Consciousness and Cognition 9, 591–617 (2000)
doi:10.1006/ccog.2000.0472, available online at http://www.idealibrary.com on IDEAL®

The Revised Transliminality Scale: Reliability and Validity Data From a Rasch Top-Down Purification Procedure

Rense Lange,*‡ Michael A. Thalbourne,† James Houran,‡ and Lance Storm†

Illinois State Board of Education, 100 N. First St., Springfield, Illinois 62777-0001; †Department of Psychology, University of Adelaide, Adelaide, South Australia 5005; and ‡Department of Psychiatry, Southern Illinois University, P.O. Box 5201, Springfield, Illinois 62705

The concept of *transliminality* ("a hypothesized tendency for psychological material to cross thresholds into or out of consciousness") was anticipated by William James (1902/1982), but it was only recently given an empirical definition by Thalbourne in terms of a 29-item Transliminality Scale. This article presents the 17-item Revised Transliminality Scale (or RTS) that corrects age and gender biases, is unidimensional by a Rasch criterion, and has a reliability of .82. The scale defines a probabilistic hierarchy of items that address magical ideation, mystical experience, absorption, hyperaesthesia, manic experience, dream interpretation, and fantasy proneness. These findings validate the suggestions by James and Thalbourne that some mental phenomena share a common underlying dimension with selected sensory experiences (such being overwhelmed by smells, bright lights, sights, and sounds). Low scores on transliminality remain correlated with "tough mindedness" in on Cattell 16PF test, as well as "self-control" and "rule consciousness," whereas high scores are associated with "abstractedness" and an "openness to change" on that test. An independent validation study confirmed the predictions implied by our definition of transliminality. Implications for test construction are discussed.

Key Words: transliminality; Rasch scaling; item bias; top-down purification; thresholds; William James.

INTRODUCTION

In his *History, psychology, and science,* Edwin Boring (1963) credited Herbart (and before him, Leibniz) with popularizing the notion of the threshold as applied to psychological phenomena. Sensory thresholds were already the subject of psychophysics, but toward the end of the 19th century there arose the idea that our everyday empirical consciousness lay above a threshold of sorts and that below that consciousness and that threshold was a whole region of mind with its own characteristics. Frederic Myers—poet, classicist, psychologist, and psychical researcher—called this region the subliminal consciousness (Myers, 1903; Thalbourne, Bartemucci, Delin, Fox, & Nofi, 1997). William James accepted this label while also referring to it as the "extramarginal" or "ultramarginal" region:

> I cannot but think that the most important step forward that has occurred in psychology since I have been a student of that science is the discovery, first made in 1886, that, in certain subjects at least, there is not only the consciousness of the ordinary field, with its usual centre and margin, but an addition thereto in the shape of a set of memories, thoughts, and feelings which

Address correspondence and reprint requests to James Houran. E-mail: rlange3726@aol.com.

1053-8100/00 $35.00

> are extra-marginal and outside of the primary consciousness altogether, but yet must be classed as conscious facts of some sort, able to reveal their presence by unmistakable signs. I call this the most important step forward because, unlike the other advances which psychology has made, this discovery has revealed to us an entirely unsuspected peculiarity in the constitution of human nature. (James, 1902/1982, p. 233)

And, elaborating:

> The most important consequence of having a strongly developed ultra-marginal life of this sort is that one's ordinary fields of consciousness are liable to incursions from it of which the subject does not guess the source, and which, therefore, take for him the form of unaccountable impulses to act, or inhibitions of action, of obsessive ideas, or even of hallucinations of sight or hearing. The impulses may take the direction of automatic speech or writing, the meaning of which the subject himself may not understand even while he utters it; and generalizing this phenomenon, Mr. Myers has given the name of *automatism*, sensory or motor, emotional or intellectual, to this whole sphere of effects, due to 'uprushes' into the ordinary consciousness of energies originating in the subliminal parts of the mind. (James, 1902/1982, p. 234)

And most relevant for our present theme:

> . . . the elementary mechanisms of our life are presumably so uniform that what is shown to be true in a marked degree of some persons is probably true in some degree of all, and may in a few be true in an extraordinarily high degree. (James, 1902/1982, p. 233)

James did not give a name to this process, but several years later Usher and Burt (1909) spoke of "transliminal" leakage between the subliminal and the supraliminal consciousness, and others since have likewise spoken of transliminal ("across the threshold") processes (Rugg, 1963; MacKinnon, 1971). However, "transliminality" apparently did not exist as a noun until the early 1990s (Thalbourne, 1991, p. 182), when it was conceived of as "an openness or receptiveness to impulses and experiences whose sources are in preconscious (or unconscious) processes."

