



## Testing alleged mediumship: Methods and results

Ciarán O’Keeffe<sup>1\*</sup> and Richard Wiseman<sup>2</sup>

<sup>1</sup>Psychology Department, Liverpool Hope University, UK

<sup>2</sup>Psychology Department, University of Hertfordshire, UK

Mediums claim to be able to communicate with the deceased. Such claims attract a considerable amount of public interest and, if valid, have important implications for many areas of psychology. For over 100 years, researchers have tested alleged mediums. This work has obtained mixed results and provoked a considerable amount of methodological debate. This paper reviews the key issues in this debate, describes how the authors devised a method of testing that aimed to prevent the many problems that have hindered past research, and how they then used this method to test several professional mediums. The results of this work did not support the existence of genuine mediumistic ability. Competing interpretations of these results are discussed, along with ways in which the methodology presented in the paper could be used to assess conceptually similar, but non-paranormal, claims made in clinical, occupational and forensic contexts.

Some individuals claim to possess mediumistic abilities that allow them to contact the ‘spirit world’ and receive information from the deceased. There are several reasons to subject these claims to rigorous and empirical investigations.

First, mediumistic abilities, if valid, would provide evidence to support the survival of bodily death, and thus have important implications for aspects of psychology. Such data would, for example, present a strong challenge to key assumptions underlying neuropsychological research, including the notion that human personality, cognition, and consciousness is dependent on a living brain. Evidence of genuine mediumistic abilities would also raise intriguing questions about the sensory mechanisms that might underlie such abilities and, on a more practical level, have important implications for the many aspects of clinical and counselling psychology concerned with bereavement and grief.

Second, demonstrations of apparent mediumistic abilities have a significant impact on public belief and behaviour. Recent opinion polls have revealed that almost 30% of Americans now believe in the existence of genuine mediumistic abilities (Newport & Strausberg, 2001), approximately 10% of Britons visit mediums to both receive messages

\*Correspondence should be addressed to Ciarán O’Keeffe, Psychology Department, Liverpool Hope University, Taggart Avenue, Hope Park, Liverpool, L16 9JD, UK (e-mail: [okeeffc@hope.ac.uk](mailto:okeeffc@hope.ac.uk)).

from the deceased and obtain general guidance for their lives (Roe, 1998), and new types of television programmes featuring such demonstrations consistently attract millions of viewers (Brown, 2001). Well-controlled tests of mediums would help the public and television programme makers assess the validity of such alleged abilities, and thus help inform their resulting decisions and behaviour.

Third, certain individuals working in non-paranormal contexts make claims that are analogous to those made by mediums, and the methods developed to test mediums could be used to examine these claims. For example, some clinicians claim to be able to gain insights into patients' backgrounds purely from their reactions to certain projective tests, some practitioners working in an occupational setting appear to be able to give detailed accounts of people's personality simply from their scores on certain assessment tools, and some individuals operating in a forensic context claim to be able to produce accurate profiles of offenders from a very limited amount of behavioural information. Several writers (e.g. Alison, Smith, & Morgan, 2003; Wood, Nezworski, Lilienfeld, & Garb, 2003) have recently noted that the anecdotal evidence supporting these claims may be the result of the same types of psychological stratagems that can underlie the apparent accuracy of mediumistic readings (i.e. the use of general statements, chance, etc.), and thus the methods developed to examine such claims may benefit from a thorough understanding of the procedures used to test mediumship.

Given the nature of the theoretical and practical issues surrounding this topic, it is perhaps not surprising that the scientific testing of mediumship has a long and controversial history.

Initial tests of mediums were carried out in the 1880s and primarily involved investigators attending sésances, noting down the comments that the mediums allegedly received from the deceased, and then attempting to assess the accuracy of this information. The majority of the resulting reports argued in favour of the existence of genuine mediumistic ability, and contained lengthy transcripts of mediumistic messages along with detailed descriptions of the evidence supporting these statements (see, e.g. Hodgson, 1892, 1898). Critics have attacked this work, arguing that it often failed to assess whether the seemingly accurate readings could have been the result of various psychological stratagems, such as the mediums engaging in shrewd guesswork or producing very general statements that would be endorsed by the majority of people (see, e.g. Gardner, 1992; Hyman, 1977; Podmore, 1901).

