

The Oxford Studies: Validating Muscle Response Testing

Part II – Results and Implications

As mentioned in Part I of this article, I completed my DPhil (PhD) at Oxford University, where my research focused on assessing the validity of Muscle Response Testing (MRT) in a specific application: to distinguish true from false spoken statements. The rationale for the methods used was outlined previously, and in this part, I will discuss the specific methods of each study, report the results of this research and discuss its implications.

This series of studies consisted of 5 diagnostic test accuracy studies (see Table 1 below), and while the general methodologies of all studies remained consistent, specific elements were changed in an attempt to better understand how they influenced MRT accuracy.

Table 1 – An outline of this series of studies assessing the validity of MRT

Study 1 – Estimating the Accuracy of MRT

Study 2 – Replication of Study 1

Study 3 – Replacing the Practitioner with Grip Strength Dynamometry

Study 4 – Using Emotionally-arousing Stimuli

Study 5 – Estimating MRT Precision using a Round-robin Format

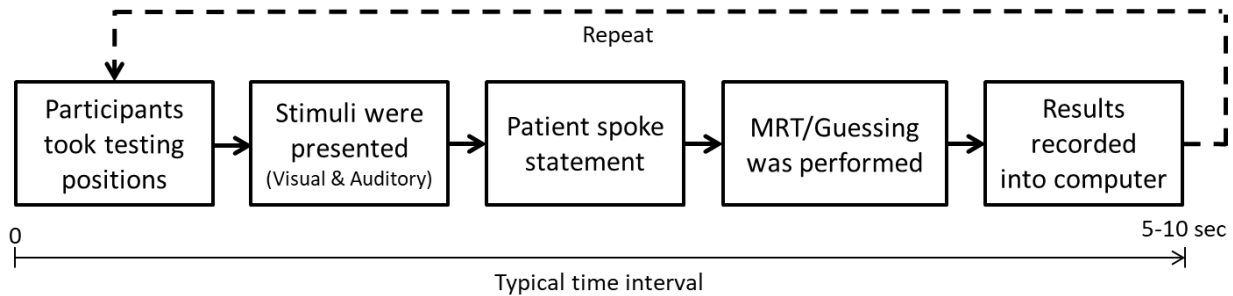
Summary of Testing Scenario

In all the studies, the patient viewed a computer screen on which was displayed pictures of everyday items (e.g. apple, basketball, tree, train, etc), and which instructed them (via an earpiece) what to say in relation to the picture. About half of the time, they were instructed to say true statements, and half of the time, false statements, and the order was randomly generated by the computer.

As mentioned in Part 1, the paradigm under investigation was that true statements result in a strong MRT outcomes, and false statements result in a weak MRT outcomes. Only the deltoid muscle was used for testing, however, how the participants were positioned was left up to the discretion of the practitioner, as long as they could not view each other's computer screen. Also, practitioners could perform any pre-testing procedures, as they were encouraged to do what they would normally do in practice.

All participants gave informed consent and completed demographic questionnaires prior to beginning the actual testing. The questionnaires also asked participants about their MRT experience, their confidence in MRT, etc. During the testing, for each MRT or intuitive guess, the sequence proceeded in this manner:

Figure 1 - The sequence of 1 repetition.



Participants proceeded in this manner until they completed all MRTs / guesses, and then completed another short questionnaire, after which they were done.

