**Psychology of Consciousness: Theory, Research, and Practice**

Anomalous Experiences and Paranormal Attributions: Psychometric Challenges in Studying Their Measurement and Relationship

Rense Lange, Robert M. Ross, Neil Dagnall, Harvey J. Irwin, James Houran, and Kenneth Drinkwater

Online First Publication, May 16, 2019. http://dx.doi.org/10.1037/cns0000187

## Abstract

Research on the psychology of paranormal, religious and delusional belief has been stifled by a lack of careful distinction between anomalous experiences and corresponding attributions. The Survey of Anomalous Experience (SAE: Irwin, Dagnall & Drinkwater, 2013) addresses this nuance by measuring Proneness to Anomalous Experience (PAE) and Proneness to Paranormal Attribution (PPA). Re-analyzing existing data (351 men, 1,026 women) from previously published studies, we examined the SAE's internal validity via Rasch scaling and differential item functioning analyses. PPA showed good Rasch model fit and no item-bias, but it lacked adequate reliability. Several PAE items showed misfit to the Rasch model or gender-bias, though deleting five items produced a scale with acceptable reliability. Finally, we failed to validate a three-category rating scale version with the goal of improving the SAE's psychometric properties. All three formulations revealed a secondary factor related to the items' extremity rather than contents, suggesting that future research should consider the *intensity* of respondents' anomalous experiences and paranormal attributions.

**Keywords:** Anomalous experience; attribution theory; paranormal belief; psychometrics; Rasch scaling

# Anomalous Experiences and Paranormal Attributions:

# Psychometric Challenges in Studying their Measurement and Relationship

## INTRODUCTION

Anomalistic psychology posits that anomalous experiences have orthodox explanations resulting from psychological or physical factors that suggest parapsychological activity to some people under certain conditions (Jones & Zusne, 1981; French & Stone, 2013). That is, paranormal beliefs and experiences in the general population are regarded as cognitive aberrations within established frameworks of problem-solving or decision-making. The moniker *anomalous* is therefore synonymous with *unexplained* and references both the *aberrant salience* (Irwin, 2014; Irwin, Schofield, & Baker, 2014) and *ambiguity* surrounding purported parapsychological events and their susceptibility to a myriad of possible interpretations (Houran, Lynn & Lange, 2017; Irwin, Dagnall & Drinkwater, 2012). This view implies two, potentially dissociable components: (i) the occurrence of anomalous experiences, and (ii) the parapsychological interpretation of these experiences. Inventories for assessing subjective parapsychological or anomalous experiences and delusion-like ideations [e.g., the Anomalous Experiences Inventory (AEI): Gallagher, Kumar & Pekala, 1994 and the Peters et al. Delusions Inventory (PDI: Peters et al., 1999, 2004)] typically do not distinguish between these two components (David, 2010; Irwin, Dagnall & Drinkwater, 2013) and instead conflate them. Likewise, an important limitation of much research in the cognitive science of religion is that measures rarely tease apart general religious beliefs (e.g., “God exists”), personal religious beliefs (e.g., “God appeared to me last night”) and experiences supporting these beliefs (van Leeuwen & van Elk, 2018).

To address these conceptual and measurement limitations, Irwin et al. (2013) introduced the Survey of Anomalous Experiences (SAE) – which comprises two subscales that distinguish *Proneness to Anomalous Experiences* (PAE) from *Proneness to Paranormal Attribution*s (PPA) in the study of subjective parapsychological experiences. Irwin et al (2013) found that PAE and PPA were positively associated; however, in this study and other research this relationship was found to be modest (*rho*'s = .29 to .40, $p < .001$; cf. Irwin, 2015, 2017, 2018; Irwin & Wilson, 2013), thereby suggesting that the subscales may be distinct, discernible factors. Moreover, PAE and PPA correlated positively with measures of schizotypy, emotion-based reasoning and suspension of reality testing. Irwin et al. (2013) found the only statistically significant difference between the two subscales concerned associations with executive dysfunction: PAE correlated significantly with executive dysfunction, but PPA did not. Overall these findings were interpreted as offering conceptual and psychometric support for the SAE. However, the authors noted that further work examining relationships between SAE subscales and executive dysfunction was required. Subsequent research has shown that both PAE and PPA correlate with an intuitive thinking style (Irwin & Wilson, 2013), whereas stress sensitivity and minimal-self dysfunction correlated with PAE but not PPA (Irwin, 2018).