Over the years, several researchers have attempted to devise procedures that eliminate the potential for such stratagems, and then used these to examine some of the best-known mediums of the day. The resulting studies have obtained mixed results, with some work finding evidence in favour of genuine paranormal abilities and other research supporting the null hypothesis (for a review of this work see Schouten, 1994). This work has provoked a considerable amount of methodological and statistical debate, much of which has focused on the degree to which the procedures employed in those studies obtaining positive results have eliminated potential biases and problems (for a recent example of this type of debate, see Hyman, 2002; Hyman, 2003; Schwartz, 2003). Unfortunately, discussion surrounding the diverse range of potential methodological and statistical issues that can bias such work is spread across the psychological and parapsychological literature, and is often presented in a piecemeal way rather than being more conceptually organized. In addition, researchers working in this area have yet to develop a relatively standard method of testing that is both practical and minimizes the potential for such artifacts. This paper addresses both of these issues. The first part of this paper reviews the main problems that have hindered previous tests

of mediumship and describes how the authors devised a method of testing that was both practical and methodologically sound. The second part of the paper presents a detailed description of how this method was then used to test several professional mediums.

### **Methodological issues: Problems and procedures**

The debate concerning the potential problems that can arise during tests of alleged mediumistic ability has centered around three key issues: (a) the need to control for potential sensory leakage, (b) the need to accurately assess the generality of the mediums' statements, and (c) the need for 'blind' judging. The following three sections briefly review each of these problems and outline the types of procedures that can be employed to overcome them.

#### ***The need to control for potential sensory leakage***

Alleged mediums may be able to gain information about their clients (often referred to as 'sitters') via normal means and then use this information to help produce accurate readings. Such information may be obtained in a variety of ways. For example, books about how to fake mediumistic abilities describe various techniques for obtaining useful information in advance of a reading, including, for example, secretly eaves-dropping on sitters' conversations, or conducting surreptitious searches of telephone directories and the Internet (see, e.g. Rowland, 1998). Other writers have described how experienced mediums may be able to unconsciously gain information from more subtle sources, such as sitters' clothing, posture, demeanor, and jewellery (see, e.g. Morris, 1986). Even a very limited amount of contact between medium and sitter has the potential to provide useful information. For example, Wiseman and O'Keeffe (2001) noted that the speed with which the sitter answers 'yes' or 'no' to the medium's questions could unconsciously provide experienced mediums with useful feedback about the accuracy of their comments during a reading.

For these reasons, any well-controlled test of mediumistic ability should prevent mediums gaining information about sitters via normal means. This usually involves experimenters taking appropriate steps to ensure that mediums cannot ascertain any information about sitters in advance of test readings, and that there are sufficient safeguards preventing them from obtaining verbal or non-verbal cues during the readings. Such safeguards should also extend to anyone involved in the study (e.g. experimenters or other participants) who are aware of any information about the identity of the sitters. Researchers examining the possible existence of telepathy have developed various procedures for eliminating potential sensory leakage between participants (see e.g. Milton & Wiseman, 1997) and many of these safeguards (e.g. placing participants in separate rooms with sufficient levels of sound attenuation) can be employed to eliminate possible leakage during tests of alleged mediumistic abilities.

#### ***The need to accurately assess the generality of the mediums' statements***

Research into the so-called 'Barnum effect' has consistently shown that people tend to rate certain types of very general personality statements (e.g. 'You have a great deal of untapped creative potential') as highly accurate (Forer, 1949; Furnham & Schofield, 1987). In addition, more recent work has revealed that even statements that do not appear especially general can be true of many people. For example, Blackmore (1994)

carried out a large-scale survey in which over 6,000 people were asked to state whether quite specific statements were true of them. Over one third of people endorsed the statement ‘I have a scar on my left knee’ and over a quarter answered yes to the statement ‘Someone in my family is called Jack’. Mediums can utilize this phenomena to produce readings that may appear highly accurate but, in reality, simply contain very general statements that are endorsed by a large number of sitters.

Attempts to deal with this issue in tests of mediumistic ability have taken many forms over the years and have been the subject of considerable debate. In perhaps the earliest attempt to solve the problem, Hyslop (1919) collated statements that had been endorsed by a sitter during various test readings, and then asked a ‘control’ group containing approximately 500 people to indicate whether each statement was true of them. Hyslop then calculated the general acceptance level of the reading on the basis of the percentage of people in the control group that endorsed each statement. For example, if 250 people in the control group endorsed the statement ‘you are male’, then Hyslop calculated the probability of acceptance as 250/500 or 0.5. To obtain an overall probability of all of the statements being endorsed, Hyslop multiplied the individual probabilities for each of the statements together (e.g. the probability of two statements being endorsed, each having a general acceptance level of 0.5, would be 0.25). Several critics have correctly noted that this approach greatly inflates the medium’s apparent accuracy because it incorrectly assumes that each of the statements are independent of one another (Schouten, 1994). Thus, if, for example, the medium stated that the sitter ‘had recently lost someone who was male’ and that this person ‘had a beard’, the probability of these statements would be multiplied together as if they were independent, whereas the probability of the first being accurate is heavily related to the probability of the second being correct.