The present article has two major purposes. We first review the extant research on the development of the notion of transliminality as a measurable psychological construct, which culminated in Thalbourne's (1998) 29-item transliminality scale. Although this scale served to establish the definition of the transliminality concept, its measurement properties remain to be determined. For instance, its development relied heavily on items' correlations with their common factor, a technique that has known methodological drawbacks (Comrey, 1978; Panter, Swygert, Dahlstrom, & Tanaka, 1997; Lange, Irwin, & Houran, 2000), and the use of factor scores does not alleviate this problem (cf. Michell, 1990). Accordingly, the second purpose of this article is to refine the current transliminality scale using the top-down purification process described in Lange et al. (2000) that combines Rasch scaling and the removal of age and gender bias at the test and item levels. The dimensionality of the resulting scale will be tested by competitive tests of one-factor vs two-factor Rasch formulations. In addition to providing interval level measures with realistic estimates of the standard error of measurement, item-purification typically decreases the number of items, thereby enhancing the practical usefulness of the resulting instrument. Being essentially a probabilistic version of Guttman's deterministic approach, Rasch scaling should also increase our understanding of the concept of transliminality as it expresses the "meaning" of this concept in terms of the response hierarchy implied

by Thalbourne's original items (cf. Wright, 1999). The following sections address each of these issues in more detail.

Development of the Transliminality Construct

Studies of perception, imagery, and memory all provide some evidence for a threshold that mediates unconscious-conscious awareness. However, we agree with Baars (1988), who noted that an activation threshold by itself is not necessarily sufficient to produce conscious experience. This is illustrated by phenomena like habituation and the automatization of conscious experiences when perceived internal or external stimuli are presented repeatedly. In these cases, people generally lose conscious awareness of the stimuli. Consequently, if we accept that conscious experience corresponds in part to activation above some threshold, as Herbart (1824/1961) and James (1902/1982) suggested, we must also accept the seemingly paradoxical idea that too much activation can lead to a *loss* of conscious experience. The notion of transliminality presented here is not inconsistent with this idea because it emphasizes a flow of psychological material *into* and *out* of conscious awareness. For more detailed thoughts on this issue, we recommend Baars' (1988) discussion of theories that simultaneously account for activation thresholds as well as availability and efficiency.

Thalbourne's (1991) theoretical intuition concerning transliminality was given some empirical flesh by Thalbourne and Delin (1994), who found a single factor underlying six relevant variables (belief in, and experience of, the paranormal; magical ideation; creative personality; mystical experience; maniclike experience; and depressive experience). The definition of transliminality was thereupon refined to "a largely involuntary susceptibility to, and awareness of, large volumes of inwardly generated psychological phenomena of an ideational and affective kind" (p. 25). Using the factor scores on this first factor, a number of correlates were uncovered, including religious variables, a proneness to hallucination, and variables related to mental illness and medication. Also correlated was attitude to dream interpretation ("A person should try to understand their dreams and be guided by or take warnings from them"), and this was confirmed by Thalbourne and Delin (1995). Based on a follow-up study conducted in 1991–1992, Thalbourne and Delin (1999) found that transliminality correlated with a dream-recall scale, general religiosity, frequency of dream interpretation, and with two other measures of mystical experience (as might be expected, given that mystical experience was a variable that was originally seen as a constituent of transliminality). Also, Thalbourne (1996) found evidence to suggest that hyperesthesia (heightened sensitivity to sensory stimulation) was a part of transliminality.

Thalbourne et al. (1997) replicated the finding of a single transliminality factor, except that depressive experience now failed to load on this factor. Transliminality correlated significantly with Claridge and Broks' (1984) STA schizotypy scale, Eysenck and Eysenck's (1991) psychoticism scale (P), a measure of fantasy proneness (Myers, 1983), Tellegen and Atkinson's (1974) absorption scale, and an index of hyperesthesia (different from the one used by Thalbourne, 1996). As predicted, transliminality did not correlate with intelligence as measured by the Raven Progressive Matrices (Raven, 1965), but, contrary to prediction, it was likewise unrelated

to repression sensitization (Byrne, Barry, & Nelson, 1963) and three of Torrance's (1966/1974) tests of verbal creativity. Thalbourne et al.'s (1997) study marked a milestone for the definition of transliminality as it now appeared to involve variables which had little to do with the subliminal feeding into the supraliminal. *First,* active fantasy proneness would seem to consist of a seed idea in the supraliminal consciousness (an idea which we could call "psychospermia") arousing material from the subliminal and by various creative processes shaping it into a theme, so transliminality seemed to be, sometimes, a two-way subliminal–supraliminal process. *Second,* and more radically, hyperesthesia has to do with thresholds of the perceptual system's interaction with the outside world. So, perception also seemed to have a transliminal component. Accordingly, the definition of transliminality was changed to "susceptibility to, and awareness of, large volumes of imagery, ideation and affect—these phenomena being generated by subliminal, supraliminal and/or external input" (Thalbourne et al., 1997, p. 327).