Ross, Hartig, and McKay (2017) used the SAE to explore the role of reasoning biases in the formation of paranormal explanations of anomalous experiences. This study built on research highlighting cognitive deficits associated with paranormal beliefs (for a review, see Irwin, 2009), i.e., the adoption of implausible explanations for anomalous experiences due to improper or biased consideration of evidence. In addition, Ross et al. (2017) suggested that this result may have implications for cognitive theories of delusions. For instance, Lange and Houran's model of subjective paranormal belief and experience (Houran & Lange, 2004; Lange & Houran, 1998, 1999, 2000) specifically draws on attribution theory in the clinical literature (e.g., Jaspers, 1923/1963; Maher, 1988, 1992; Kihlstrom & Hoyt, 1988), which explains delusions as a byproduct of an individual's failure to find a standard explanation for ambiguous events or anomalous experiences.

Other research challenges the adequacy of a "one-factor" model of delusions and recommends a "two-factor" model (Coltheart, Langdon, & McKay, 2011; Davies, Coltheart, Langdon, & Breen, 2001). According to two-factor models, a second factor – an impairment of the belief evaluation system – is required to explain the process by which unusual (ambiguous or anomalous) experiences lead to delusional beliefs. Consistent with two-factor models, Ross et al. (2017) found that individuals low in "analytic cognitive style" (i.e., the willingness or disposition to critically evaluate outputs from intuitive processing and engage in effortful analytic processing) are more likely to invoke paranormal or esoteric labels for anomalous experiences. Thus, empirically speaking, this study lends further credence to Irwin et al.'s (2013) SAE instrument.



*The Present Study*

Although the available literature seems to recommend the SAE as an improved conceptual and empirical measure for researching self-reported parapsychological experiences, its psychometric qualities remain open to examination. In particular, it is well-established (Bond & Fox, 2001; Wright & Stone, 1979) that there can be serious measurement problems with instruments developed using Classical Test Theory (CTT), and unfortunately, the SAE and most measures used in the paranormal belief and anomalous experience literature have been developed using these methods (see e.g., Drinkwater, Denovan, Dagnall & Parker, 2017; Gallagher et al., 1994; Jinks, 2012; Mehmet & Yesilyurt, 2014; Reiner & Wilson, 2015; Schofield, Baker, Staples & Sheffield, 2018; Storm, Drinkwater & Jinks, 2017; Tobacyk, 2004). Specifically, CTT has at least four main limitations:

- The usual approach within CTT is to develop a test consisting of a number of items, and to assume that the sum of the scores received on the test items defines the latent trait. Consequently, summed scores do not provide linear (i.e., interval-level) measures of the underlying trait. Raw scores are ordinal measures at best and, as a consequence, it is possible that group differences and treatment effects are distorted;

- Traditional scaling approaches essentially treat all questionnaire items as equivalent, thereby ignoring how the difficulty of the stimuli interacts with respondents' trait levels to produce the outcome of the test. As a result, it is difficult to select those questions that are most appropriate for a specific population of respondents;

- The standard raw score approach does not recognize that some items may be biased such that subjects with *identical trait levels* receive systematically *different* scores. This might be the case for instance when women (or younger respondents) endorse some questions more (less) often then do men (or older respondents) with *equal* trait levels. To the extent that subject groups show such differential response tendencies, the questions involved are said to be biased;

- Since traditional scaling approaches do not model respondents' expected answers to each of the items, aberrant response records cannot be identified. For instance, it cannot be determined whether low scores are due to low trait levels rather than malingering, item biases, or respondents' misunderstanding or incomplete processing of the questions. In other words, traditional scaling approaches offer no indicators of the internal validity of respondents' scores.

In a psychometric study of paranormal belief that avoided these limitations, Lange, Irwin and Houran (2000) introduced a series of statistical analyses grounded in Modern Test Theory (MTT) that they described as a "top-down purification" process. This method is used increasingly in consciousness studies (see e.g., Irwin & Marks, 2013; Preti, Vellante, & Petretto, 2017), and it combines Rasch (1960/1980) scaling with the removal of age- or gender-related response biases at the test and item levels. Controlling for such biases is critical, because statistical theory (Stout, 1987) and computer simulations (Lange, Irwin, & Houran, 2000) alike demonstrate that response biases can lead to spurious factor structures of constructs, significant distortions in scores, and consequently erroneous reliability and validity findings. For an overview of the advantages of MTT, and specifically Rasch scaling, as applied to anomalistic psychology and consciousness studies we refer readers to Lange (2017).

