. Then we report accuracies using all stimuli, since these results parallel the previous studies in this series.

Using only emotionally arousing stimuli, the mean (95% CI) MRT accuracy for MRT was 0.632 (0.544–0.720), which was significantly different from chance (0.500; p = 0.01) but not from the mean (95% CI) IG accuracy, 0.545 (0.491–0.599; p = 0.09). In addition, the mean MRT accuracy using only emotionally arousing stimuli was not significantly different from MRT accuracy using only affect-neutral stimuli (p = 0.35). Using only affect-neutral stimuli, the mean (95% CI) MRT accuracy for MRT was 0.659 (0.560–0.757), which was significantly different from both chance (p = 0.01) and the mean (95% CI) IG accuracy, 0.508 (0.459–0.556; p = 0.01). For a summary of these results, see Table 2.

Using all stimuli, the mean (95% CI) accuracy for MRT was 0.648 (0.558–0.737), which was significantly different from the mean (95% CI) IG accuracy, 0.526 (0.488–0.564; p = 0.01), and also chance (p < 0.01). It was also noted that MRT accuracy scores ranged from 20% correct to 100% correct within this sample of 20 pairs.

Potential Influencing Factors

A post-participation questionnaire was administered where TPs were asked if they guessed the paradigm under investigation, and 9 TPs reported guessing the paradigm, whereas 11 TPs did not report guessing the paradigm. For all stimulus types (all, emotionally arousing and affect-neutral stimuli), these 2 groups did not differ significantly in MRT accuracy (p = 0.10, p = 0.06, and p = 0.21, respectively). In addition, comparing the mean MRT accuracies of those pairs whose TPs were MRT-naïve (n = 13) to those pairs whose TPs were not MRT-naïve (n = 7) for all stimuli, for just emotionally arousing stimuli and for just affect-neutral stimuli also found no significant differences (p = 0.26, p = 0.50, and p = 0.21). Lastly, comparing the mean MRT accuracies of those pairs who knew each other (n = 8) to those pairs who did not know each other (n = 12) for all stimuli, for just emotionally arousing stimuli and for just affect-neutral stimuli also found no significant differences (p = 0.48, p = 0.22, and p = 0.70).

Another interesting result that was achieved was that during correlation testing, this study did not find any significant correlations between MRT accuracy and the amount of a practitioner's clinical experience (p = 0.75).

Discussion

Statement of Principal Findings

The main finding of this study was that the use of emotionally arousing stimuli was found not to improve MRT accuracy scores. Furthermore, using only emotionally arousing stimuli, the MRT accuracy scores were not significantly different from IG scores either. One interesting result was that regardless of the valence of stimuli (i.e., affect-neutral, emotionally arousing, or both combined), MRT accuracy was consistently found to be significantly better than chance.

In addition, findings of this study also replicated those of previous studies in this series, in that (when using the data from all stimuli) MRT accuracy was significantly better than IG and chance at distinguishing lies from truth.

Strengths and Limitations

The main strengths of this study include the heterogeneity of the sample, the use of a true gold standard as a reference standard, and its simple but robust methodology. The limitations include its lack of generalizability to other applications of MRT and its lack of similarity to true clinical settings. Another limitation of this study was that the sample size did not allow for meaningful subgroup analyses.

One aspect of this study that could have been improved is the type of lie that TPs were asked to tell. In previous studies of this series, TPs were asked to tell lies about affect-neutral stimuli (i.e., pictures and words); however, in this study, some of the neutral stimuli were exchanged for emotionally arousing ones, with the expectation that the stress response would be enhanced, making it easier for MRT to distinguish lies from truth. Nevertheless, because MRT accuracy was not influenced by valence of stimuli, simply changing the valence of the stimuli was not sufficient to invoke a change in MRT accuracy. O'Sullivan et al. [3] found that police professionals were more successful at detecting high-stakes lies than low-stakes lies. She defines a high-stakes lie as a lie that has personal relevance or that is important to the liar, such as a strongly held opinion or about a highly stressful personal event or one which had a significant consequence (positive or negative). On the other hand, she defined a low-stakes lie as one that is relatively trivial, such as a "white lie," or one where the reward or punishment is immaterial. Therefore, it is possible that this current study would have been strengthened if the lies that the TPs were asked to tell were personally relevant to the TP.