Given the above, Thalbourne (1998) administered to a sample of more than 300 psychology students the basic transliminality variables, together with the significant correlates uncovered by Thalbourne et al. (1997), plus measures of dissociation (Bernstein & Putnam, 1986; Riley, 1988) and of hallucination proneness (Launay & Slade, 1981). Using factor analysis, a single factor was again found to underlie the (now five) original variables constituting transliminality. Other variables that appeared to be part of transliminality were fantasy proneness, attitude toward dream interpretation, absorption, and hyperesthesia: These variables correlated not only with each and every one of the original variables, but also with each other. These nine variables, by way of a factor score and its correlations with the individual scale-items, provided the material for a unidimensional (i.e., in the factor analytic sense) 29-item Transliminality Scale. This new scale showed several correlates, including schizotypal personality, dissociation, proneness to hallucination, psychoticism, and interest in Eastern religion (but not in Bible reading).

Finally, Thalbourne and Houran (2000) administered this new set of items to large general-population samples in both Australia and the United States, along with the Mental Experience Inventory (Kumar & Pekala, 1992) and (in the case of 100 American participants) Lange and Houran's (1999a) Rasch adaptation of MacDonald's (1970) AT-20 scale of tolerance of ambiguity. This last scale showed a low and non significant correlation with transliminality. However, in both countries transliminality showed high positive correlations with paranormal belief and experience, sense of being high, daydreaming and fantasizing, sense of mental potency, introspection (as predicted by Thalbourne & Delin, 1994), and altered consciousness. This article presents the latest definition of transliminality, as namely "a hypothesised tendency for psychological material to cross thresholds into or out of consciousness." Most recently, Storm and Thalbourne (1998–1999) administered the Transliminality Scale together with the Cattell Sixteen Personality Factor test (16 PF; Russell & Karol, 1994) and the findings of this research are reanalyzed in a later section.

We note that explication of the transliminality construct is also important because it resembles related concepts proposed by others. For instance, Thalbourne and Delin (1994) already mentioned the relatively old concepts of openness to experience, flexibility of repression, and ego permissiveness. More recently, Virtanen (1990) sug-

gested that anomalous/paranormal experiences apparently reflect physiological processes associated with the movement of information from nonawareness to awareness during various states of consciousness. This notion coincides with Hartmann's (1991) theory that "boundaries in the mind" affect important behavioral and cognitive variables. He has devised a standardized questionnaire measuring various types of cognitive boundaries and presents evidence indicating that persons with "thin" (permeable) boundaries are more likely to be highly hypnotizable, to experience anomalous events, and to experience certain forms of psychopathology. Virtanen's and Hartmann's theories thus involve thresholds of awareness and to that extent their basic concepts are cousins of transliminality.

Rasch Scaling

While the transliminality research reviewed above yielded a clear pattern of results, the indices that were used previously are essentially weighted counts of "positive" answers. As is the case in general, such indices are ordinal at best and they cannot be relied on to produce measures that are conjoint additive (Michell, 1990; Wright, 1999) and unbiased. Following the earlier lead provided by research in educational testing (Lord & Novick, 1968; Rasch, 1960; Wright & Stone, 1979), the psychological literature now increasingly recognizes that classic test theory often fails in these respects (Embretson, 1995), as it does not explain how respondents' answers can be understood as a function of the latent variable addressed by the items. The following describes how Rasch scaling provides a solution to these problems. Since Rasch scaling may be unfamiliar to many readers, we describe the assumptions underlying this approach in some detail (general introductions can be found in Wright, 1999; van der Linden & Hambleton, 1997; Embretson & Herschberger, 1999; Wright & Mok, 2000).

