This document is protected by copyright. It is published with permission and all rights are reserved.

Usage of any items from Buckinghamshire New University's institutional repository must follow the usage guidelines.

Any item and its associated metadata held in the institutional repository is subject to

**Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)**

Please note that you must also do the following;

- the authors, title and full bibliographic details of the item are cited clearly when any part of the work is referred to verbally or in the written form
- a hyperlink/URL to the original Insight record of that item is included in any citations of the work
- the content is not changed in any way
- all files required for usage of the item are kept together with the main item file.

You may not

- sell any part of an item
- refer to any part of an item without citation
- amend any item or contextualise it in a way that will impugn the creator's reputation
- remove or alter the copyright statement on an item.

If you need further guidance contact the Research Enterprise and Development Unit
ResearchUnit@bucks.ac.uk
ABSTRACT

We review conceptualizations and measurements of base (or core) experiences commonly attributed to haunts and poltergeists (i.e., “ghostly episodes”). Case analyses, surveys, controlled experiments, and field studies have attempted to gauge anomalous experiences in this domain, albeit with methods that do not cumulatively build on earlier research. Although most approaches agree, to an extent, on the base experiences or events that witnesses report, the literature lacks a standard operationalization that can be used to test the factor structure of these occurrences or allow meaningful comparisons of findings across studies. Towards filling this gap, our review and deliberation identified 28 base experiences that include subjective (or psychological) experiences more typical of haunts, and objective (or physical) manifestations more common to poltergeist-like disturbances. This qualitatively-vetted list is proposed as the foundation for new measurement approaches, research designs, and analytical methods aimed to advance model-building and theory-formation.

Keywords: ghost, haunt, phenomenology, poltergeist, psychometrics
QUANTIFYING THE PHENOMENOLOGY OF GHOSTLY EPISODES: PART I – NEED FOR A STANDARD OPERATIONALIZATION

“If you can’t measure it, you can’t manage it.”

INTRODUCTION

Several early and prominent scientists – including Oliver Lodge, William Crookes and Frederick W. H. Myers – were keen to apply pioneering science in the study of spontaneous cases, and notably apparitions, haunts, and poltergeist-like outbreaks. Their activities with organizations such as the Society for Psychical Research (SPR) and the famous Ghost Club sparked a legacy of instrumentation that today is most closely associated with technical hardware (e.g., Hill, 2017; Houran & Lange, 1998; Potts, 2004). The literature from this perspective shows that ghostly experiences can involve measured or inferred physical events such as object movements, raps and knocking, electrical disturbances, malfunctioning equipment and anomalies on various types of recording media. That said, other important instrumentation seems to garner less widespread attention or appreciation, i.e., psychometric tools and test theory (Houran, 2017; Laythe & Owen, 2012).

To be sure, parapsychologists have long been interested in advances with questionnaire and survey design, data collection, and analytical techniques. For example, Gurney, Myers and Podmore (1886) conducted a large-scale survey of non-pathological ‘hallucinations’ – the first of its kind in the mind sciences – that revealed apparitions to be relatively common experiences in the general population. Shortly after, Gurney initiated an international replication with the SPR’s

---

1 This ubiquitous quote is ascribed to various sources, including Peter Drucker. See Lucas Coffeen, “If you can’t measure it, you can’t manage it,” MarketCulture Blog (March 20, 2009), http://blog.marketculture.com/2009/03/20/if-you-can’t-measure-it-you-can’t-manage-it-peter-drucker/, archived at http://perma.cc/3VVC-SGHT.
classic ‘Census of Hallucinations’ (Sidgwick, Johnson, Myers, Podmore, & Sidgwick, 1894). This was the first truly international project commissioned by the International Congress of Psychology and arguably set the stage for fresh reviews of historic and spontaneous cases dealing with ghosts and kindred phenomena (e.g., Alvarado & Zingrone, 1995; Finucane, 1996; Gauld & Cornell, 1979; Puhle, 2001; Roll, 1977), as well as opinion polls on global and specific paranormal beliefs and experiences in contemporary society (e.g., Palmer, 1979; Haraldsson, 1985; Ross & Joshi, 1992; McClenon, 2013; Dagnall, Drinkwater, Parker, & Clough, 2017). These collective efforts have documented a variety of subjective or psychological experiences inherent to ghostly encounters, such as visual apparitions or other entities and forms, voices or various auditory perceptions, unusual bodily sensations or sudden temperature changes, abrupt emotional disruptions and sensed presences.

As we address, research on ghosts and related phenomena is not sparse (although recent developments are seemingly sparse, see Laythe, Houran, & Ventola, 2018), but unfortunately the literature is not always uniform with operationalizations. Ghostly episodes do not seem to be a simple phenomenon but rather a multifaceted mixture of psychological, cognitive, environmental, and potentially parapsychological factors (for a review, see Houran & Lange, 2001b). Comprehensive and data-driven models are lacking arguably due to this complexity, as well as, in part, the absence of easily comparable and representative measurements of the fundamental features to these occurrences. As such, we first review various methods used to assess the phenomenology of ghostly episodes and then outline the practical and theoretical merits to a more robust psychometric approach.