We collated data from seven published studies (Irwin et al 2013; Irwin, Schofield, & Baker, 2014; Irwin & Wilson, 2013; Irwin, 2015, 2017, 2018; Ross et al., 2017) to conduct a comprehensive evaluation of the SAE's measurement properties via "top-down purification" analysis. We explored two primary issues:

- *Hypothesis 1*: Self-reported anomalous experiences (PAE) and paranormal attributions (PPA) on the SAE will conform to a hierarchical Rasch (1960/1980) scale, thereby providing evidence of discernible factors and conceptually replicating previous findings of inherent *probabilistic* structures underlying *paranormal-esoteric beliefs* (Lange, Irwin & Houran, 2000; Lange & Thalbourne, 2002; Irwin & Marks, 2013) and subjective *anomalous-paranormal experiences* (Houran & Lange, 2001; Houran, Wiseman & Thalbourne, 2002; Lange & Thalbourne, 2007; Lange, Thalbourne, Houran & Storm, 2000; Funkhouser, 2007; Lange, Greyson & Houran, 2008).

- *Hypothesis 2*: The PAE and PPA subscales will evidence (or can be readily modified to produce) interval-level, bias free measurements, which will improve the reliability of validity of Irwin et al.'s (2013) original SAE approach.



## METHOD

### *Rasch Scaling*

Readers might benefit from a summary of the basic features of the (binary) Rasch (1960/1980) model used in the present study, although more detailed information is provided by Bond and Fox (2015) and Lange (2017). This model focuses on $P_{ij}$ (the probability that person $i$ will endorse item $j$) given that the person has trait level $T_i$ and the item assesses the trait at level $D_j$. The latter is also referred to as the item's difficulty. The quantities are related as:

$$log(P_{ij} / (1 - P_{ij})) = T_i - D_j \qquad (1)$$

Items' fit to Equation 1 are assessed via their Outfit values, where values in the range 0.5 to 1.4 are generally deemed acceptable. Other factors can be included in the model to assess their impact (see e.g., Linacre, 2018a, b), either as main effects (e.g., men vs women), or as interaction effects. Of particular interest are interactions involving the $D_j$, as this indicates that items' difficulties vary across sub-groups, thus creating different item hierarchies. Such interactions are also referred to as "biases" or as Differential Item Functioning (DIF). For instance, we might find that some items are easier (or harder) to endorse for men than for women with equal trait levels; and these items are said to be biased or show DIF.

Rasch scaling defines reliability ($R$) analogous to the approach in CTT. However, with a Rasch scaling framework, reliability computations rely on the test-takers' estimated trait levels $T$ in Equation 1 rather than their raw scores. As a result, $R$ values tend to be generally lower than those for KR-20 or coefficient alpha. Rasch scaling allows for (non-systematic) missing data; hence, in contrast to the CTT approach, R can be estimated in the presence of missing data. We note that the trait levels T cannot strictly be estimated for "perfect" scores (i.e., when test-takers' responses are either *all 0* or *all 1*), and R is typically computed twice, once with and once without extreme scores. The former R value typically being higher than the latter.

All Rasch analyses performed here use the *Winsteps* and *Facets* software (Linacre, 2018a, b) whose JMLE estimation algorithm does not make any assumptions about the distribution of trait levels and item difficulties. As a result, the shapes of the item and person distributions does not materially affect the estimates of the *D* and *T* parameters. It is finally noted that Rasch scaling assumes that the test items should define a unidimensional dimension, and this issue is addressed here via factor analysis of items' Rasch residuals (see Linacre, 2018a).

*Data and Respondents*

We collated data ($n$ = 1,377) from seven published studies for analysis (i.e., Irwin, 2015, 2017, 2018; Irwin, Dagnall & Drinkwater, 2013; Irwin, Schofield & Baker, 2014; Irwin & Wilson, 2013; Ross et a., 2017), consisting of 351 men and 1026 women ($M_{age}$ = 32.3 yrs, $SD$ = 19.5, range = 18 - 75 yrs). These reflect predominantly convenience samples of students from Australian and British universities. We refer readers to the original papers for further details on the respective recruitment methods and respondents' demographic backgrounds. Data were originally collected in compliance with the Human Research Ethics Committees at Manchester Metropolitan University, the University of New England, and Royal Holloway, University of London. Conventions for obtaining informed consent required by each investigator's research institution, as well as IRB or ethical committees were followed. All participants provided written informed consent prior to participation. The study was conducted in accordance with the guidelines of the Declaration of Helsinki (World Medical Association, 2013).