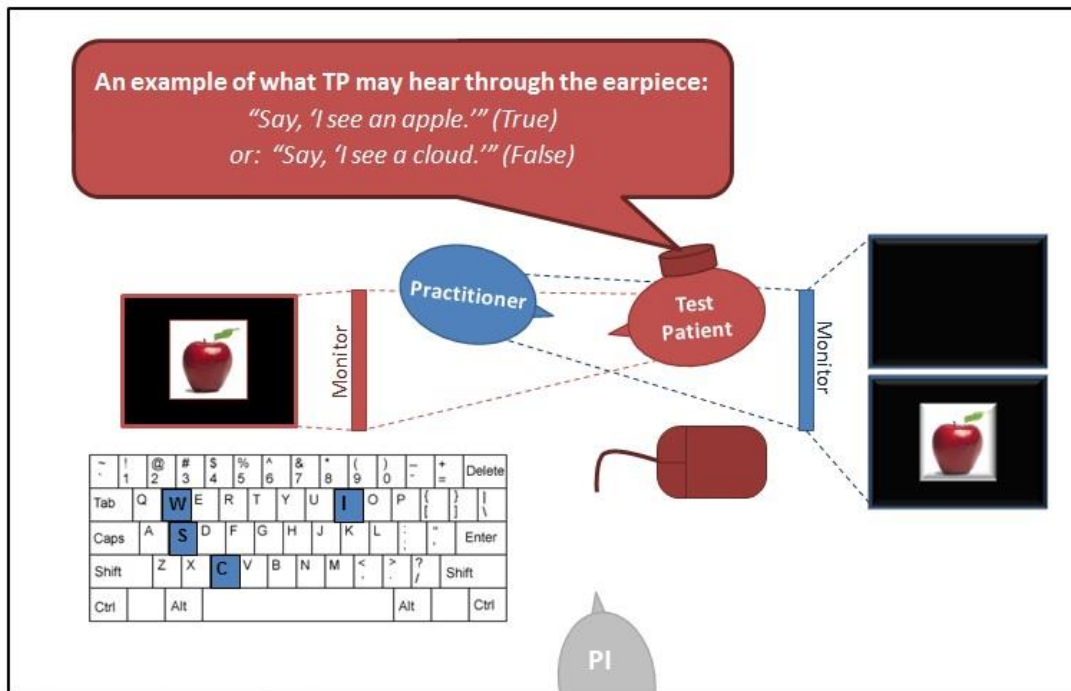
Specific Study Methods

As mentioned above, while each study followed the same basic protocol, certain details were changed to investigate different factors. Now I will summarize the details of each study.

Study 1 (Main Study) – Estimating the Accuracy of MRT

In the first study, 48 pairs of participants were recruited: 48 practitioners who use MRT and 48 patients who had no previous MRT experience. During this study, the practitioner also viewed a computer screen on which was shown either the same picture as the patient or a blank black screen (see Figure 2 below). In the second case, the practitioner was blind to the statement's truth, and it was only the blind tests that were used to calculate accuracy. Pairs performed 40 MRTs (broken up into blocks of 10), which alternated with blocks of 10 intuitive guesses.

Figure 2 - Study 1 testing scenario layout.



TP, Test Patient; PI, Principal Investigator

Study 2 – Replication of Study 1

The results of Study 1 (described below) were so impressive that my supervisors insisted I repeat it (i.e. replicate it). So, from the data collected in Study 1, another sample size calculation was performed and it was determined that only 20 pairs were needed to obtain similar results. Also, the practitioner's computer was removed, making the practitioners blind in all repetitions. I found in the first study that the second computer added an unnecessary degree of complexity, and by removing it, made data collection flow much smoother. One other change that my supervisors asked me to implement was for me to leave the testing room while the pairs were testing. In Study 1, I was present – mainly to ensure the smooth running of things – but they thought that my presence may have had an influence on the results. So, in Study 2, I left the room. All other aspects of Study 2 were identical to Study 1.

Study 3 – Replacing the Practitioner with Grip Strength Dynamometry

One of the criticisms of MRT is that it is not objective, meaning that practitioners (and patients) could seemingly influence the outcome. With this in mind, we sought a way to standardize the assessment, and therefore, improve its objectivity. One of my supervisors suggested to run a study where muscle strength assessed by a practitioner was replaced by muscle strength assessed by a machine – with the view determine if a device could be useful in distinguishing false from true statements. So, in Study 3, MRT was replaced by grip strength testing via dynamometry, however aside from that, the same protocol was followed. Twenty patients were recruited to perform 20 grip strength tests (10 right hand, 10 left hand) following the speaking of true and false statements. In this study, the average grip strengths after true statements were compared to the average grip strength after false statements. A statistically significant difference would mean that the dynamometer could also be used to detect false statements.

Study 4 – Using Emotionally-arousing Stimuli

In Studies 1-3 & 5, the pictures that were shown to patients, were of ordinary, neutral items (e.g. an apple, a bucket, a fence, a basketball, etc). While the results obtained were very good, we thought perhaps they might even improve if the pictures were emotionally-arousing or stressful. So, in Study 4, the database of pictures presented included a combination of neutral and emotionally-arousing images, and we followed the same protocol as Study 2: 20 practitioner-patient pairs, 40 MRTs, 40 intuitive guesses, and I left the room during testing. All other elements also remained the same.

Study 5 – Estimating MRT Precision using a Round-robin Format

For a test to be considered valid, it must be both accurate and precise. So, it was also necessary to assess MRT's precision, which would mean investigating if (under similar conditions) can MRT achieve the same results consistently. In other words, if a practitioner achieved 85% correct with one patient, did s/he achieve approximately 85% correct with other patients. One could also consider this the *stability of MRT accuracy*.