We are not suggesting that important developments have not been made. In fact, meaningful progress has come from in-depth treatments of surveys, historical accounts and free-
response data, which go beyond basic content or affective theme analyses (e.g., Houran, 2013; Persinger & Makarec, 1992). For instance, techniques in forensic linguistics purport to appraise the internal veracity of witness accounts (e.g., Kohnken, 2004; Chaski, 2013; Kang & Lee, 2014), whereas the use of semiotics (Machado, 2001) or the more powerful method of computerized Latent Semantic Analysis (e.g., Lange, Greyson, & Houran, 2015) have been used as types of “factor analysis of language” to model the semantics and structure of paranormal narratives. Other approaches, such as Interpretive Phenomenological Analysis (e.g., Drinkwater, Dagnall, & Bate, 2013; Simmonds-Moore, 2016) and Conversation Analysis (Murray & Wooffitt, 2010; Wooffitt, 1992), take a qualitative, phenomenological approach combining hermeneutics and idiography to understand how percipients construct meaning from their experiences and likewise how experiences affect individuals. This last approach is particularly relevant given the way that witnesses may ‘mold’ their accounts in the face of skepticism (Ohashi, Wooffitt, Jackson, & Nixon, 2013).

Finally, we emphasize the use of leading-edge psychometrics grounded in Modern Test Theory, including Item Response Theory (IRT), Rasch (1960/1980) and differential item/test functioning (response bias testing). Often misunderstood, IRT allows questionnaire variables to be modeled into probabilistic interval-level and bias-free measurements similar to those in the physical and biomedical sciences. More importantly, IRT-constructed scales provide the benefit of consistent measurement with a mathematically set hierarchy that is fixed in actual constant measurement between items and respondents’ abilities or experiences. To borrow a metaphor from Bond and Fox (2015), the measure becomes an actual ruler in which an inch-is-an-inch across scores of the measure, regardless of comparing item-one to item-two or item-six to item-seven.
More technical overviews on Modern Test Theory are beyond the present scope, but resources are readily available that explain how these statistics transcend mere issues of measurement quality of questionnaires to speak directly to model-building and theory-formation (e.g., Bond & Fox, 2015; Wright & Mok, 2000). More specifically, Lange (2017) and colleagues (Houran, 2017, pp. 191-193; Houran, Lynn & Lange, 2017) have summarized several examples of IRT applications in parapsychology, including hierarchical models of constructs or outcomes such as paranormal belief and experience, success rates in experimental psi research, phenomenological aspects of near-death experiences, and discrete perceptions in haunt and poltergeist episodes. On this latter point, Rasch scaling of questionnaire data from both surveys (Houran & Lange, 2001a) and field studies (Houran & Lange, 2009; Houran, Wiseman, & Thalbourne, 2002) suggests that the physical manifestations and psychological experiences comprising ghostly experiences are inherently structured as suspected by earlier investigators (e.g., Palmer, 1974; Playfair, 1980; Pratt & Palmer, 1976). We discuss the nuances and implications of this and other important psychometric issues below.

**Previous Approaches to Assessing Ghostly Episodes**

Psychometric studies in this domain face the immediate challenge that generic claims of “encountering a ghost” can unwittingly entail separate classes of events unfamiliar to lay-people but differentiated by parapsychologists. On one hand, a witness could simply refer to a singular experience of an apparition or other anomaly, while, on the other hand, accounts could allude to a more complex haunt or poltergeist episode. Such issues of operationalization and the theories that drive nomenclatures are critically important. Some researchers clearly differentiate *haunts* and *poltergeists* (e.g., Gauld & Cornell, 1979; Dixon, 2016) or suggest they involve a constellation of
different phenomena (Houran & Lange, 1996; Cardeña, Lynn, & Krippner 2014). We argue that a firm distinction between the two episodes is currently problematic. That is, identifiable features of these episodes substantially overlap (cf. Williams & Ventola, 2011), and they sometimes seem to occur in tandem within individual cases (e.g., Dixon, 2016; McHarg, 1973). In fact, the content of both haunt and poltergeist episodes have been Rasch (1960/1980) scaled to form a collective hierarchy or continuum (Houran & Lange, 2001a; Houran, Wiseman, & Thalbourne, 2002). These patterns suggest a common underlying phenomenon or set of mechanisms might be operating across both types of episodes, consistent with others’ speculations (e.g., Evans, 1987, 2001; Hansen, 1988, Houran, 2000; Hufford, 1982, 2001). For these reasons and the purposes of this paper, we refer to apparitions and corresponding anomalies, haunts and poltergeists collectively as ghostly episodes.