---- Insert Table 1 about here ---

*Measure*

The *Survey of Anomalous Experience* (SAE) is a 20-item inventory developed by Harvey Irwin to gauge self-reported anomalous experiences (Irwin et al., 2013, pp. 51-53), and the items' text is shown in Table 1. For each item, respondents are presented with an anomalous experience described without any explicit reference to its possible cause.

Respondents who report having had such an experience are asked to further clarify their position by stating whether they attributed their experience to a specified paranormal process or alternatively, to a specified non-paranormal process. Thus, for each item addressing an anomalous experience, the respondent has three response options of the following general form: Option 1, "Yes, and it must have been [specific paranormal attribution]"; Option 2, "Yes, but it was probably just [specific non-paranormal attribution]"; and Option 3, "No."[1]

The index of *Proneness to Anomalous Experiences* (PAE) is defined in terms of the "yes" responses (i.e., Option 1 or 2 in any item) to each of the 20 items. Thus, Rasch scaling proceeds from a complete response record of 0's and 1's. Second, each participant's *Proneness to Paranormal Attributions* (PPA) to anomalous experiences was defined in terms of the "yes" (Option 1 or 2) responses with a paranormal attribution (i.e., the answer was "yes, paranormal", Option 1). By its very nature this introduces missing data, as scoring is conditional on at least one "yes" response. However, these "missing data" cases are rare, less than 1% of the total sample. While missing data are problematic when using CTT approaches, they are not generally problematic when using Rasch scaling, and this allows the computation of realistic estimates of scale reliability, even for the PPA. The preceding assumes that missing values occur in a non-systematic fashion, which is not the case here. Fortunately, it has been found that the impact of violations of this assumption is typically quite small (Linacre, 1994).

Since the PAE and PPA respectively omit or combine data, it seemed desirable to have an instrument that exhausts all available observations. For this reason, we introduce the SAE3, a three-category rating scale version of the SAE, which treats the three possible SAE answers as defining an ordinal progression toward paranormality, i.e., "no" (not observed) <



[1] Ross et al. (2017) lightly edited some of the items (see supplementary materials file included with their paper for precise wording) and the response options. Response items had the following general form: Option 1, "Yes, and it was probably [specific paranormal attribution]"; Option 2, "Yes, but it was probably just [specific non-paranormal attribution]"; and Option 3, "No".

"yes" (with *normal* attribution) < "yes" (with *paranormal* attribution). The psychometric performance of this formulation will be tested analogous to the PAE and PPA, i.e., we expect that Hypotheses 1 and 2 apply.

## RESULTS

Table 1 summarizes the Rasch scaling results for each of the three variables. Here, the items' location (difficulty) parameter are listed in the "Item Loc" columns and Outfit values in columns labeled "Outfit". DIF effects (if any) are shown in the columns so labeled. Other statistics are discussed later in context: Table 1 shows "Loading" columns, and Table 2 shows statistics related to test reliability, DIF, main effects, residual loadings, and loading correlations.

----- Insert Table 2 about here ----

### *Proneness to Paranormal Attribution (PPA)*

All twenty SAE items scored according to the PPA rules show good fit to the Rasch model as none of their Outfit values exceed the criterion value of 1.4 (range = 0.70 to 1.24, see PPA Outfit column Table 1). Next, we performed an overall statistical test for DIF (i.e., variation in items' $D_j$) due to age and gender using Linacre's (2018b) *Facets* software. Table 2 shows that neither Age- nor Gender-DIF reached statistical significance (see columns under DIF $\chi^2$). In other words, we failed to reject the null hypotheses that the item hierarchy is the same across younger and older people, as well as across women and men.

Table 2 further shows that the observed average person differences (i.e., $T_j$ in Equation 1) across gender and age are trivial (0.02 logits and 0.06 logits, respectively) and these differences fail to reach statistical significance (we used $p < .01$ throughout given the exploratory nature of our analyses). While the preceding is very encouraging, we note that the PPA has poor reliability. As is shown in Table 2, the Rasch reliability $R$ is 0.24 when extreme scores are included and 0.55 when extreme scores are excluded.

*Proneness to Anomalous Experience (PAE)*

When focusing on anomalous experiences (i.e., regardless of these being interpreted as normal or paranormal) the SAE items show much poorer fit. Specifically, as is indicated by the boldface entries in the PAE Outfit column in Table 1, the Outfit of items 1 and 5 (1.45 and 1.49, respectively) exceed the criterion of 1.4 by less than 0.10. However, these minor excursions are dwarfed by the Outfit values of 2.01 and 5.77 observed for items 3 ("being contacted by people one had been thinking about") and item 4 ("knowing what someone says before they say it"), respectively. Clearly, these items received very noisy (i.e., inconsistent) answers.