Comparisons to Other Studies

We have previously shown that MRT is better than chance alone or intuitive guessing alone at distinguishing truth from lies, in a similar set-up using affect-neutral stimuli [2]. The present study confirms this finding for affect-neutral stimuli and shows that it also holds for emotionally arousing stimuli with similar effect size. We have commented previously [2] that other studies in the field have had weaknesses in methods or reporting such as unclear blinding [22] or lack of a gold standard [23].

Possible Explanation of Results

Using emotionally arousing stimuli did not achieve MRT accuracies that were any better or worse than using affect-neutral stimuli. The most plausible explanation for this is that it is not the *emotional valence* of the picture being presented that is important, but how *personally relevant* the lie is to the TP that may be a mediator [3].

Implications for Clinical Practice

From the results of this study, two recommendations for clinical practice surfaced. Firstly, it is commonly thought that learning MRT is a clinical skill and as such is learned just like any other clinical skill (e.g., taking blood pressure readings, taking a spinal radiograph, or performing an otoscopic exam). However, the results of this study, supporting the findings of the previous studies of this series, suggest that length or extent of clinical experience with MRT did not correlate with MRT accuracy. In effect, new practitioners and potential practitioners should be encouraged by this.

Second, it is suggested that MRT practitioners have their own MRT accuracies assessed. This information may inform clinical practice. Each can then use this information to adapt his MRT sessions accordingly.

In addition, this study supported the findings of previous studies in this series, in that once again it demonstrated that rigorous methodology does exist that can consistently estimate the diagnostic accuracy of MRT for distinguishing lies from truth.

Unanswered Questions and Future Research

One of the most important targets of future MRT research must still be to ascertain what factors promote more accurate MRT. We noted that MRT accuracy scores can range from 20% correct to 100% correct within a sample of 20 pairs. One cannot help but wonder what features the 100% accurate pair possess that the 20% accurate pair does not – and vice versa. Also, it would be interesting to know whether these characteristics are innate or are acquired (i.e., learned).

In addition, instead of focusing on the valence of the stimuli, future studies may want to compare MRT accuracy using personally relevant and high-stakes lies to MRT accuracy using low-stakes, irrelevant lies, such as those used in previous studies. Researchers may want to use a multiprong approach toward devising high-stake lies: (1) by using highly stressful, personal events, and/or (2) by amplifying the value of the reward and/or the severity of the punishment, and/or (3) by a combination of both [3]. If using high-stakes lies does turn out to improve the accuracy of MRT, then if a practitioner tailors the statements he asked a patient to speak so that they are personally relevant, it is conceivable that his accuracy score may improve.

Conclusion

This study used a mixture of emotionally arousing stimuli and affect-neutral stimuli as pairs to perform MRT, and the differences in MRT accuracies were compared. The results of this study suggest that MRT using emotionally arousing stimuli is no more or less accurate than MRT using affect-neutral stimuli, contradicting the initial study hypothesis. However, it was found that (using all stimuli combined) MRT can be used with significant accuracy to distinguish lies from truths, compared to both IG and chance. The study would have been strengthened by adding personally relevant, high-stakes lies instead of lies about impersonal yet emotionally arousing stimuli. The primary limitation of this study is its lack of generalizability to other applications of MRT. The main strengths of this study were its choice of a "gold standard" reference standard and its high degree of blinding. Finally, this study is further evidence that a simple yet robust methodology for assessing the value of MRT as a diagnostic tool can be developed and implemented effectively.

Acknowledgments

The authors would like to thank those who volunteered their time to participate in this study and those practitioners who offered the use of their facilities for data collection.

Statement of Ethics

This protocol received ethics committee approval by the Oxford Tropical Research Ethics Committee (OxTREC; Approval #41-10) and the Parker University Institutional Review Board for Human Subjects (Approval # R17_10). Also, this study protocol was registered with two clinical trials registries: the Australian New Zealand Clinical Trials Registry (ANZCTR; www.anzctr.org.au) and the US-based ClinicalTrials.gov. Written informed consent was obtained from all participants, and all other tenets of the Declaration of Helsinki were upheld.

Disclosure Statement

No competing financial interests exist. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare that there are no conflicts of interest.

Author Contributions

All authors made substantial contributions to conception and design, and/or acquisition of data, and/or analysis and interpretation of data; all authors participated in drafting the article or revising it critically for important intellectual content; and all authors give final approval of the version to be submitted and any revised version. Concept development: A.M.J., A.J.B.; design: A.M.J., R.J.S., A.J.B.; supervision: A.M.J., R.J.S., A.J.B.; data collection: A.M.J.; data processing: A.M.J., R.J.S.; analysis/interpretation: A.M.J., R.J.S.; literature search: A.M.J.; writing: A.M.J.; critical review: A.M.J., R.J.S., A.J.B.; all authors read and approved the final manuscript.