Several authors have proposed that these occurrences have a structured or cumulative pattern of events. Sudre (1960) noted that ghostly events do not manifest continuously over the life of the phenomenon. Although they tend to occur in the same area, the actual display is episodic. Palmer (1974; Pratt & Palmer, 1976) was perhaps the first to propose the interesting notion that ghostly phenomena might progress systematically over time. Similarly, Nisbet (1979) suggested that, like some illnesses, poltergeist-like episodes have an ‘incubation’ period before phenomena begin. Experiences may then subsequently build upon themselves like a contagious process. ‘Contagious’ processes in conjunction with ghostly episodes are a well-known proposal (Kerner, 1836; Bauer, 1989; Houran & Lange, 1996; Romer, 2013). According to Guy Lyon Playfair (1980), there are approximately 19 “symptoms” of a poltergeist outbreak, beginning with raps and ending with equipment failure of cameras, tape recorders, and so forth. Individual cases may involve only half a dozen of these symptoms, but Playfair (quoted in Wilson, 1993) asserted rather
dogmatically that, “You always get them in the same order. You don’t get puddles of water before stone throwing, you don’t get fires before raps…there is a predictable behavior pattern. They appear to be random to us, but they’re obeying some sort of rules that they understand even if we don’t” (pp. 388-389). That assertion might sound overly deterministic, but Houran and Brugger (2000) similarly argued that haunts and poltergeists could form a hierarchy and that determining the probability of certain anomalies should provide clues to the nature of these phenomena or provide critical insights into specific cases. This hierarchy might also characterize séance-type phenomena. Recent work by Laythe and colleagues, in a post hoc examination of frequency distributions of internal and external perceptions within an occult-themed séance, found a repeated pattern of sensations and observations that suggested a hierarchical structure (Laythe, Laythe, & Woodward, 2017).

Indeed, ghostly episodes seem more structured than random. In the first of a series of studies on context effects in spontaneous cases, Lange, Houran, Harte, and Havens (1996) analyzed a large set of purportedly sincere ghost narratives published in commercial books. They developed a checklist (p. 757) of seven distinct types of anomalous experience commonly reported across these accounts, along with brief definitions:

1. **Visual**: Perception of a Form, e.g., a moving shadow, amorphous light, or a defined apparition which is mistaken as a real person.

2. **Auditory**: Sound phenomena that cannot be accounted for, e.g., footsteps, percipient's name being called out, or knockings.

3. **Olfactory**: Anomalous or unaccountable odor, e.g., the smell of flowers or cigars.

4. **Tactile**: Physical sensations, e.g., cold, heat, or a touch on the shoulder.

5. **Sensed Presence**: Feeling of being watched or not alone.
6. **Object Movement**: Subjective certainty that an object either unaccountably disappeared, appeared from seemingly nowhere, physically moved while in sight, or an inferred movement, e.g., losing a personal item, finding an object in your residence which does not belong to you, or a door opening on its own accord.

7. **Erratic Functioning of Apparatus**: Unaccountable malfunction or irregular operation of mechanical fixtures or electrical equipment, e.g., electrical current surges, telephone rings, light bulb failures, jammed door locks, and film processing difficulties.

These types reflect the definition of apparitional experiences provided by Baker (2002), which was adapted from Thalbourne’s (1982) *Glossary*: “A sensory experience in which there appears to be present a person or animal (deceased or living) who is in fact out of sensory range of the experient…” (p. 110). This checklist was subsequently used to study other types of “entity encounters,” including *angelic visitations* (Lange & Houran, 1996), *deathbed visions* (Houran & Lange, 1997), and *shamanic-trance journeys* (Houran, Lange & Crist-Houran, 1997). Consistent with earlier psychological thinking on apparitions (e.g., De Boismont, 1853; Jaffé, 1978; Tyrrell, 1943/1973), Houran’s (2000) meta-analysis of these studies revealed: (i) a strong congruence between the content of the experiences and the nature of the contextual variables (i.e., psychological or environmental cues) available to percipients; (ii) the number of contextual variables was related to percipients’ state of arousal immediately preceding the experience; and (iii) the number of contextual variables was also associated with the number of perceptual modalities involved in experiences.

Harte’s (2000) basic replication of Lange et al. (1996) amended the above checklist to include the new category of *Emotional Feeling*, defined as an “unaccountable onset of emotion (e.g., becoming inexplicably depressed, irritable, or fearful).” This amended checklist was later used to code free-response narratives of the research participants in a field study of Edinburgh’s historic South Bridge Vaults (Houran, Wiseman & Thalbourne, 2002). Other studies have utilized
more general measures of anomalous experience. For instance, Lange and Houran (1997) studied the role of expectation and suggestion effects in a purported haunt using an experiential questionnaire with 10 subscales related to psychological and physiological perceptions, which was originally designed for research on mirror-gazing and the influence of magnetic fields (Green et al., 1992).

A parallel study by French, Hague, Bunton-Stasyshyn, and Davis (2009) examined the role of expectation-suggestion, electromagnetic stimulation and infrasound for inducing ghostly experiences in a controlled, artificially-constructed “haunted chamber (room).” These researchers used a total score from a 20-item, true/false EXIT scale, adapted from a three-point Likert version by Granqvist et al. (2005), to measure anomalous experiences (e.g., “felt dizzy or odd, felt a presence, tingling sensations,” etc.) associated with the manipulation of environmental variables in the room. Although many of their participants reported anomalous sensations of various kinds, the number reported was unrelated to experimental condition but was related to scores on a measure of signs (or symptoms) of temporal lobe stimulation or activity in the general population. As a result, French et al. (2009) concluded that suggestibility was the most parsimonious explanation for their findings, although there has been criticism of their infrasonic measurement (Parsons, 2012). We should further emphasize that the EXIT scale does not have robust internal consistency, with Cronbach’s alpha coefficients ranging from .68 to .71, so its usefulness for ongoing research is questionable.