Rasch scaling assumes that the probability (P) with which a person i endorses some item j depends *solely* on (a) this person's position on an underlying latent variable θ (in our case, transliminality) as expressed in theoretical units called "logits" (for a discussion see, e.g., Ludlow & Haley, 1995; Wright & Stone, 1979) and (b) the location δ_j of this item's characteristic curve (ICC) on this latent variable. The ICC is assumed to follow a logistic function, which is centered on the point δ_j. Thus, the likelihood that a person with transliminality level θ will endorse this item is given by the conditional probability:

$$P(\theta|\delta_j) = (1 + e^{-\theta+\delta_j})^{-1}. \quad (1)$$

The solid lines in Fig. 1 show the ICC for a hypothetical three-item test with $\delta_1 = -1$, $\delta_2 = 0$, and $\delta_3 = 2$. It can be seen that a person with $\theta = 0$ is likely to endorse Item 1 but not Item 3. Note that all solid ICCs have the same slope and that $P(\theta|\delta_j) = 0.5$ when $\theta = \delta_j$, as is illustrated in Fig. 1 for the case of Item 2. In the present research, the quantities θ and δ both increase with greater transliminality. Accordingly, items with greater δ are *less* likely to be endorsed.

Most importantly, whereas classic methods essentially produce ordinal scales of measurement, Rasch models yield measures at an *interval* level. For instance in Fig. 1, one may assume that the difference between the locations of Items 2 and 3 (i.e.,

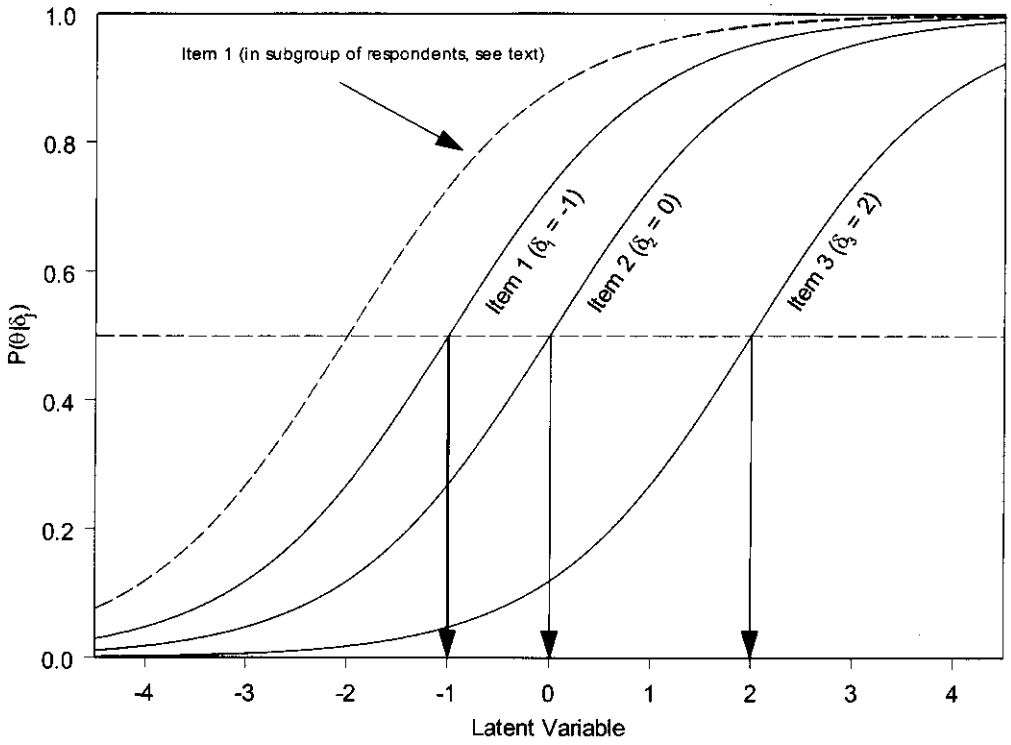

FIG. 1. Hypothetical three-item test (Item 1 shows DIF).

$\delta_3 - \delta_2 = 2$) is indeed twice as large as the difference between the δ values of Items 1 and 2 (i.e., $\delta_2 - \delta_1 = 1$). Similar considerations apply to the positions θ of individuals on the latent variable.