Over the years, researchers have devised various forms of analyses that attempt to overcome this problem (see e.g. Pratt, 1936; Saltmarsh & Soal, 1930). Probably the most widely endorsed and employed is that developed by Pratt and Birge (1948). In the Pratt and Birge procedure, a small number of sitters each receive a reading from a medium. The sitters are then asked to rate the accuracy of statements from both their own reading (often referred to as the ‘target’ reading) and those from the readings of other sitters (referred to as ‘decoy’ readings). If the medium is accurate, then the ratings assigned to the target readings will be significantly greater than those assigned to the decoy readings. If, however, the medium is simply producing general statements, then the sitters will assign similar ratings to both the target and decoy readings. Pratt and Birge noted that the results of experiments using this procedure can perhaps best be viewed as shown in Table 1, with the numbers on the diagonal of

**Table 1.** Standard way of representing data from experiments employing the Pratt and Birge technique

Sitter present during reading	Sitter judging accuracy of reading				
	John	Eric	Bill	Tony	Tom
John	<b>58</b>	23	46	6	56
Eric	25	<b>73</b>	14	45	53
Bill	18	41	<b>67</b>	33	39
Tony	61	22	40	<b>49</b>	30
Tom	11	39	26	28	<b>72</b>

the table (shown in bold) representing the scores that each sitter gave to their own readings, and the numbers on the off-diagonal numbers representing the ratings that sitters assigned to the readings of others (Pratt, 1969).

It is widely recognized that the statistical analyses used to test whether the numbers on the diagonal are significantly greater than those on the off-diagonals do not assume that the statements within the readings are independent (Pratt, 1969). To this end, researchers have recommended creating a distribution of the sum of the numbers on the diagonal for each possible permutation of the matrix, and then calculating the probability of the experimental outcome by examining where the sum of the numbers on the diagonal actually obtained in the experiment lies within this distribution (for further discussion about such analyses see Greville, 1949; Pratt & Birge, 1948; Pratt, 1969; Scott, 1972; Thouless, 1949).

### ***The need for 'blind' judging***

The way in which sitters rate the accuracy of mediumistic readings is highly subjective (Hyman, 1977). For example, Wiseman and O'Keeffe (2001) note that the statement 'The spirits are talking about the younger woman who has now passed away', is open to several interpretations (e.g. the word 'younger' could refer a young child, a teenager, or even someone who died in their forties), and that the degree to which a sitter is prepared to think through these alternative interpretations will influence the perceived accuracy of the statement. The process of assessment can also be biased by selective recall. For example, the medium saying 'Your daughter was an extrovert' may cause sitters to selectively recall certain life events (i.e. the times that his or her daughter went to parties), forget other events (e.g. the times that she wanted to be alone), and thus assign a spuriously high accuracy rating to the statement. The degree to which the sitter thinks about alternative interpretations of ambiguous statements and engages in selective recall may be influenced by several factors, including, for example, their need to believe in the afterlife or please the medium.

Researchers testing alleged mediumistic ability have attempted to eliminate such biases by having sitters rate the accuracy of statements without informing them whether the statements are drawn from target or decoy readings (Pratt, 1969). However, such procedures may not fully eliminate some of the more subtle temporal cues that might help sitters distinguish target from decoy readings. Imagine, for example, that the sittings are scheduled for different days and that in one reading the medium refers to a memorable news story (e.g. 'The spirits are upset by that horrible train crash today'). When the sitters are subsequently presented with the readings for assessment, they may see this comment and correctly deduce the day on which the reading took place, and thus know whether this is their target reading. Similar problems may arise even if the sitters are scheduled at different times on the same day, if the medium's comments allow a sitter to figure out when a reading was made (e.g. during a lunchtime sitting the mediums remarks 'The spirits always get hungry around now'), or both the sitter and medium experience an idiosyncratic event during a reading (e.g. a crash of lightning outside) and the medium makes reference to this event (e.g. 'The lightning is making it difficult to contact the spirits').

To our knowledge, previous tests of alleged mediumistic ability have failed to recognize, and therefore control for, this potential artefact (although see Milton & Wiseman, 1997, for a discussion regarding how the same type of temporal cues could bias the outcome of certain types of extra-sensory perception experiments). Various

procedures could be employed to minimize the problem. For example, the sitters can be scheduled on the same days, and the time of their readings can be counterbalanced across the days (i.e. each sitter has one session scheduled at 11:30 a.m., one at 12:30 a.m., etc.). Also, the sitter and/or the medium can be located in rooms that isolate them from obvious sources of idiosyncratic external events, such as unusual weather conditions or noise from surrounding rooms and corridors. Finally, the statements that make up the reading can be separated and randomly ordered before being presented to sitters for assessment, thus minimizing the possibility of a subtle cue in one statement influencing the way in which sitters evaluate an entire reading.