Taking a different approach, Kumar and Pekala (2001) presented an 8-item ‘Poltergeist’ subscale developed from existing items on their Anomalous Experiences Inventory (AEI: Kumar, Pekala., & Gallagher, 1994; cf. Gallagher, Kumar, & Pekala, 1994), which itself was adapted from the Mental Experiences Inventory (Kumar & Pekala, 1992). This subscale takes a more liberal
view of these episodes – one that places ghosts within a wider context of entity encounter narratives and traditions (Hufford, 1982; Evans, 2001; Houran, 2000). These items reference the themes of “seeing a ghost, possessed by an outside force, having a scary psychic experience, objects appearing or disappearing, objects floating in the air, communicating with the dead, and seeing fairies or other folklore type entities.” Kumar and Pekala (2001) found that scores on this subscale were positively related to several hypnosis-specific attitudes and behaviors.

A related subscale, an index of Encounter Experiences, comprises 11 “true/false” items from the AEI that encompass an even broader definition of encounter experiences (Pekala, Kumar, & Marcano, 1995, p. 323):

1. I am able to communicate with supernatural forces
2. I have experienced other planes of existence beyond the physical
3. I have had an out of body experience
4. I have tried channeling or have been a medium
5. I have communicated with the dead
6. I have seen a ghost or apparition
7. At times, I have felt possessed by an outside force
8. I can leave my body and return at will
9. I have experienced or met an extraterrestrial
10. I am able to communicate with the dead
11. I have seen elves, fairies, and other types of little people

We are aware of only a few studies using this index (i.e., Pekala, Kumar, & Marcano, 1995; Houran, Ashe, & Thalbourne, 2003; Houran, Kumar, Thalbourne, & Lavertue, 2002), which was originally designed for a study on shamanistic phenomena. However, this index consistently correlates with paranormal belief and permeability in mental boundaries, echoing patterns reported by Kumar and Pekala (2001) for the AEI Poltergeist subscale. Note that the Encounter index, like other methods reviewed here, aims to measure the number of different properties or modalities of encounter experience, as opposed to the frequency of each modality.
Consistent with the idea of a structured or cumulative phenomenology, Houran and Lange (2001a) found that the AEI Poltergeist subscale conformed to a Rasch (1960/1980) model. That is, Rasch scaling produced a linear measure of the experiences’ perceptions with the frequency by which each experience was reported being modeled as the outcome of a Poisson process. The notion that ghostly experiences define a probabilistic response hierarchy does not simply mean that experiences differ with respect to their endorsement rates. Rather, Rasch scaling requires that a scale of ghostly experiences forms a (latent) quantitative dimension on which each respondent and each type of ghostly perception assume a constant position (Bond & Fox, 2001). These positions reflect respondents’ trait-levels and the trait-level implied by the item, respectively, and together they determine the probabilistic likelihood of a given response on the scale. These item and person locations (also called, item and person measures) are expressed in a common Logit (δ) metric, which creates measurable interval-level positions on the scale (Wright & Masters, 1982).

Fit of the Rasch model implies that items form a hierarchy that reflects the structure of the variable, thus establishing construct validity and unidimensionality of a construct (Bond & Fox, 2015) by defining the variable’s semantics (for discussions, see Lange, 2017; Lange, Irwin, & Houran, 2000). For instance, given the item and person locations described above, the Rasch model implies that higher response categories should have a greater probability of being selected for items with lower locations than for items with higher locations. Also, respondents with higher trait levels should be more likely to give higher ratings than are respondents with lower trait levels. Finally, both properties should hold across all respondents, items, and response categories, creating a robust underlying unidimensional construct which is applicable across diverse populations. It can be shown (Wright & Masters, 1982) that when these requirements are fulfilled, the resulting...
variables have the property where the responses of those with lower trait-levels are probabilistic subsets of those with higher trait-levels.

Simplified in terms of the present context, Rasch scaling of ghostly episodes represents a fitted interaction of the respondent’s “sensitivity” levels to anomalous experiences (i.e., person’s inherent ability or receptiveness) and the rarity of a specific type of experience (an endorsement rate of experiencing a specific feature of ghostly episodes). The underlying assumption of a unidimensional Rasch scale is an assumption often made (incorrectly) with non-Rasch purified measures. Applied to ghostly episodes, it is not a rational issue but a probabilistic one that a person with little ghostly episode experience (low person experience) will endorse items that occur more rarely (item rarity). Rather, as a person has more ghostly episode experiences, the overall odds that experiences of rarer features to ghostly episodes become more likely, akin to the number of trials in a binomial or Poisson trials experiment with a set probability of an event occurring (Rozanov, 1969). Rasch scaling allows the verification of the above model by mathematically ensuring that ghostly episodes align with the rarity of their features (cf. Bond & Fox, 2015).

Rasch scaling was also used to examine the phenomenology of ghostly experiences in Houran, Wiseman, and Thalbourne (2002), drawing on the 8 categories of experience from content analyses (Harte, 2000; Lange et al., 1996) that we previously outlined in this paper. However, the hierarchy of experiences associated with the South Bridge Vaults did not fully agree with the Rasch order of four other items, sharing similar themes, from the AEI that Houran and Lange (2001a) reported previously. There are many possible reasons for this discrepancy. Differences between the two studies in terms of instruments, environments, instructional sets, country of origin and implicit demands could be confounding variables. The discrepancy could also imply that a probabilistic hierarchy of haunt perceptions is idiosyncratic to specific environments or that there
are distinct types of haunt cases whose features differ (cf. Gauld & Cornell, 1979). Nevertheless, there does appear to be some agreement on the sequence of general themes common to the two Rasch models. In both hierarchies, experiences of a more subjective (or psychological) nature tend to be endorsed most easily, followed by perceptions of objective (or physical) events in the intermediate part of the Rasch hierarchy, and finally the most extreme types of experiences (i.e., most rarely reported or occurring) correspond again to subjective (or psychological) features (see also: Houran, 2013).