Equation (1) has strong implications for the meaning of the transliminality measure. In particular, the fit of this equation entails that the items form a probabilistic hierarchy of items such that they can be renumbered according to their likelihood of being endorsed, i.e., $P(\theta|\delta_1) < \ldots < P(\theta|\delta_i) < \ldots < P(\theta|\delta_n)$. And, since the ICCs never cross, *this ordering is the same across respondents.* Accordingly, if person *a* is known to posses a transliminality level θ_a, then this person is more likely than not to endorse all items *i* with $\delta_i < \theta_a$, but *not* to endorse all items *k* with $\delta_k > \theta_a$. Additionally, some person *b* with $\theta_b > \theta_a$ is also likely to endorse the items with $\delta_i < \theta_a$, but *b* more likely than not will also endorse the items *j* with $\theta_a < \delta_j < \theta_b$. In other words, the items endorsed by respondents with lower trait values are a probabilistic subset of those endorsed by respondents with higher trait values. Thus, Rasch scaling uniquely identifies those items that distinguish higher levels of transliminality from lower ones. As is illustrated in a later section, this allows a unique "mapping" of the item locations across the latent dimension. Naturally, neither the locations of the items on the latent variable, nor those of the respondents, are known

beforehand. Therefore, they must be estimated from the data using specialized software such as Bigsteps (Linacre & Wright, 1997).[1]

Reliability. To simplify the following it is assumed that the items being scaled are unidimensional, but we touch on multidimensional extensions in a later section. Regardless of the number of dimensions involved, it is assumed that the items defining a particular factor are "locally independent," i.e., the responses for all respondents *with the same* θ should not contain any statistical dependencies. Local independence allows the positions of the items and those of the respondents to be estimated separately and independently of each other (Wright & Stone, 1979). As a result, one can distinguish between the reliability of the estimation of the item parameters δ and that of the estimates of respondents' locations θ on the latent variable (R). The latter type of reliability corresponds most closely to the notion of score reliability (r_{xx}) in classic test theory as expressed by the KR-20 coefficient (i.e., coefficient alpha for binary items).

The reliability of a Rasch measure depends exclusively on the fact that the information about respondents' trait levels increases to the extent that their θ coincide with the δ_i (Wright & Stone, 1979). In other words, the errors (SE_θ) made in estimating persons' transliminality *do not depend on the sample being studied* (except through sampling error). Instead, SE_θ is a "local" quantity that is determined by the position of θ relative to the item locations δ. In practice, the range of item locations is often smaller than that of the respondent locations and therefore extreme person measures (high or low) tend to have larger standard errors than intermediate measures. Further, although respondents who endorse all (none of the) items must have a high (low) trait value, we do not know *how high (or how low)* and their SE_θ can only be inferred. Given its relation to the reliability (Lord, 1980), the local nature of SE_θ has direct implications for R. For instance, while reporting the overall R may be useful for purposes of quick communication, it is more appropriate to also use a local (i.e., θ-specific) pseudo-Rasch reliability R_θ. Analogous to the classic reliability, r_{xx}, this local reliability is defined as (Daniel, 1999):

$$R_\theta = 1 - \frac{SE_\theta^2}{S_\theta^2}, \tag{2}$$

where S_θ^2 represents the variance in the Rasch person measures. Typically, R_θ is 0 for the highest and lowest possible scores.

Model fit. As was explained earlier, persons with $\theta > \delta_j$ are likely to endorse item j, while those with $\theta < \delta_j$ are not, and this forms the basis for computing items' *infit* (i.e., the fit relative to items at nearby locations) and *outfit* (i.e., the fit relative to items at more distant locations) (for details, see Wright & Stone, 1979). The infit and outfit statistics computed by the Bigsteps software have an average theoretical model value of 1, and values in the range 0.7 to 1.3 are generally deemed acceptable (Linacre & Wright, 1997). It is a major advantage of the Rasch model that the estima-

[1] A publicly accessible respository of information related to Rasch scaling can be found at the website www.rasch.org. Visitors may also download a free copy of the Bigsteps software.

tion of the δ_j parameters is essentially "population free" because their estimates depend little on the particular sample in which they were obtained. An important quality indicator is the items' capability to spread the respondents across the Rasch dimension. For this reason, we also report the distributions of the θ and the δ_j across their common logit scale.

Differential Item Functioning. Local independence entails that extraneous factors such as gender or age should have no effect on the responses once transliminality is taken into account. Formally, this implies that the conditional probabilities of the following type should hold for each item j:

$$P(\theta|\delta_j, E) = P(\theta|\delta_j, \bar{E}), \tag{3}$$

where θ is defined as before and where E denotes one of the values of a binary variable such as gender or age (e.g., obtained by classifying the respondents as either "young" or "old") and $\bar{E}$ denotes E's complementary value. To the extent that Eq. (3) is violated for a particular item, this item is said to exhibit Differential Item Functioning (DIF) as related to E.[2] Alternatively, if the δ_j are computed separately for respondents with E and $\bar{E}$, DIF manifests itself by shifts in the item locations δ_j as a function of group membership (see dotted ICC in Fig. 1). Despite the fact that Eq. (3) is couched in the language of Rasch modeling, we note that analogous considerations apply within the framework of classic test theory (although, unfortunately, DIF is rarely considered there).