## **An experimental test of mediumship**

The previous section outlined the main methodological and statistical problems that can hinder tests of mediumship, and some of the procedures that can be employed to eliminate these potential problems. Some of these procedures have been used during previous tests of mediumship, whilst others (e.g. those concerned with ways of eliminating potential temporal cues that might help sitters distinguish target from decoy readings) have not. The authors recently devised a method for testing mediumship that incorporated all of the procedures described above, and then used this method to test several professional mediums. This section outlines the methods and results of that test.

The test involved five professional mediums giving readings for five sitters under conditions that eliminated any potential sensory leakage between medium and sitter. The sitters were then asked to assess the accuracy of the mediums' statements without knowing whether the statements were drawn from target or decoy readings. Extortive permutation analyses were then used to assess whether the ratings assigned to target readings were significantly higher than the ratings assigned to decoy ratings.

### **Participants**

#### *Mediums*

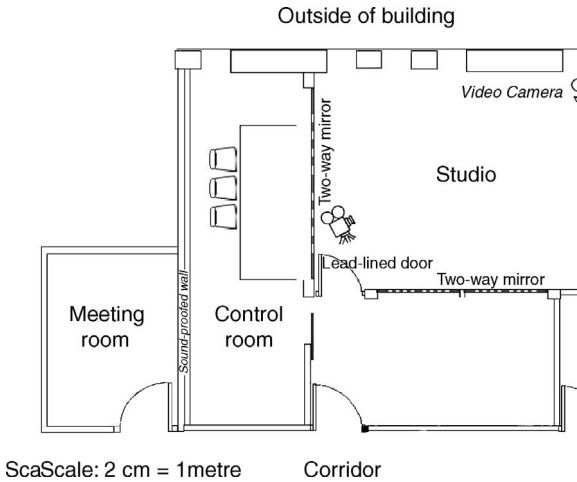
The five mediums (3 female, 2 male; age range 42-55) were recruited via a list of certified mediums provided by the Spiritualists Nationalist Union (SNU). The SNU stated that all of the mediums on this list had undergone a rigorous selection procedure and were subject to continual assessment. Each medium was initially contacted by telephone, and then sent a detailed description of the protocol and consent form.

#### *Sitters*

The five sitters (all male, age range 25-30) were either students or staff from the university. They were selected from a pool of individuals who responded to a general e-mail, circulated within the university, asking for volunteers to be involved in a scientific test of mediumship. The sitters were chosen using the following criteria; (a) they did not know one another, (b) they were the same gender, and (c) they were approximately the same age. Each sitter was initially contacted by telephone, and then sent a detailed description of the protocol and consent form. None of the sitters were paid for their involvement in the study.

### **Rooms and apparatus**

The experiment took place in a suite of rooms located within the university's Psychology Department (see Fig. 1). The medium was located in the studio area and the sitter was



**Figure 1.** Floor-plan of rooms.

placed in the meeting room. These rooms were acoustically isolated from one another, such that the sitter could not hear the medium and vice versa. Events happening outside the building (weather effects, etc.) could not be heard in the meeting room, and noise from the corridor directly outside the meeting room could not be heard in the studio. The mediums' comments were recorded via video-cameras operated remotely by E1 from the control room. The sitter was supplied with a portable stereo system and headphones so that they could listen to music throughout the session.

### Procedure

The experiment was run by two experimenters, E1 and E2. E1 initially selected and contacted the five mediums, whilst E2 selected and contacted the five sitters. E1 then liaised with the mediums to arrange a day on which each of them could visit the university to participate in the study (referred to as 'experimental days'). E1 then passed these five dates to E2, who arranged for each sitter to visit the Psychology Department at a specified time on each of these days. E2 scheduled the five sitters in hourly slots from 11:00 a.m. to 4:30 p.m. The order of the sitters was counterbalanced across the 5 days (see Table 2).

At the start of each experimental day, E1 met the medium and took him or her to the studio. E1 outlined the nature and design of the study, and ensured that the medium was

**Table 2.** Counterbalanced order of sitters A–E on each experimental day

	Mediums				
	M1	M2	M3	M4	M5
11:00–12:00	A	E	D	C	B
12:00–13:00	B	A	E	D	C
13:30–14:30	C	B	A	E	D
14:30–15:30	D	C	B	A	E
15:30–16:30	E	D	C	B	A

comfortable. At 11:00 a.m., E1 informed the medium that the first session was just about to start and then left the studio for the control room. The medium was asked to attempt to contact one or more spirit(s) associated with the sitter and relate any information that they believed appropriate. The medium was allowed to say as much or little as he or she wished during the next 60 minutes. All of these comments were recorded on videotape and monitored by E1 from the control room. At approximately 12:00 midday, E1 re-entered the studio and informed the medium that the session had ended, and that the next session would commence after a short break. This procedure was repeated five times throughout the day.