In addition, statistically significant overall item hierarchy differences were observed across gender (but not age, see Table 2), and we suggest that the misfit of Items 3 and 4 is related to their Gender-DIF. As is shown in the PAE "Gender-DIF" column of Table 1, women endorse items 4 and 5 more often than do men, the difference being 0.94 and 0.55 logits respectively. By contrast, men endorse items 8 ("envelope of light reflecting well-being"), 16 ("'self' was moving through a tunnel of light"), and 18 ("having abilities not inherited from parents") more often than women.

The preceding reflects important differences between men and women's item hierarchies. For instance, it follows from Equation 1 after solving for $P_{ij}$, that for men and women with below average trait levels (say, -2 logits) men interpret Item 3 as paranormal with probability 34% whereas women do so with probability 77%. The analogous probabilities for Item 4 are 55 and 88%, respectively. Table 2 shows that PAE has far greater reliability than does the PPA, and that there exist statistically significant gender and age main effects. However, given the items' misfit and biases, these findings should be interpreted with caution.

*SAE3*

Recall that the SAE3 redefines the responses as forming a 3-category ordinal scale. This yields a set of items that fits the Rasch rating scale model, because Outfit values for all items fall below 1.4. As was the case for the PPA and PAE, no Age-DIF was observed, but significant overall item hierarch differences were obtained for women versus men. Table 1 shows Gender DIF in Items 3 and 4 similar to the PAE, but now Item 5 ("impression of a figure") is over-endorsed by women as well. By contrast, 8 ("envelope of light reflecting well-being"), 11 ("being outside of physical body"), 13 ("healed by power of mind"), 14 ("object unaccountably moved"), and 16 ("'self' was moving through a tunnel of light") receive higher endorsement from men.

The pervasive Gender-DIF suggested that the SAE formulation is untenable and hence no further analyses were performed.

---- Insert Figure 1 about here ----

*Ancillary Analyses*

*Dimensionality*. To clarify the source of the various limitations in the three SAE formulations, we performed separate factor analysis of response's residuals for each model as obtained via *Winsteps* (Linacre, 2018a). Not all factor analyses are shown, but for the case of the PAE Figure 1 plots items' loadings according to their locations in logits (X-axis) and their loadings on the first (and most important) residual factor (X-axis). It can be seen that as items become more difficult (i.e., going from left to right along X) their loadings decrease systematically. Table 2 (Residual Factors, Loadings) shows the loading ranges for the three scale formulations, and it can be seen that the three-category formulation (SAE3) has the widest loading variation (range = -0.68 to 0.71).

Note that some residual factors are quite strong, they explain from 9 to 15% of the variance remaining after the items and persons are accounted for (see Table 2), the maximum occurring for the SAE3. Also, considerable agreement among the three residual factors' loadings was observed. In particular, Table 2 ("Loading Correlation") shows that items' residual loadings are highly correlated ($0.72 < r < 0.96$), i.e., each of the SAE scaling formulations showed similar residual loading patterns on the first residual factor across items' extremity.

*Study Differences*. Recall that we combined data from seven published studies and then treated this collection as a single data set. It is reasonable to ask therefore whether the item hierarchies vary across the studies that were used. This proved indeed to be the case as items' locations showed significant Study x Item interactions for each of the PPA, PAE, and SAE3 (all $p < .001$). These hierarchy shifts proved difficult to interpret and are not reported here. However, this clearly introduces additional caveats for using the SAE.



## Summary & Discussion

Our hypotheses received mixed support. Confirming Hypothesis 1, we replicated previous findings that anomalous experiences and paranormal attributions exhibit *hierarchical* structures (Funkhouser, 2007; Houran & Lange, 2001; Houran, Wiseman, & Thalbourne, 2002; Lange, Irwin, & Houran, 2000; Lange & Thalbourne, 2002; Irwin & Marks, 2013; Lange & Thalbourne, 2007; Lange, Thalbourne, Houran, & Storm, 2000; Lange, Greyson, & Houran, 2008). This implies that individuals perceive or endorse specific aspects of these constructs to systematically different degrees. As such, we seem to observe qualitative differences as these phenomena progress along their respective quantitative dimensions.