While several statistical methods to assess DIF have been developed (for a comprehensive review, see Clauser & Mazor, 1998), the present research uses the SIBTEST method proposed by Shealy and Stout (1993). The SIBTEST software (Stout & Roussos, 1996) quantifies *DIF* in terms of the β-statistic, which captures the overall group differences in the probabilities on the left- and right-hand sides of Eq. (3). Specifically, β is a direct estimate of the overall differences in groups' response probabilities, given θ. SIBTEST also provides statistical tests to determine whether the obtained β-values differ significantly from zero. Simulation studies (Shealy & Stout, 1993) indicate that SIBTEST possesses good Type I error control, while providing acceptable power in detecting DIF when relatively small samples are used. The biasing variables of primary interest in the present research are the respondents' ages and gender.[3]

Differential Test Functioning. We note that the item-level distortions introduced by DIF may combine to introduce systematic biases into the measure derived from the entire test as well. If so, we speak of Differential Test Functioning (DTF). In the present context, DTF manifests itself in a differently shaped item sum to Rasch mea-

[2] The IRT literature distinguishes between uniform and nonuniform DIF. As the latter results in poor item fit statistics, it is likely to be excluded during the scale construction. Accordingly, only uniform DIF is considered in the following.

[3] In educational contexts (see e.g., Ackermann, 1992), a distinction is made between "benign" and "adverse" DIF. Differential item functioning is "benign" when the second dimension reflects an unavoidable or desirable feature of the test (e.g., Reading Comprehension unavoidably also reflects knowledge related to the subject matter). No such considerations apply in the present context, and all DIF is considered adverse for our purposes.

TABLE 1
Summary of Scaling Analyses

	Original (29 items)	Purified (17 items)	Excluded (12 items)
Classic			
KR-20 reliability	0.89	0.85	0.75
Rasch			
Overall reliability *R*	0.88	0.82	0.72
Item separation	2.68	2.14	1.61
Range—item infit	0.81–1.31	0.88–1.28	0.82–1.29
Range—item outfit	0.66–1.47	0.72–1.25	0.56–1.58
Items shown in table(s)	2 + 4	2	4

sure translations as computed by Bigsteps for different age and gender groups. DIF does not necessarily produce DTF, as it is possible that the DIF in one or more items may cancel (Ackermann, 1992, 1996; Waller, Thompson, & Wenk, 2000). For such cancellation to occur reliably, it is required that all items are always administered and that subjects' response records contain no missing data. Such limitations can be avoided by eliminating *all* biased items, i.e., even when keeping such items would not result in noticeable DTF. As this point is often misunderstood, we further stress that the absence of DIF and DTF does *not* imply that men and women, or older and younger respondents, should have similar group means. Instead, as DIF may act "for" as well as "against" a particular group, its removal can actually *accentuate* group mean differences.

Multidimensionality. While the assumption of unidimensionality is supported by a satisfactory item fit (Hattie, 1985; Linacre, 1998a), explicit dimensionality tests are provided by Stout's DIMTEST (Nandakumar & Stout, 1993) and the powerful ConQuest software (Wu, Adams, & Wilson, 1998). ConQuest has the advantage of being able to fit multidimensional as well as unidimensional Rasch models and it provides χ^2 tests of fit, thus allowing for competitive model testing. Both of these features are exploited here. We point out, however, that almost all item sets are multidimensional to some extent (Hattie, 1985; Linacre, 1998a) and that multidimensionality is fueled by DIF (Stout, 1990; Lange et al., 2000). Therefore, we additionally rely on ConQuest's direct (i.e., unattenuated) estimates of the correlation between the Rasch factors to evaluate the actual impact of any statistically significant multidimensionality.

METHOD

The Transliminality Scale data (items and total scores) were obtained from 318 individuals (126 known men and 189 known women) with a mean age of 35.8 years (SD = 13.75; Md = 36.00 years; range: 17–84 years). The exact wording of the 29 transliminality items is given in Tables 2 and 4 below. These persons had participated in either the experiment of Storm and Thalbourne (1998–1999) or the study of Thalbourne and Houran (2000). In the former study participants also completed the Cattell 16PF (Russell & Karol, 1994), while participants in the latter study participants also completed Kumar and Pekala's (1992) Mental Experience Inventory.