We cannot push comparisons and contrasts between the two hierarchies too far given the caveats noted above. Still, hierarchies might well vary according to the environment in which the experiences occur. Interesting theoretical and methodological benefits would follow from this inquiry. For example, different Rasch hierarchies of ghostly experience might differentiate cases with evidential value for parapsychology from those grounded in fraud, imagination, or exposure to conventional environmental variables. Or, we might discover that certain people are differentially sensitive to certain features of ghostly episodes, thereby seeming to violate the hierarchy. For example, Houran and Lange (2009) found that experients’ levels of transliminality (or permeable mental boundaries) were associated with systematic distortions in the perceived phenomenology of their ghostly experiences. Detecting these types of differences becomes a straightforward case of comparing item placement (ranking) in contrast to the common measure.

In line with this approach, Houran (2002, p. 124) introduced a 25-item “Haunt Experiences Checklist” to uniformly document perceptions across participants \((n = 20)\) in a field study of a putative haunt. It was designed by collating (i) the AEI Poltergeist subscale; (ii) descriptions of various anomalous experiences reported by experients during the South Bridge Vaults field study (Houran, Wiseman, & Thalbourne, 2002); and (iii) reports of anomalous experiences from the case
collection used by Lange et al. (1996). Houran (2002) further proposed that the items on the checklist could be divided into two Classes of Experience (Psychological Experiences vs. Physical Manifestations) according to the rationale outlined in Houran, Wiseman, and Thalbourne (2002). Psychological Experiences comprised the summed total of visual apparitions (and related visual imagery), physiological alterations, emotional responses, and sensed presences (i.e., items #1, 2, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24), whereas Physical Manifestations consisted of the summed total of temperature changes, auditory experiences, bodily injury of some kind and olfactory experiences (i.e., items #3, 4, 5, 6, 7, 8, 9, 10, 23, 25). The coefficient alpha (measure of internal consistency) for Houran’s (2002) Checklist was .83, which well exceeds the traditional (Classical Test Theory) criterion of .70 for satisfactory reliability (Kline, 1986).

Unfortunately, there were too few data to perform robust Rasch analyses, but further refinement appears promising in terms of a comprehensive measure. Nevertheless, such analyses of this checklist are confounded by the fact that witnesses reported experiences involving multiple modalities or frequencies of those modalities, as do participants in other studies. Other researchers have adapted witness questionnaires from this earlier work. For instance, Dixon (2016, pp. 28-29) used a 25-item anomaly checklist as part of a study-specific Paranormal Investigation Survey used in his 13-month, longitudinal investigation of a ghostly episode. Comparison of these endorsements, particularly comparing numbers of apparitional experiences versus physical events (interpreted as ‘recurrent spontaneous psychokinesis: RSPK’ by Dixon) led the author to conclude that this spontaneous case represented a haunt-poltergeist hybrid, although this conclusion has been since tempered (Dixon, Storm, & Houran, 2018). Other studies have created study-specific questions or measures of ghost-like experiences (e.g., Wiseman, Watt, Greening, Stevens, & O’Keeffe, 2002; Wiseman, Watt, Stevens, Greening, & O’Keeffe, 2003), or used clinical measures
of constructs such as absorption, dissociation or schizotypy to gauge anomalous experiences attributed to apparitions and kindred phenomena (e.g., Parra, 2007; Parra & Argibay, 2006, 2016).

Furthermore, context can muddle inferences drawn from psychometric analyses. This is what Helms (1992) termed *contextual equivalence*, similar to cultural equivalence, which is broadly defined as, “the extent to which a cognitive ability is assessed similarly in different contexts in which people behave” (Reynolds & Suzuki, 2013, p. 105). The instruction set of polls and questionnaires that ask broadly about individuals’ past experiences do not capture the nuances and patterns of discrete perceptions that can occur in tandem within individual ghostly episodes and therefore best measured closer to real-time. In other words, Rasch hierarchies might differ based on responses to questionnaires that take an inventory of one’s anomalous experiences over a lifetime (e.g., Houran & Lange, 2001a; Kumar & Pekala, 2001) versus responses that reflect an individual’s collective perceptions within specific episodes (e.g., Dixon, 2016; Houran, 2002; Houran, Wiseman, & Thalbourne, 2002; Laythe & Owen, 2012). Therefore, instructions on questionnaires should be worded to be *case-specific* to gain valid insights about the phenomenology of singular episodes so that cross-study comparisons are interpretable.