Contrary to Hypothesis 2, we could neither validate nor readily modify the SAE to serve as a robust "top-down purified" aggregate measure. Instead, we showed that each of the SAE's three derivative factors had its own set of limitations. First, the PPA fares quite well as it fits the Rasch model admirably. Yet, it should not be used in its current form due to poor reliability, and this cannot be rectified without adding more items. Second, the PAE showed acceptable reliability, but at least two of its items (i.e., "Sometimes I've been thinking of a person I haven't heard from in ages, and later in the day I received a phone call, email or letter from that very person." and "With someone I know intimately I sometimes know what they are about to say before they say it. ") are problematic given unacceptable item fit and powerful Gender-DIF, both of which distort measurement. It is not clear at this point whether these items can simply be reworded, or whether they simply do not belong to the same domain as the other items. Finally, while showing acceptable item fit, the SAE3 three-category formulation performed poorly due to pervasive Gender-DIF. Together with the finding of variation in items' locations across studies, the preceding implies that the SAE3 should not be used in its current form. Given the pattern displayed in Figure 1 for the PPA, and the similar patterns for the PAE and SAE3 (not reported), the extremity of items' difficulties should play a major role in this endeavor.

The SAE did not bear out as an adequate measure of its intended constructs, but our psychometric scrutiny of Irwin et al.'s (2013) approach revealed valuable, new insights that potentially transcend mere measurement issues. In particular, the item locations of specific PAE and PPA questions in their respective Rasch hierarchies revealed some interesting trends. In this context, we can distinguish between *internal* (*S*, subjective or internalized) and *external* (*O*, objective or externalized) categories of experience, following previous studies in anomalistic psychology (e.g., Houran, Wiseman, & Thalbourne, 2002; Hufford, 1982). The types of events in the Rasch hierarchies (i.e., those with lower logit values) that were most easily given credence (PAE) and attributed to the paranormal (PPA) seemed to involve personally-relevant incidents with at least some external (or objective) basis. Accordingly, we propose that our respondents are best characterized as predominantly striving to interpret private experience or information that was connected to (or corroborated by) information in their social or physical environments, i.e., amalgams of *S* and *O* features. This idea should be studied in more depth, since it adds a new dimension to the basic SAE approach and broader operationalizations encompassing religious, esoteric or even delusional ideations.

Moreover, across the PAE, PPA and SAE3 we observed a powerful residual or secondary factor that varied with the items' difficulty levels (i.e., extremity-bias). Establishing the prevalence and impact of this effect will be critically important, because, to our knowledge, it has not been previously reported in the literature, and its validation would have far-reaching implications. We note that earlier models have stressed the similarities and differences between anomalous experiences and unorthodox attributions (Lange & Houran, 1998, 1999), as well as their relation within linear and non-linear process models (Lange & Houran, 2000, 2001). However, it will become necessary to re-assess and re-interpret the assumptions currently held about the causes, correlates and clinical import of anomalous experience and paranormal attributions if measurement in this entire domain has been systemically distorted or confounded by extremity-biases.

As a first step we propose conceptual replications across independent samples and using different instruments pertinent to anomalous experience and paranormal belief (for listings of pertinent measures, see e.g., Goulding & Parker, 2001; Irwin, 2009, Appendix). These efforts should also include clinical measures of delusional beliefs, since the concepts related to the formation and labeling of anomalous experiences in the general population are germane to cognitive theories of delusions (cf. Houran & Lange, 2004; Irwin, Dagnall, & Drinkwater, 2012; Ross et al., 2017). Ultimately, we argue that the theoretical insights of the present research, in conjunction with Rasch (1960/1980) scaling, will aid greatly in refining the conceptualization and measurement of anomalous experiences and paranormal attributions, and ultimately promote more robust theory-formation grounded in anomalistic psychology.



## References

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, *62*, 271-298.

Dagnall, N., Drinkwater, K., Parker, A., & Clough, P. (2017). Paranormal experience, belief in the paranormal and anomalous beliefs. *Paranthropology*, *7*, 4-14.

David, A. S. (2010). Why we need more debate on whether psychotic symptoms lie on a continuum with normality. *Psychological Medicine*, *40*, 1935-1942.

Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, *8*, 133-158.

Drinkwater, K., Denovan, A., Dagnall, N., & Parker, A. (2017). An assessment of the dimensionality and factorial structure of the Revised Paranormal Belief Scale. *Frontiers in Psychology*, *8*, 1693.

French, C. C., & Stone, A. (2013). *Anomalistic psychology: exploring paranormal belief and experience*. New York, NY: Palgrave Macmillan.

Funkhouser, A. (2007). “Been there, done that:” Recent research into the phenomena commonly called “déjà vu.” Paper presented at the Psiber Dreaming Conference (September 23-October 7), Sonoma State University, Rohnert Park, CA.