E2 met the first sitter at 10:45 a.m. and took them to the meeting room. They were given a choice of music and asked to listen to this music, via the headphones provided, between 11:00 a.m. and midday. At approximately 11:00 a.m., E2 left the meeting room and remained outside until the end of the session. At approximately midday, E2 re-entered the meeting room, terminated the session and thanked the sitter for participating. This procedure was repeated five times throughout the day.

In the event of any unexpected events (e.g. a session time having to change slightly due to the late arrival of a sitter), the two experimenters communicated with one another via a series of simple coded text messages.

Throughout the study, both E1 and the mediums did not receive any information about the identity of the sitters, E2 did not have any contact with the mediums, and all of the mediums and sitters remained blind to each others' identities.

### **Transcription and rating of readings**

After the 5 experimental days had been completed, E1 transcribed the mediums' comments, removed any extraneous details from the transcripts (e.g. questions to the experimenter, pauses) and broke each reading down into a series of statements (see Appendix A for an example). The statements from all of the mediums were then sent to the sitters, who were asked to rate the accuracy of each statement between 1 (*not applicable*) and 7 (*very applicable*). Each sitter independently rated all of the statements, and an overall score for each reading was created by summing the individual ratings assigned to each of the statements making up that reading.

## **Results**

Table 3 contains the number of statements produced by the mediums during each of the readings. Although the length of the readings produced by any one medium is reasonably consistent, there is considerable variation between mediums. For example, Mediums 2 and 5 tended to produce long readings that contained around seventy statements, whilst Mediums 3 and 4's readings were much shorter.

**Table 3.** Total number of statements in each reading

	Medium 1	Medium 2	Medium 3	Medium 4	Medium 5
Sitter A	55	92	6	24	80
Sitter B	62	56	9	19	67
Sitter C	60	78	11	26	52
Sitter D	64	82	8	20	61
Sitter E	58	69	6	28	76
<b>Mean</b>	<b>59.8</b>	<b>75.4</b>	<b>8.00</b>	<b>23.4</b>	<b>67.2</b>



Table 4 contains the means of the sitters' ratings collapsed across the five readings produced by each medium. These ratings display large but consistent variability with, for example, almost all of the readings produced by Medium 5 obtaining relatively high ratings, whilst those from Medium 3 consistently being assigned far lower ratings.

**Table 4.** Means of the sitters' ratings collapsed across the 5 readings produced by each medium

	Medium 1	Medium 2	Medium 3	Medium 4	Medium 5	Mean
Sitter A	3.33	3.72	1.52	3.67	5.24	<b>3.50</b>
Sitter B	2.88	4.15	2.25	2.58	4.66	<b>3.30</b>
Sitter C	2.77	3.42	1.55	2.92	5.15	<b>3.16</b>
Sitter D	2.67	3.75	2.27	3.11	4.10	<b>3.18</b>
Sitter E	3.43	3.15	1.30	4.46	3.15	<b>3.10</b>
<b>Mean</b>	<b>3.01</b>	<b>3.64</b>	<b>1.78</b>	<b>3.35</b>	<b>4.46</b>	<b>3.25</b>

As noted at the start of this paper, many researchers have argued that the degree to which sitters accept a reading is determined, at least in part, by various psychological stratagems (e.g. the number and diversity of topics mentioned in the reading, the generality of statements). The data obtained in this study support this notion. For example, Reading A from Medium 5 received the highest mean rating, and contained a relatively large number of general statements relating to a diverse range of topics. During the eighty statements that make up this reading, the medium allegedly contacted the following six spirits: a woman who had had five children, a grandfather, a large dog, Henry VIII, a man from East Africa, and a smaller dog. Some of the statements about these spirits were extremely general. For example, the woman was described as working in a shop that 'had something to do with pots and pans', the grandfather was seen as tall and balding, the dog was medium-sized and of a light colour with spots, and the man from East Africa was 'involved in the construction business'. Perhaps not surprisingly, many of these very general statements received high ratings. For example, the two statements that received maximum ratings from all five sitters were 'Yes, a relative. Is it a man? Will he let me shake his hand? I can feel your hand in my hand' and 'What can we call you? Are you father? Grandfather? You're a grandfather.'. In marked contrast, Reading E from Medium 3 obtained the lowest ratings of any reading, and consisted of just six statements, all of which referred to just one alleged spirit – a man from Lancaster. Most of these statements were highly specific, describing, for example, how the man had long black hair, a pain in his left foot, a big right toe, and was 167 cm tall. It seems very likely that the low ratings assigned to this reading were due to the relatively unambiguous and unusual nature of these statements, as reflected in the fact that the statement receiving the lowest possible rating from all sitters was also extremely specific, namely that the spirit was somehow connected to the name 'Bilger'.