TABLE 2
Item Locations, Fit Statistics, and Significance of the β-Values in DIF Tests by Gender and Age for the 17 Purified Items

Item	Item location (δ, in logits)	Item location (scaled)	Infit	Outfit	Gender β[a]	Age β[b]
2. At times I perform certain little rituals to ward off negative influences	0.05	25.45	1.16	1.18	0.00	0.01
3. I have experienced an altered state of consciousness in which I felt that I became cosmically enlightened	0.73	27.42	0.89	0.76	−0.08*	−0.05
4. At the present time, I am very good at make-believe and imagining	−0.53	23.76	1.16	1.25	−0.07	0.04
5. I have felt that I had received special wisdom, to be communicated to the rest of humanity	1.87	30.72	0.90	0.77	0.03	−0.07**
8. I have sometimes sensed an evil presence around me, although I could not see it	0.16	25.76	0.96	0.93	0.05	0.03
9. My thoughts have sometimes come so quickly that I couldn't write them all down fast enough	−1.41	21.21	1.07	1.13	−0.05	0.10**
12. It is sometimes possible for me to be completely immersed in nature or in art and to feel as if my whole state of consciousness has somehow temporarily been altered	−0.91	22.66	0.89	0.79	−0.06	−0.01
13. Often I have a day when indoor lights seem so bright that they bother my eyes	1.17	28.69	1.28	1.17	0.01	0.04
16. I have experienced an altered state of consciousness which I believe utterly transformed (in a positive manner) the way I looked at myself	0.80	27.62	0.91	0.94	0.08*	−0.05
18. I think that I really know what some people mean when they talk about mystical experiences	−0.36	24.26	0.89	0.87	0.05	−0.07*
19. I have gone through times when smells seemed stronger and more overwhelming than usual	−0.45	24.00	1.06	1.08	0.06	−0.07
20. I can clearly feel again in my imagination such things as: the feeling of a gentle breeze, warm sand under bare feet, the softness of fur, cool grass, the warmth of the sun and the smell of freshly cut grass	−1.03	22.31	1.00	1.14	0.01	−0.02
21. A person should try to understand their dreams and be guided by or take warnings from them	−0.76	23.10	0.99	0.95	0.04	0.08*
22. At times I somehow feel the presence of someone who is not physically there	−1.77	20.14	1.02	0.79	0.02	−0.05
25. For several days at a time I have had such a heightened awareness of sights and sounds that I cannot shut them out	2.05	31.25	0.88	0.72	−0.03	0.00
26. I sometimes have a feeling of gaining or losing energy when certain people look at me or touch me	−0.19	24.75	0.89	0.80	−0.06	0.04
29. When listening to organ music or other powerful music, I sometimes feel as if I am being lifted up into the air	0.59	27.01	0.99	0.96	0.06	0.07*

[1] A positive value indicates that women's (N = 189) likelihood of endorsing the item is greater than that of comparable men (N = 126). A negative β-value indicates the opposite effect.

[2] A positive value indicates that younger respondents' (N = 156) likelihood of endorsing the item is greater than that of comparable older ones (N = 157). A negative β-value indicates the opposite effect.

* $.10 < p < .20$.