Gallagher, C., Kumar, V. K., & Pekala, R. J. (1994). The Anomalous Experiences Inventory: Reliability and validity. *Journal of Parapsychology, 58*, 402-428.

Goulding, A., & Parker, A. (2001). Finding psi in the paranormal: Psychometric issues used in research on paranormal beliefs/experiences and in research on psi-ability. *European Journal of Parapsychology*, *16*, 73-101.

Houran, J., & Lange, R. (2001). A Rasch hierarchy of haunt and poltergeist experiences. *Journal of Parapsychology*, *65*, 41-58.

Houran, J., & Lange, R. (2004). Redefining delusion based on studies of subjective paranormal ideation. *Psychological Reports*, *94*, 501-513.

Houran, J., Lynn, S. J., & Lange, R. (2017). Commentary on Stokes’ (2017) quest for “white crows” in spontaneous cases of psi. *Australian Journal of Parapsychology*, *17*, 61-88.

Houran, J., Wiseman, R., & Thalbourne, M. A. (2002). Perceptual-personality characteristics associated with naturalistic haunt experiences. *European Journal of Parapsychology*, *17*, 17-44.

Hufford, D. J. (1982). *The terror that comes in the night: an experience-centered study of supernatural assault traditions*. Philadelphia: University of Pennsylvania Press.

Irwin, H. J. (2009). *The psychology of paranormal belief: a researcher's handbook*. UK: Univeristy of Hertfordshire Press.

Irwin, H. J. (2014). Aberrant salience and motivation as factors in the formation of belief in scientifically unaccepted phenomena. *Journal of Scientific Exploration*, 28, 633-645.

Irwin, H. J. (2015). Paranormal attributions for anomalous pictures: A validation of the *Survey of Anomalous Experiences*. *Journal of the Society for Psychical Research, 79,* 11–17.

Irwin, H. J. (2017). Empathy and parapsychological experiences: A constructive replication. *Journal of the Society for Psychical Research, 81,* 1-16.

Irwin, H. J. (2018). Stress sensitivity and minimal-self dysfunction as predictors of anomalous experiences and paranormal attributions. *Journal of the Society for Psychical Research, 82,* 1-14.

Irwin, H. J., Dagnall, N., & Drinkwater, K. (2012). Paranormal beliefs and cognitive processes underlying the formation of delusions. *Australian Journal of Parapsychology, 12*, 107-126.

Irwin, H. J., Dagnall, N., & Drinkwater, K. (2013). Parapsychological experience as anomalous experience plus paranormal attribution: A questionnaire based on a new approach to measurement. *Journal of Parapsychology, 77*, 39-53.

Irwin, H. J., & Marks, A.D.J. (2013). The 'survey of scientifically unaccepted beliefs': A new measure of paranormal and related beliefs. *Australian Journal of Parapsychology*, *13*, 133-167.

Irwin, H. J., Schofield, M. B., & Baker, I. S. (2014). Dissociative tendencies, sensory-processing sensitivity, and aberrant salience as predictors of anomalous experiences and paranormal attributions. *Journal of the Society for Psychical Research, 78*, 193-206.

Irwin, H. J., & Wilson, K. (2013). Anomalous experiences and the intuitive-experiential style of thinking. *Journal of the Society for Psychical Research, 77*, 65-71.

Jaspers, K. (1963). *General psychopathology*. Chicago, IL: University of Chicago Press (original work publ. 1923).

Jinks, T. L. (2012). Paranormal and alternative health beliefs as quasi-beliefs: Implications for item content in paranormal belief questionnaires. *Australian Journal of Parapsychology*, *12*, 127-158.

Jones, W. H., & Zusne, L. (1981). Teaching anomalistic psychology. *Teaching of Psychology*, *8*, 78-82.

Kihlstrom, J. F., & Hoyt, I. P. (1988). Hypnosis and the psychology of delusions. In T. F. Oltmanns & B. A. Maher (Eds.), *Delusional beliefs: interdisciplinary perspectives* (pp. 66-109). New York, NY: Wiley.

Kline, P. (1986). *A handbook of test construction*. London: Methuen.

Lange, R. (2017). Rasch scaling and cumulative theory-building in consciousness research. *Psychology of Consciousness: Theory, Research and Practice*, *4*, 135-160.

Lange, R., Greyson, B., & Houran, J. (2004). A Rasch scaling validation of a "core" near-death experience. *British Journal of Psychology*, *95*, 161-177.

Lange, R., & Houran, J. (1998). Delusions of the paranormal: A haunting question of perception. *Journal of Nervous and Mental Disease*, *186*, 637-645.