To explore the accuracy of each medium, the sum of the sitters' ratings for each reading produced by that medium were entered into a Pratt and Birge table (see Appendix B), and the significance level calculated via a permutation analysis based on the 120 possible arrangements of each matrix. This analysis involves summing the ratings that the sitters assigned to the target readings (i.e. the readings made by the medium when they were the sitter), creating a distribution from the sums of ratings that could have been obtained had different combinations of decoy readings

been the target reading for each sitter, and then calculating a  $p$  value by examining where the sum of the actual target ratings sits within this distribution.

None of the analyses were significant, and the resulting  $p$  values (1 tailed) were as follows: Medium 1, .89; Medium 2, .27; Medium 3, .27; Medium 4, .77; Medium 5, .66; all mediums combined, .63. Inspection of the data revealed that there was only one occasion (Medium 2 reading for Sitter B) when the sitter for whom a reading was intended assigned a higher rating to the reading than the other four sitters. On all other occasions, the ratings assigned by sitters who were not present at the time of the reading were higher than the rating assigned by the sitter for whom the reading was intended.

## Discussion

This paper first briefly outlined the major methodological and statistical problems that have hindered previous test of alleged mediumistic abilities, and described procedures that can be used to minimize these problems. It then described the way in which these procedures were implemented during the authors' recent test of five professional mediums. This test involved five sitters each receiving five readings and then rating the accuracy of those readings. The results revealed that the ratings that sitters assigned to their own readings were not significantly different from the ratings they assigned to others sitters' readings, and thus did not support the existence of mediumistic ability.

These findings can be interpreted in various ways. It is possible that genuine mediumistic ability does not exist, and that the apparent accuracy of mediums' readings are entirely due to the type of psychological stratagems outlined in the first section of this paper. This interpretation is consistent with much of the sceptical literature on alleged mediumship (see e.g. Gardner, 1992; Hyman, 1977), previous work that has also failed to find evidence of such abilities under controlled conditions (see Schouten, 1994, for a review of this work) and those arguing that the studies that have obtained positive results are methodologically flawed (e.g. Hyman, 2002, 2003). This interpretation is also consistent with the fact that the presence of some of these strategies (e.g. length and diversity of reading, generality of statements) appeared to be associated with the ratings assigned by sitters to the readings obtained in this study. If this interpretation is correct, then the most productive direction for future work in this area is to examine these stratagems more closely, examining, for example, the types of people that tend to endorse mediumistic readings and the forms of rhetoric that alleged mediums use to convince sitters that they are receiving messages from their deceased friends and relatives (see e.g. Wooffit, 1992, 2001). Alternatively, it is possible that genuine mediumistic abilities do exist, but that this study failed to find evidence of them because, for example, the mediums involved in the experiment do not possess such abilities or the setting in which the study was conducted did not elicit such abilities. These hypotheses can only be evaluated by systematically varying these factors in future work, providing that such work also eliminates the various methodological problems discussed in this paper.

On a methodological level, the study eliminated the various types of bias that can hinder research in this area. Whilst some of these procedures have been used in several previous tests of mediumship (e.g. safeguards against sensory leakage between mediums and sitters, and use of the Pratt-Birge technique), others have not been utilized in this context before (e.g. safeguards against potential temporal cues). The resulting methodology was both practical and straightforward, and it is hoped that

other researchers will employ this method to investigate other individuals claiming similar types of paranormal abilities, and that this work will help tease apart the competing interpretations outlined above.

Finally, as noted in the Introduction, certain individuals working in clinical, occupational, and forensic contexts make claims that are analogous to those made by mediums (i.e. being able to ascertain highly accurate information about a person or situation on the basis of very limited data), and thus tests of such claims could benefit from many of the methodological procedures described in this paper. For example, the efficacy of psychoanalytic projective tests could be evaluated by asking clinicians to produce descriptions of five people solely on the basis of their responses on a projective test, having all five people rate the accuracy of these descriptions without knowing which pertained to them, and then using the analysis described in this paper to evaluate the accuracy of the clinicians involved. The same type of test could also be used to evaluate individuals claiming to be able to produce accurate descriptions of criminals (e.g. likelihood of reoffending, whether they are socially isolated, likely interests and hobbies, etc.) based on relatively limited information about their behaviour whilst committing a crime. This could involve presenting profilers with information about five crimes that have already been solved, and asking them to produce a profile of the criminals that they believe committed these crimes. Experienced police officers would then be shown these profiles, along with information about the criminals who committed the crimes, and be asked to rate the degree of correspondence between the profile and each of the actual criminals. Once again, the analyses described in this paper could then be used to assess the accuracy of the profilers involved in the study.