Of course, these conceptual, measurement, empirical and pragmatic issues are largely moot until an effective measure of ghostly episodes is created and subjected to a range of validity tests. That process begins with face- and content-validities. Our review therefore included a mapping exercise that aimed to build upon the existing literature by identifying the degree of alignment in content among previous measurement approaches. As we discuss next, this served as the foundation for item selection in constructing a standard operationalization.
Elements Defining the Narratives of Ghostly Episodes

Despite specific differences in previous approaches to gauging “ghostly episodes,” we found reasonable agreement in the literature on the general core or base experiences characterizing witness reports. Table 1 lists these base experiences along with additional supporting references. We emphasize that this collection was subjected to an iterative process of review, discussion, and agreement among the authors, who sometimes argued for refining or distinguishing among subtleties in anomalies based on previous and independent research, ideology, as well as findings from field studies by several of the present authors. Lastly, we note that this list includes anomalies that are inferred or documented via recording media, which is a distinction not consistently made in the literature. We recommend this final collection of 28-core experiences as the basis for a standard operationalization in this domain.

To promote model-building and theory-formation, the base experiences are grouped according to their presumed point of origin. Specifically, manifestations are conceptualized as either Subjective (S, psychological) or Objective (O, physical) phenomena, consistent with previous work (Belz & Fach, 2015; Dixon, 2016; Dixon et al., 2018; Houran, 2002; Houran, Wiseman, & Thalbourne, 2002; Hufford, 1982, 2001; Laythe & Owen, 2013; Persinger & Cameron, 1986). Manifestations categorized S are those experienced by a singular observer via their own senses, or could be explained via artifacts of imagination, cognition and personality. A disturbance classified as O would be objectively experienced by a group of people, recorded on a device such as a camera/audio recorder, or measured with equipment such as a thermometer.

Additional research must validate this proposed S/O classification system, since distinctions between the nature of base experiences is neither always straightforward nor
unambiguous. For example, some haunt or poltergeist cases involve phenomena such as bites, cuts, scratches, welts, and possession-like trances (see e.g., Amorim, 1990; Mulacz, 1999). These externalized anomalies are objective and measurable, yet potentially psychosomatic in nature (e.g., Houran, Kumar, et al., 2002). On the other hand, psychological experiences—including abrupt onsets of emotions and “sensed presences”—can allegedly derive from environmental (i.e., objectively physical) agencies; for example, stimulation from electromagnetic fields (Persinger & Koren, 2001) or infrasound (Tandy, 2000; Tandy & Lawrence, 1998). Further, seemingly indisputable physical events like object movements, electrical failures and photographic anomalies are subject to orthodox causes if not extensively vetted but are commonly interpreted as “paranormal” via demand characteristics, expectation or suggestion effects, or other contextual variables (e.g., Houran, 1997, Houran & Brugger, 2000; Houran & Lange, 1996; Dagnall, Drinkwater, Denovan, & Parker, 2015; Irwin, 2015).

Some items in Table 1 could be refined further to tease out the best-fitting S/O categories, or better establish the “aberrant salience” (Irwin, 2014; Irwin, Schofield, & Baker, 2014) of specific anomalies. For example, auditory phenomena might be better understood as S vs. O events by asking about corroborating evidence from instrumentation, e.g., “I heard on an audio recorder mysterious sounds that could be recognized or identified, such as ghostly voices or music (with or without singing)” and “I heard on an audio recorder mysterious “mechanical” or non-descript noises, such as tapping, knocking, rattling, banging, crashing, footsteps or the sound of opening/closing doors or drawers” (see Laythe & Owen, 2013, for examples of S/O instrumentation-based vetting procedures). Likewise, there are likely to be critical differences between the implications of endorsements to statements such as, “I saw objects moving on their own across a surface or falling” versus “I saw objects flying or floating in midair” (see Gauld &
Cornell, 1979, for various phenomenological distinctions). We propose that answering the question of S/O phenomena first requires the reliable modeling of the observed types of phenomena themselves. Only then can we reliably begin to examine their contextual, environmental, psychological and potentially parapsychological facets.

**DISCUSSION**

Modeling the phenomenology of perceptions or reports in this domain has only just begun with any rigor from Lange’s (2017) “quality of measurement” perspective, but improvements in this area are critical to theory evolution. Undeniably, our understanding of other anomalous experiences has benefited greatly from detailed quantitative analyses of their features --- most especially near-death experiences (NDEs). In particular, Lange, Geyson, and Houran’s (2004) Modern Test Theory studies verified that Greyson’s (1983, 1985, 1990) NDE Scale captures an “NDE core experience.” That is, these experiences seemed to be (i) comprised of a probabilistic hierarchy of cognitive, affective, transcendental and paranormal components that was (ii) invariant across experiencers’ gender, current age, age at time of NDE, as well as years elapsed since the NDE (latency) and intensity of the NDE. Moreover, these same authors later showed that scores on the NDE Scale could be predicted from the specific elements in first-hand NDE narratives (Lange et al., 2015: cf. Houran, Lange, & Greyson, 2017). Altogether, this was among the first times that an empirical, quantitative description of NDEs substantiated insights only previously assumed from spontaneous cases alone (e.g., Moody, 1975; Ring, 1980).