In order to assess this, we gathered 16 practitioners and 7 patients, in the same location. Each practitioner performed 20 MRTs and 20 intuitive guesses on each patient, following the same basic format as the previous 2 studies.

Summary of Results & their Implications

With over 400 participants evaluated in these studies, the data collected was extensive, and because of this, only some of the results can be reported here. However, the main findings are described below.

Table 2 - Summary of Accuracy Results

	<i>n</i>	MRT Accuracy*		Guessing Accuracy*		<i>p</i> -value
		Mean	95% CI	Mean	95% CI	
Study 1 - Main Study	48	0.659	0.623 - 0.695	0.474	0.449 - 0.500	<0.01*
Study 2 - Replication Study	20	0.594	0.541 - 0.647	0.514	0.483 - 0.544	0.01*
Study 4 - Using Emotional Stimuli	20	0.648	0.558 - 0.737	0.526	0.488 - 0.564	0.01*
Study 5 - Stability Study	96	0.616	0.578 - 0.654	0.507	0.484 - 0.530	<0.01*

MRT, Muscle Response Testing; CI, Confidence Interval; * Accuracy, defined as Overall Fraction Correct.

Result #1: MRT can accurately distinguish false from true statements.

In Table 2 (above), it can be seen that the average MRT accuracies in this series of studies ranged from 59.4% correct to 65.9% correct, and that in each study the MRT accuracy was significantly better than the average accuracies of intuitive guessing (i.e. each *p*-value was less than 0.05). While it is not shown in this table, it was also found that the MRT accuracies were significantly better than chance (i.e. 50-50). Therefore, it can be said that MRT is consistently more accurate than either guessing or chance at distinguishing false from true statements. This implies that MRT is not a chance occurrence, and that the success of MRT cannot be attributed to a practitioner's ability to "read" people.

Result #2: Factors that influence degree of MRT accuracy remain unknown.

When examining the accuracy scores of all participating pairs, we found the range of MRT accuracies surprisingly wide: 25-100%. This means that some practitioners got every MRT correct, while other practitioners got a score of half as good as chance. This lead us to wonder what the 100%-practitioners were doing that the 25%-practitioners were not, and vice versa. Because previous research found that experienced practitioners were more accurate than less experienced practitioners, we wanted to know if we could replicate these results, and if there was any factors or participant characteristics that may be attributed to better or worse accuracy scores. Therefore, we tracked on those factors listed in Table 3 (below). Correlation analyses revealed that none of these tracked characteristics consistently had any influence on MRT accuracy, including practitioner experience. That is, there was no difference in MRT accuracies between novice and experienced practitioners. Moreover, the results of our correlation analyses mean we do not know why some practitioners perform better than others.

Result #3: Practitioners did not seem to influence (or bias) the MRT outcome.

Another criticism of MRT is that it may appear that practitioners can influence (or bias) the result of the MRT, so I believe that it was important to specifically address this concern. In Study 1, the practitioners were blind to the verity of the statement during approximately ½ of the repetitions (total number of repetitions was 40). It was hypothesized that when the practitioners were not blind (that is, when they were viewing the same picture as the patient), their accuracy would be close to 100%, or at least significantly more accurate than when they were blind. However, this was not the case. There was no difference in MRT accuracies when they were blind and not blind (*p*=0.52). This may suggest that

practitioners did not consciously influence the outcome of the MRT, or in other words, they were doing honest MRT.

Table 3 - Participant characteristics tracked.

Practitioner characteristics	Patient characteristics
Age	Age
Gender	Gender
Profession	If patient had previous MRT experience
Number of years in practice	Confidence in practitioner
Years of MRT experience	Confidence in practitioner's MRT ability
Self-ranked MRT expertise	If patient knew practitioner
Practising status (full-time, part-time, not practising)	Self-rated degree of connection with practitioner
If practitioner knew patient	How well patient liked practitioner
Confidence in own ability to perform MRT	Self-rated anxiety
Self-rated degree of connection with patient	
How well practitioner liked patient	
Self-rated anxiety	

Result #4: Patients did not seem to influence (or bias) MRT accuracy.