Up to this point in time, the literature discussing the potential problems that can hinder research attempting to assess mediumistic claims, and the possible procedures that can be employed to overcome them, has been widely distributed across a range of highly specialist publications within parapsychology. It is hoped that the conceptual grouping of these problems and procedures presented in this paper, along with an example of how they were combined into a practical and methodologically sound method that was used to assess several professional mediums, will help bring this work to psychologists working in a broad range of applied contexts.

In short, the present study found no evidence to support the notion that the professional mediums involved in the research were, under controlled conditions, able to demonstrate paranormal or mediumistic ability. However, the authors have developed a practical, straightforward, and methodologically sound way of testing such claims, and it is hoped that this approach will be employed by researchers to test other individuals who appear to have mediumistic or psychic abilities, and conceptually similar claims being made within clinical, occupational, and forensic contexts.

## Acknowledgements

The authors would like to thank Dr Caroline Watt for her feedback on an earlier version of this paper, the mediums and sitters who kindly participated in the study, and the Perrott-Warrick Fund for helping to fund the research.

## References

- Alison, L. J., Smith, M. D., & Morgan, K. (2003). Interpreting the accuracy of offender profiles. *Psychology, Crime and Law*, *9*, 185–195.
- Blackmore, S. (1994). Probability misjudgement and belief in the paranormal: Is the theory all wrong? In D. Bierman (Ed.), *Proceedings of the 37th Annual Convention of the Parapsychological Association* (pp. 72–82).
- Brown, I (2001, March 5). Hearing from dearly departed proves a hit on Sci-Fi Channel. *Los Angeles Times*.
- Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology*, *44*, 118–123.
- Furnham, A., & Schofield, S. (1987). Accepting personality test feedback: A review of the Barnum effect. *Current Psychological Research and Reviews*, *6*, 162–178.
- Gardner, M. (1992). *On the wild side*. Buffalo, NY: Prometheus Books.
- Greville, T. N. E. (1949). On the number of sets required for testing the significance of verbal material. *Journal of Parapsychology*, *13*, 137–138.
- Hodgson, R. (1892). A record of certain phenomena of trance. *Proceedings of the Society for Psychical Research*, *8*, 1–167.
- Hodgson, R. (1898). A further record of observations of certain phenomena of trance. *Proceedings of the Society for Psychical Research*, *13*, 284–582.
- Hyman, R. (1977). Cold reading: How to convince strangers that you know all about them. *Skeptical Inquirer*, *1*, 18–37.
- Hyman, R. (2002). How not to test mediums: Critiquing the afterlife experiments. *Skeptical Inquirer*, *26*(1).
- Hyman, R. (2003). Hyman's reply to Schwartz's 'how not to review mediumship research'. *Skeptical Inquirer*, *27*(3), 61–64.
- Hyslop, J. H. (1919). Chance coincidence and guessing in a mediumistic experiment. *Proceedings of the American Society for Psychical Research*, *13*, 5–88.
- Milton, J., & Wiseman, R. (1997). *Guidelines for extrasensory perception research*. Hatfield, England: University of Hertfordshire Press.
- Morris, R. (1986). What psi is not: The necessity for experiments. In H. T. Edge, R. L. Morris, J. Palmer & J. H. Rush (Eds.), *Foundations of parapsychology*. London: Routledge & Kegan Paul.
- Newport, F., & Strausberg, M (2001, June 8). *Americans' belief in psychic and paranormal phenomena is up over last decade*. Report by The Gallup Organisation.
- Podmore, F. (1901). On Professor Hyslop's report on his sittings with Mrs. Piper. *Proceedings of Society for Psychical Research*, *17*, 374–388.
- Pratt, J. G. (1936). Towards a method of evaluating mediumistic material. *Bulletin of the Boston Society for Psychic Research*, *23*, 5–53.
- Pratt, J. G (1969). *On the evaluation of verbal material in parapsychology* (Parapsychology Monograph No 10). New York: Parapsychology Foundation.
- Pratt, J. G., & Birge, W. R. (1948). Appraising verbal test material in parapsychology. *Journal of Parapsychology*, *12*, 236–256.
- Roe, C. A. (1998). Belief in the paranormal and attendance at psychic readings. *Journal of the American Society for Psychical Research*, *92*(1), 25–51.
- Rowland, I. (1998). *The full facts book of cold reading*. London, England: Ian Roland.
- Saltmarsh, H. F., & Soal, S. G. (1930). A method of estimating the supernormal content of mediumistic communications. *Proceedings of the Society for Psychical Research*, *39*, 266–271.
- Schouten, S. A. (1994). An overview of quantitatively evaluated studies with mediums and psychics. *Journal of the American Society for Psychical Research*, *88*, 221–254.
- Schwartz, G. E. (2003). How not to review mediumship research. *Skeptical Inquirer*, *27*(3), 58–61.