Similar progress with ghostly episodes remains elusive, even though many case studies, surveys, and field investigations have attempted to define and measure base experiences in various ways. Consequently, consistent quantitative modeling arguably has been stifled. However, a
standard operationalization combined with opportunities for “big data” on these anomalies via collaborations with self-styled ghost-hunting groups or large-scale online (open-source) surveys could facilitate unprecedented advances (Houran, 2017). Speaking to future research directions and collaborations, a well-validated measure of ghostly episodes ultimately allows researchers in this domain to connect other significant components of the ghostly episode together in ways not previously possible. For instance, what appears to be a modest benefit of probabilistic interval-level measurement, is in fact, a vast untapped resource for validation, selection, and analysis of ghostly episodes. In terms of hypothesis formation, a probabilistic model allows an informed hypothesis towards the most likely features of other studies whereby nominal variables have been used to assess these episodes (e.g., believers or non-believers, Wiseman et al., 2002). Assuming a reliable probabilistic structure to ghostly episodes exists, and evidence thus far suggests this is likely, we now have a probabilistic breakdown of the most plausible phenomena that occurred within any measurement of “yes, a ghostly episode occurred.” In other words, a generalization of additional experiential details can be assumed when data specificity is lacking via simple probability applications (Rice, 1995; Rozanov, 1969).

In another vein, simple “yes/no” questions of environmental conditions, general locations, topography, geology, and other non-psychological variables, because they are often collected as nominal data, can be used to construct a form of “spooktistics.” Applications of using a probabilistically-standardized set of ghostly episode phenomena could easily lead researchers to areas, conditions, or populations that show abnormally high or low areas or conditions of activity. Such an application permits a reliable empirical assessment of environmental features and their effect on ghostly episodes (for an applied example of binomial modeling, see Laythe et al., 2017, Appendix B).
As we stated previously, two potential measurement and theoretical differences revolve around the issue of time-span in which ghostly episode experience is collected and the theoretical distinction of subjective versus objective classification of specific features of the ghostly episode. Both issues, without standardized measurement, create significant confounds and inhibits model-building, particularly in the association of psychological or environmental variables.

Regarding the former, it should be apparent that ghostly episodes, if measured collectively over a respondents’ lifetime (e.g., Houran & Lange, 2001a; Houran, Kumar, et al., 2002), will not equate to data collected from a single occurrence (e.g., Wiseman et al., 2002, 2003). The former will likely yield greater endorsement of features compared to the latter. However, as no consistent method has been applied to determine if life-time inventories of experiences substantively differ in content from single occurrences, a myriad of confounds exist in any study performing inferential statistics on these reports.

Maybe a tedious venture, but future research should empirically explore differences between these two perspectives, as predictions from studies using one type of operationalization are neither readily nor reliably applicable to studies with different ones. Conversely, with a Rasch purified measure, we may find that an underlying hierarchy of experience frames both, or even other, conditions (e.g., Lange et al., 2015). The essential point is that this confound of time-span in ghostly episodes cannot be fully addressed without a standardized measurement.

Regarding S/O distinctions in witness reports, we rest on the primary claim that first we need a comprehensive set of features that can be measured in a psychometrically-sound manner before empirical distinctions or confounds can be properly addressed. We believe the distinction is valuable in terms of separating psychological versus environmental based phenomena, but these classifications cannot remain arbitrary. It is our hope that once a standardized measure is achieved,
field research and comprehensive studies on the psychological correlates of ghostly episodes (e.g., Laythe et al., 2018) might provide empirical clues about which features of these anomalous experiences are predominantly psychological in nature (but possibly psi mediated, e.g., Rock & Storm, 2015), and which features are embedded in external stimuli, which may in turn contain parapsychological predictors. It remains the case that if we wish to successfully understand the individual features of ghostly episodes, we must specify the features we are examining, and their correlates, and conditions.

Related to the above, while our focus here is on identifying the base elements or experiences that should form a thorough psychometric assessment of ghostly episodes, comprehensively modeling these occurrences is likely to require big data and the use of certain technology once reliable features are established. The features or characteristics of a ghostly episode lend themselves to survey and psychometric assessment, but the eternal debate over the nature or core constituent components of ghostly phenomena will not be solved by surveys. Rather, some pieces of the puzzle can only be achieved by fieldwork applied to the features of ghostly episode, e.g., the vetting of phenomena by reliable and consistent means (e.g., Laythe & Owen, 2013), as well as measuring potentially important environmental variables such as electromagnetic fields, infrasound, or infrared or ultraviolet light (see e.g., Braithwaite, 2004, 2006, 2008; Braithwaite & Townsend, 2006; Joines, Baumann, & Kruth, 2012). The more robust, comprehensive, and simple the psychometric measure, the easier it will be to involve laypeople who actively investigate purportedly haunts or “poltergeist” cases. A standardized measure, in conjunction with the development of applications for mobile devices, could lead to a critical mass of connected data and evidence collected in a simple, but constant method, and designed in such a
way to prevent error in data collection. The number of empirical answers collected from such a “big data design” would be myriad and profound.