References

1 Jensen AM. Estimating the prevalence of use of kinesiology-style manual muscle testing: A survey of educators. Adv Integr Med. 2015; 2(2):96–102.
2 Jensen AM, Stevens RJ, Burls AJ. Estimating the accuracy of muscle response testing: two randomised-order blinded studies. BMC Complement Altern Med. 2016;16(1):492.
3 O'Sullivan M, Frank MG, Hurley CM, Tiwana J. Police lie detection accuracy: the effect of lie scenario. Law Hum Behav. 2009 Dec;33(6): 530–8.
4 Walther DS. Applied kinesiology: synopsis, vol. 1, 2nd edn. Pueblo: Systems DC; 2000.
5 Lundberg U, Kadefors R, Melin B, Palmerud G, Hassmen P, Engstrom M, et al. Psychophysiological stress and EMG activity of the trapezius muscle. Int J Behav Med. 1994;1(4): 354–70.
6 Brouwer AM, van Wouwe N, Mühl C, et al. Perceiving blocks of emotional pictures and sounds: effects on physiological variables. Front Hum Neurosci 2013 Jun 21;7:295.
7 Kuperman V, Estes Z, Brysbaert M, Warriner AB. Emotion and language: valence and arousal affect word recognition. J Exp Psychol Gen. 2014 Jun;143(3):1065–81.
8 Holmes EA, Mathews A, Mackintosh B, Dalgleish T. The causal effect of mental imagery on emotion assessed using picture-word cues. Emotion. 2008 Jun;8(3):395–409.
9 Bradley MM, Lang PJ. The International Affective Picture System (IAPS) in the study of emotion and attention. In: Coan JA, Allen JJ, editors. Handbook of Emotion Elicitation and Assessment. Oxford: Oxford University Press; 2007. p. 29–46.
10 Bradley MM, Lang PJ. Affective Norms for English Words (ANEW): stimuli, instruction manual and affective ratings. Technical report C-1. Gainesville: The Center for Research in Psychophysiology, University of Florida; 1999.
11 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al.; Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem. 2003 Jan;49(1):7–18.
12 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al.; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ. 2003 Jan;326(7379):41–4.
13 Bossuyt PM, Leeflang MM. Chapter 6: developing criteria for including studies. In Cochrane handbook of systematic reviews of diagnostic test accuracy, version 40 [updated September 2008]. London: The Cochrane Collaboration; 2008.
14 Jensen AM, Stevens RJ, Burls AJ. Muscle testing for lie detection: grip strength dynamometry is inadequate. Eur J Integr Med. 2018 Jan; 2017(17):16–21.
15 Jensen AM, Stevens RJ, Burls AJ. Exploring the variation in muscle response testing accuracy through repeatability and reproducibility. PLoS One. Under review 2018.
16 Jensen AM, Stevens RJ, Burls AJ. Investigating the validity of muscle response testing: an attempt at blinding the patient using subliminal visual stimuli. Complement Ther Clin Pract. Under review 2018.
17 Schmitt WH Jr, Cuthbert SC. Common errors and clinical guidelines for manual muscle testing: "the arm test" and other inaccurate procedures. Chiropr Osteopat. 2008 Dec; 16(1):16.

18 Lang PJ, Bradley MM, Cuthbert BN. International Affective Picture System (IAPS): affective ratings of pictures and instruction manual. Technical Report A-8. Gainesville: University of Florida; 2008.
19 Jensen AM. Emerging from the mystical: muscle response testing is not an Ideomotor effect. Energy Psychol J. Forthcoming 2018.
20 Jensen AM. The accuracy and precision of kinesiology-style manual muscle testing. DPhil. Oxford, UK: University of Oxford; 2015.
21 Bossuyt PM. Defining biomarker performance and clinical validity. J Med Biochem. 2011;30(3):193–200.
22 Monti DA, Sinnott J, Marchese M, Kunkel EJ, Greeson JM. Muscle test comparisons of congruent and incongruent self-referential statements. Percept Mot Skills. 1999 Jun;88(3 Pt 1):1019–28.
23 Caruso W, Leisman G. A force/displacement analysis of muscle testing. Percept Mot Skills. 2000 Oct;91(2):683–92.