Lange, R., & Houran, J. (1999). The role of fear in delusions of the paranormal. *Journal of Nervous and Mental Disease*, *187*, 159-166.

Lange, R., & Houran, J. (2000). Modeling Maher's attribution theory of delusions as a cusp catastrophe. *Nonlinear Dynamics, Psychology, and Life Sciences*, *4*, 235-254.

Lange, R., & Houran, J. (2001). Ambiguous stimuli brought to life: the psychological dynamics of hauntings and poltergeists. In J. Houran and R. Lange (Eds.), *Hauntings and poltergeists: multidisciplinary perspectives* (pp. 280-306). Jefferson, NC: McFarland & Co.

Lange, R., Irwin, H. J., & Houran, J. (2000). Top-down purification of Tobacyk's Revised Paranormal Belief Scale. *Personality and Individual Differences*, *29*, 131-156.

Lange, R., & Thalbourne, M. A. (2002). Rasch scaling paranormal belief and experience: Structure and semantics of Thalbourne's Australian Sheep-Goat Scale. *Psychological Reports*, *91*, 1065-1073.

Lange, R., & Thalbourne, M. A. (2007). The Rasch scaling of mystical experiences: Construct validity and correlates of the Mystical Experience Scale (MES). *International Journal for the Psychology of Religion*, *17*, 121-140.

Lange, R., Thalbourne, M. A., Houran, J., & Storm, L. (2000). The Revised Transliminality Scale: Reliability and validity data from a Rasch top-down purification procedure. *Consciousness and Cognition*, *9*, 591-617.

Linacre, J. M. (1994). *Many Facet Rasch measurement*. Chicago: MESA Press.

Linacre, J. M. (2018a). *Winsteps® Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com.

Linacre, J. M. (2018b). *Facets® Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com.

Maher, B. A. (1988). Anomalous experience and delusional thinking: the logic of explanations. In T. F. Oltmanns and B. A. Maher (Eds.), *Delusional beliefs* (pp. 15-33). New York: Wiley-Interscience.

Maher, B. A. (1992). Delusions: Contemporary etiological hypotheses. *Psychiatric Annals*, *22*, 260-268.

Mehmet, S., & Yesilyurt, E. (2014). The development of paranormal belief scale (PBS) for science education in the context of turkey. *International Journal of Education in Mathematics, Science and Technology*, *2*, 107-115.

Peters, E. R., Joseph, S. A., Day, D. S., & Garety, P. A. (2004). Measuring delusional ideation: The 21-Item Peters et al. Delusions Inventory (PDI). *Schizophrenia Bulletin, 30*, 1005-1022.

Peters, E. R., Joseph, S. A., & Garety, P. A.  (1999). Measurement of delusional ideation in the normal population: Introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia Bulletin*, *25*, 553-576.

Preti, A., Vellante, M., & Petretto, D. R. (2017) The psychometric properties of the "Reading the Mind in the Eyes" test: An item response theory (IRT) analysis. *Cognitive Neuropsychiatry*, *22*, 233-253.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: University of Chicago Press.

Reiner, T. W. & Wilson, A. M. (2015). Assessing belief in the paranormal: An exploratory and confirmatory factor analysis. *Journal of Scientific Psychology*, August, 11-24.

Ross, R. M., Hartig, B., & McKay, R. (2017). Analytic cognitive style predicts paranormal explanations of anomalous experiences but not the experiences themselves: Implications for cognitive theories of delusions. *Journal of Behavior Therapy and Experimental Psychiatry, 56*, 90-96.

Schofield, M. B., Baker, I. S., Staples, P., & Sheffield, D. (2018). Creation and validation of the Belief in the Supernatural Scale. *Journal of Parapsychology*, *82*, 41-64.

Storm, L., Drinkwater, K., & Jinks, A. L. (2017). A question of belief: An analysis of item content in paranormal belief questionnaires   *Journal of Scientific Exploration, 31*, 187–230.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika*, *55*, 293-326.

Thalbourne, M. A., & Delin, P. S. (1993). A new instrument for measuring the sheep-goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research*, *59*, 172-186.

Tobacyk, J. (2004). The Revised Paranormal Belief Scale. *International Journal of Transpersonal Studies, 23,* 94-98.

van Leeuwen, N. & van Elk, M. (2018). Seeking the supernatural: The interactive religious experience model. *Religion, Brain, Behavior*.

Wright, B. D., & Stone, M.H. (1979). *Best test design*. Chicago, IL: MESA Press.