To reiterate, the practical goal is the development of a multi-purpose, “top-down purified” Rasch scale that systematically quantifies the phenomenology and intensity of specific ghostly episodes, while simultaneously controlling for potential response biases related to witnesses’ age, gender, and ideology (cf. Lange, Irwin, & Houran, 2000). We have hopefully, made the case for a standardized operationalization based on our review of previous methods, proposed incorporation of previous research in the proposed scale, and discussion of the numerous research benefits of such an operationalization. This tool would also be applicable to surveys, big data field studies and investigations that codify free-response data or spontaneous case material for quantitative analysis. Moreover, researchers would have a standardized method of conceptualizing and evaluating witness accounts with a measure robust enough to use with sophisticated and iterative statistical methods, e.g., power analysis, meta-analysis, path analysis (structural equation modeling), confirmatory factor analysis, and nonlinear dynamics. We argue this is the next generation approach for building conceptual and empirical frameworks in this domain. After all, if researchers cannot robustly measure the base experiences in question, it seems unlikely that academia will ever be able to manage or adequately explain the underlying causes, correlates and consequences of these anomalous phenomena.

REFERENCES


Murray, C., & Wooffitt, R. (Eds.) (2010). *Anomalous experiences and qualitative research. Special Issue, Qualitative Research in Psychology, 7*(1).


Psychometric ghost scale 38.


Table 1: Themes Represented in Narratives of Ghostly Episodes

<table>
<thead>
<tr>
<th>Putative Personal Experiences</th>
<th>SUPPORTING REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-descript anomalous image, like fog, shadow, cloud or streak of light</td>
<td>Daniels (2002), Drinkwater, Dagnall, &amp; Bate (2013), Gurney, Myers, Pease, &amp; Dawson (1883)</td>
</tr>
<tr>
<td>“Covert” ghost or apparition – anomalous presence that looked like a living person</td>
<td>Emmons, (1982), Daniels (2002), Gurney et al. (1883), Morton (1892)</td>
</tr>
<tr>
<td>Pleasant odor</td>
<td>Daniels (2002), Haraldsson (2009)</td>
</tr>
<tr>
<td>Unpleasant odor</td>
<td>Betty (1984), Haraldsson (2009)</td>
</tr>
<tr>
<td>Recognizable sounds, e.g., voices or music</td>
<td>Daniels (2002), Green &amp; McCreery (1975)</td>
</tr>
<tr>
<td>Onset of positive emotion</td>
<td>Drinkwater, Dagnall, Grogan, &amp; Riley (2017), Osis &amp; Haraldsson (1977)</td>
</tr>
<tr>
<td>Bodily sensations, e.g., dizziness, tingling, electrical shock, or nausea</td>
<td>Houran, Kumar et al. (2002), Wiseman, Watt, Stevens, Greening, &amp; O'Keeffe (2003), Young (2015)</td>
</tr>
<tr>
<td>Strange taste in mouth</td>
<td>Green &amp; McCreery (1975), Wiseman, Watt, Greening, Stevens, &amp; O'Keeffe (2002)</td>
</tr>
<tr>
<td>Guided, controlled or possessed by an outside force</td>
<td>Hess (1988), McClennon (1994, 2002)</td>
</tr>
<tr>
<td>Mystical beings, such as angels or demons</td>
<td>Daniels (2002), Dingwall (1947), McHarg (1973)</td>
</tr>
<tr>
<td>Folklore-type beings, such as elves, fairies, or other types of “little people”</td>
<td>Evans (1987); Jacome (1999); Puhle (2001)</td>
</tr>
<tr>
<td>Communication with the dead or other outside force</td>
<td>Braude (2014), Colvin (2008), Gurney et al. (1883), Hallowell &amp; Ritson (2008)</td>
</tr>
<tr>
<td>Déjà vu-type thoughts or feelings</td>
<td>Ember &amp; Ember (1988)</td>
</tr>
<tr>
<td>Putative Physical Events</td>
<td>References</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Electrical or mechanical appliances or equipment functioning improperly or not at all</td>
<td>Betty (1984), Daniels (2002), Persinger &amp; Cameron (1986), Roll &amp; Nichols (2000), McCue (2002)</td>
</tr>
<tr>
<td>Pictures from cameras/ mobile device with unusual images, distortions or effects</td>
<td>Daniels (2002), Lange &amp; Houran (1997); Maher &amp; Hansen (1995)</td>
</tr>
<tr>
<td>Plumbing equipment or systems functioning improperly or not at all</td>
<td>Betty (1984), Bugaj (1996)</td>
</tr>
<tr>
<td>Breaking of objects, e.g., glass, mirrors or housewares</td>
<td>Hess (1988), Roll (1977), Puhle (2001)</td>
</tr>
<tr>
<td>“Mechanical” or non-descript noises, e.g., rapping, knocks, rattling, apparent footsteps or sound of opening/closing doors or drawers</td>
<td>Betty (1984), Playfair (1980), Ellis (1978), Roll (1991), McCue (2002), Gurney et al. (1883)</td>
</tr>
<tr>
<td>Strange breeze or a rush of wind or air</td>
<td>Gauld &amp; Cornell (1979), Maher (2000), Tyrell (1973), Morton (1892)</td>
</tr>
<tr>
<td>Non-threatening physical touches, like a tap or light pressure on the body</td>
<td>Green &amp; McCreery (1975), Maher (2000)</td>
</tr>
<tr>
<td>Threatening physical touches, e.g., cut, bite, scratch, burn, shove or strong pressure on the body</td>
<td>Hallowell &amp; Ritson (2008), Hess (1988), Hufford (1982), Amorim (1990); Mulacz (1999)</td>
</tr>
</tbody>
</table>