- Schwartz, G. E. R., Russek, L. G. S., Nelson, L. A., & Barentsen, C. (2001). Accuracy and replicability of anomalous after-death communication across highly skilled mediums. *Journal of the Society for Psychical Research*, 65(862), 1–25.
- Scott, C. (1972). On the evaluation of verbal material in parapsychology: A discussion of Dr Pratt's monograph. *Journal of the Society for Psychical Research*, 46(752), 79–90.
- Thouless, R. H. (1949). Review of Journal of Parapsychology, Vol. 12, Number 4. *Journal of the Society for Psychical Research*, 35, 48–49.
- Wiseman, R., & O'Keeffe, C. (2001). Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A critique. *Paranormal Review*, 19, 3–6.
- Wooffit, R. (1992). *Telling tales of the unexpected: The organisation of factual discourse*. London: Harvester Wheatsheaf.
- Wooffit, R. (2001). Raising the dead: Reported speech in medium-sitter interaction. *Discourse Studies*, 3, 351–374.
- Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). The Rorschach Inkblot test, fortune tellers, and cold reading. *Skeptical Inquirer*, 27(4), 29–33.

Received 28 March 2004; revised version received 19 August 2004

## Appendix A: Example of a reading being converted into corresponding statements

### Reading

I think there is a lady in the room. Who are you? Mother? Yes, mother. About 5 foot 4'. I can see a pot, a cooking pot, a brass cooking pot. That's a rather large pot, isn't it? It's got a lid. You [*spirit*] worked in cooking, dinner cooking. You [*spirit*] worked in a shop selling pots and pans. You had something to do with a shop, pots and pans. Did you have a favourite piece in your shop? Ah, yes, I can see it now. Is it a long pan for cooking fish or something? That's very nice indeed. It looks like a fish cooker to me. You're English aren't you? Yes (*refers to Mother*).

### Statements

- S1: I think there is a lady in the room. Who are you? Mother? Yes, Mother. About 5 foot 4'.
- S2: I can see a pot, a cooking pot, a brass cooking pot. That's a rather large pot, isn't it? It's got a lid.
- S3: You [*spirit*] worked in cooking, dinner cooking.
- S4: You [*spirit*] worked in a shop selling pots and pans. You had something to do with a shop, pots and pans.
- S5: Did you have a favourite piece in your shop? Ah, yes, I can see it now. Is it a long pan for cooking fish or something? That's very nice indeed. It looks like a fish cooker to me.
- S6: You're English aren't you? Yes (*refers to Mother*).

**Appendix B: Results tables for individual mediums and all mediums combined**

Sitter present during reading	Sitter judging accuracy of reading				
	A	B	C	D	E
A	<b>156</b>	230	131	176	252
B	202	<b>183</b>	223	192	301
C	170	111	<b>145</b>	106	161
D	348	240	196	<b>166</b>	175
E	120	97	134	159	<b>137</b>

**Results for Medium 1**

Sitter present during reading	Sitter judging accuracy of reading				
	A	B	C	D	E
A	<b>354</b>	410	304	294	322
B	198	<b>277</b>	210	243	230
C	234	289	<b>282</b>	318	264
D	452	341	220	<b>269</b>	113
E	166	250	275	291	<b>258</b>

**Results for Medium 2**

Sitter present during reading	Sitter judging accuracy of reading				
	A	B	C	D	E
A	<b>15</b>	21	13	7	12
B	11	<b>17</b>	9	24	12
C	13	13	<b>12</b>	15	7
D	8	26	18	<b>24</b>	11
E	14	13	10	21	<b>10</b>

**Results for Medium 3**

Sitter present during reading	Sitter judging accuracy of reading				
	A	B	C	D	E
A	<b>72</b>	42	98	43	109
B	51	<b>59</b>	49	67	89
C	88	61	<b>77</b>	110	146
D	97	84	54	<b>61</b>	91
E	122	57	64	84	<b>87</b>

**Results for Medium 4**

Sitter present during reading	Sitter judging accuracy of reading				
	A	B	C	D	E
A	<b>257</b>	318	410	194	181
B	410	<b>364</b>	312	286	243
C	355	289	<b>297</b>	321	176
D	312	320	354	<b>331</b>	238
E	427	276	358	246	<b>220</b>

## Results for Medium 5

Sitter present during reading	Sitter judging accuracy of reading				
	A	B	C	D	E
A	<b>854</b>	1,021	956	714	876
B	872	<b>900</b>	803	812	875
C	860	763	<b>813</b>	870	754
D	1,217	1,011	842	<b>851</b>	628
E	849	693	841	801	<b>712</b>

Results for all mediums combined.