A similar criticism to the one posed above is that it may also be possible for patients to influence (or bias) the result of MRT. During Study 1, all patients recruited had no previous experience with MRT (that is, they were MRT-naïve), and they were blind to the paradigm under investigation: they were not told that a strong MRT result indicated a true statement and a weak MRT result indicated a false statement. However, since it was not possible to blind them to the verity of the statements they were speaking, and since they may have been paying attention to each MRT outcome and deduced (i.e. guessed) the paradigm, and as a result, may have influenced the results. However, this also was not the case: those pairs whose patients reported guessing the paradigm (n=21) were no more accurate than those pairs whose patients did not report guessing the paradigm (n=27), reaching statistical significance ($p=0.38$)

In other studies in this series, a mixture of MRT-naïve and non-naïve patients were recruited, and it was hypothesized that those pairs with non-naïve patients may achieve higher accuracies than those with MRT-naïve patients. However, again this was consistently not the case. For example, in Study 2, when comparing the MRT accuracies of those pairs with MRT-naïve patients (n=11) and non-naïve patients (n=9), there was no significant difference in their average accuracies (0.634 and 0.544 respectively, $p=0.07$). It is also interesting to note that the naïve group had a higher accuracy, but the difference did not reach significance.

Result #5: MRT is not an ideomotor effect.

Psychologist/physiologist, William B. Carpenter, described the nonconscious modulation of muscular movement mediated by a heightened belief as the *Ideomotor Effect*, arguing that muscular movement can be nonconsciously initiated by the mind. It is common to attribute the Ideomotor Effect to any unproven, puzzling phenomena, such as dowsing, Ouija boards, automatic writing, the motion of a pendulum, Facilitated Communication and muscle testing. However, since the practitioners were blind to the verity of the spoken statement, it is unlikely that practitioners could be unwittingly responsible for

an ideomotor action. Likewise, since there was no significant difference between the pairs whose patients reported guessing the paradigm, and those who did not, it is unlikely that patients caused an ideomotor response either. Furthermore, since ideomotor responses are said to be related to a *heightened belief*, and since no correlation was found between MRT accuracy and increase in any confidence rating, it is especially unlikely that MRT represents an Ideomotor Effect.

Result #6: Truths were easier to detect than lies.

There was one more interesting result that is directly useful to clinical practice, and that was that true statements were easier to detect than false. In other words, the average accuracy of all the true statements was consistently higher than that of all the false statements. The clinical implication of this finding is that practitioners should use more true statements than false statements when comparing spoken statements during MRT sessions.

Summary

This series of studies shows that MRT is consistently more accurate than either guessing or chance at distinguishing true from false statements. A strength of this series is that consistent results were achieved across multiple studies (see Figure 3 below). Other strengths of this series that contribute to their rigorousness include the use of a true “gold standard” as a reference standard, and a high degree of blinding. In addition, we used heterogeneous samples; that is, a broad range of practitioners with varying levels of experience, and patients with different backgrounds as well. Finally, these studies used simple yet robust methodologies, which would make their replication straightforward. It is my hope that potential researchers reading this will be encouraged and carry out additional MRT research themselves.

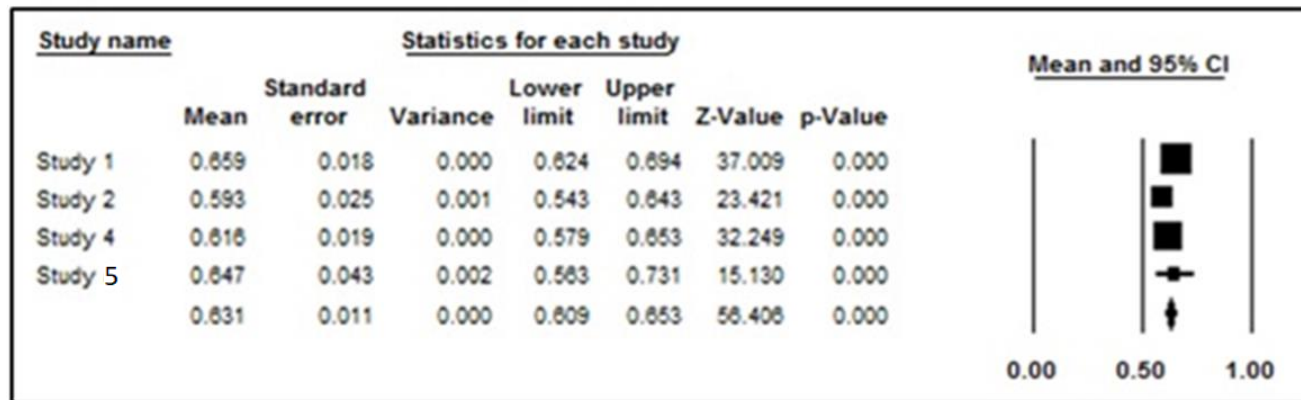
There are also limitations of these studies, including the results are not generalizable to other applications of MRT and to other types of manual muscle testing (MMT). This means that just because MRT has been shown to accurately detect lies, it does not mean that MRT can accurately detect other conditions, such as organ dysfunction, vertebral subluxation or the need for a specific nutritional supplement. In order to make these claims, further specific research is required. Another limitation is that these studies may have been under-powered for subgroup analysis because no factors that influenced accuracy were identified.

Directions of future research

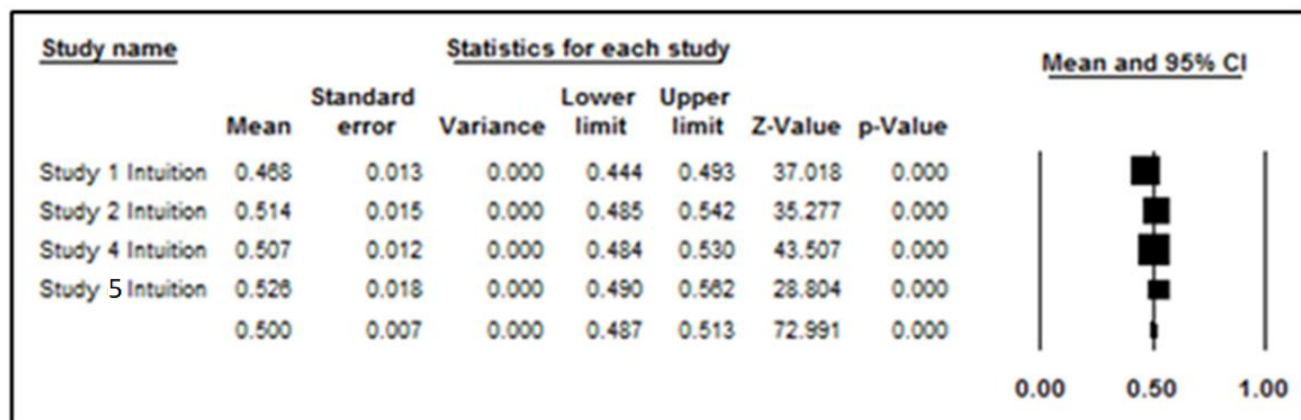
This series of studies offer encouraging first steps toward the validation of MRT, however further research is certainly required. For instance, it would be very useful to determine what factors influence MRT accuracy, and in order to achieve this, future research would require larger sample sizes. It would also be interesting to compare the results of MRT to detect lies with other lie detection tests (e.g. polygraph).

In addition, because these studies achieved MRT accuracies in the 60%-range, it is important to ascertain if this is “good enough” clinically. In order to accomplish this, MRT technique systems must be assessed for their effectiveness. This is accomplished through rigorous clinical trials (e.g. randomized

Figure 3 - Forrest Plots: (A) MRT Accuracy, and (B) Intuitive Guessing Accuracy



(A)



(B)

clinical trials, RCTs). This will *not* be accomplished through case studies – regardless of the number of case studies generated. In the world of evidence-based health care, case studies, while they may be interesting, they are considered poor evidence, similar to editorials and testimonials. They are largely ignored by those who make decisions about healthcare policies and funding. It is my suggestion that the effort and resources that are currently put in to generating case studies now be put toward running clinical trials. Only this will advance the evidence base.

Many practitioners who volunteered for these studies asked me *how* MRT worked. During my 10 years at Oxford I was only asked this question once – and that was at the very end, during my DPhil viva (PhD exam). My colleagues at Oxford were not interested in how or why an intervention or test works, but rather they were interested in learning *if* it works, how well it works, and if it causes harm. That's all. Yet, after 3 hours of questioning, one of my examiners finally asked me, "So how does muscle testing work anyway?" My response to him was, "That wasn't my research question." He was happy with that and we moved on. The truth is: We do not know how MRT works. However, if an Oxford examiner does not care, an expert in the field of clinical research, then hopefully you readers will not care as well. We also spend a lot of time, money and other resources in an attempting to figure this out, whereas I believe that our limited resources can be better spent elsewhere. Yes, this series of studies may be a start in the validation process of MRT, but we have a long way to go.

Dr Anne Jensen, DC, PGCert, PGDip, MSc, MS, DPhil