** $.05 < p < .10$ (all other $p > .20$).

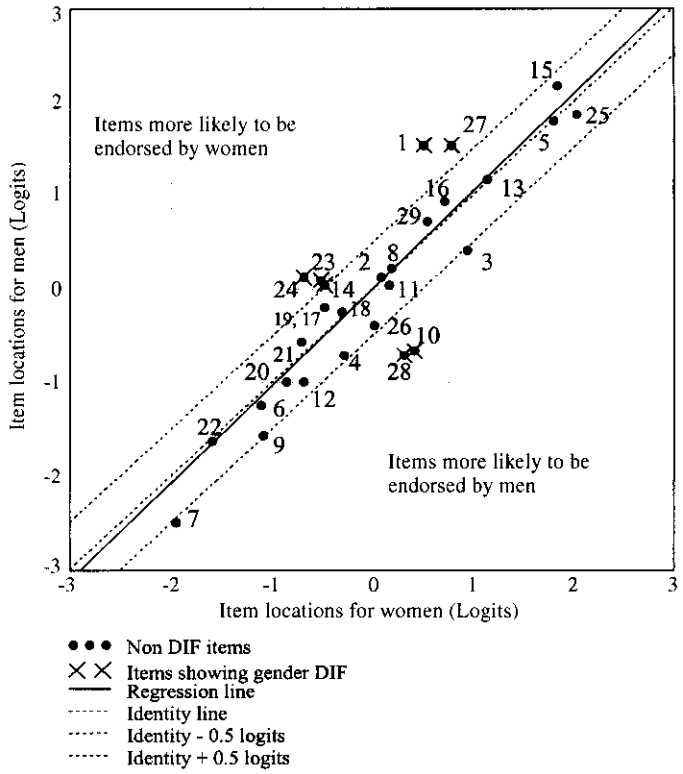

FIG. 2. Item locations for men vs item locations for women before item purification.

RESULTS

Preliminary Analyses

Statistical tests for the presence of age and gender related DIF were obtained via SIBTEST (Stout & Roussos, 1996).[4] Also, the results are presented visually in Figs. 2 and 3 by plotting the item locations in one group against those in another. For instance, as is indicated by the crossed entries in Fig. 2, seven items show statistically significant gender DIF (all $p < .05$). Specifically, the β values for items 1, 14, 23, 24, and 27 are all significantly positive ($\beta_1 = .17$, $\beta_{14} = .14$, $\beta_{23} = .13$, $\beta_{24} = .16$, and $\beta_{27} = .11$), indicating that women are more likely to endorse these items than comparable men (i.e., men who endorsed a similar number of items). By contrast, men are more likely to endorse items 10 and 28 than comparable women, resulting in negative β-values ($\beta_{10} = -.21$ and $\beta_{28} = -.19$). Additionally, the crossed entries

[4] Since no a priori DIF hypotheses were tested concerning the specific subgroups of respondents, a pooled variance approach was used in all SIBTEST runs.

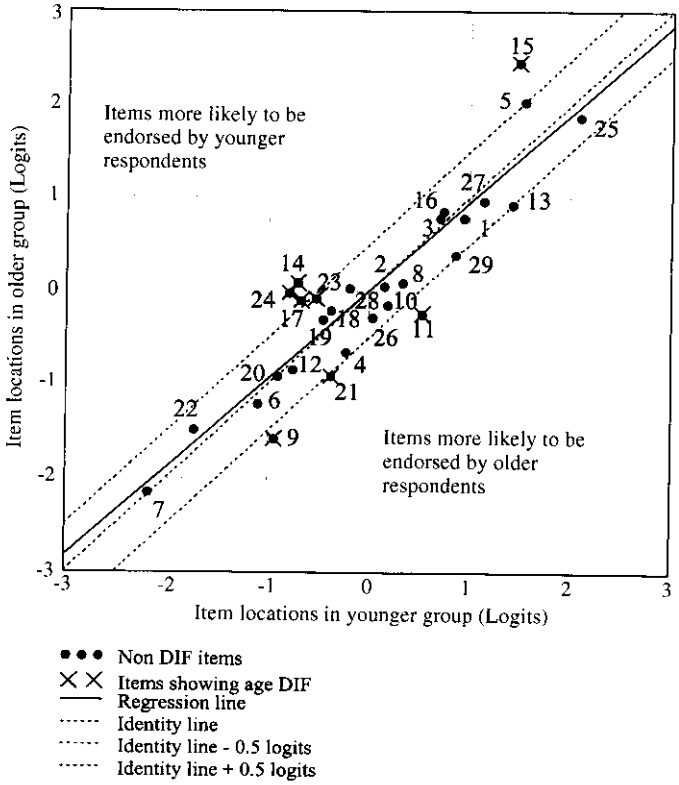

FIG. 3. Item locations for older respondents vs item locations for older respondents before item purification.

in Fig. 3 show that eight items show significant age DIF ($p < .05$). In particular, the negative weights $\beta_9 = -.12$, $\beta_{11} = -.15$, and $\beta_{21} = -.10$ imply that younger respondents are more likely to endorse items 9, 11, and 21 than comparable older respondents. By contrast, older respondents are more likely to endorse items 14, 15, 17, 23, and 24 than comparable younger respondents as the β-weights for these items (.19, .14, .13, .12, and .16, respectively) are significantly positive.

Unfortunately, even when all marked items in Figs. 2 and 3 are removed, DIF remains because the β-values of some other items now become statistically significant. For this reason we postpone a substantive interpretation of the DIF findings until a later section.

As the classic test reliability is quite high (KR-20 = 0.89), we realize that the scale defined by the 29 items may appear acceptable from a classic test theory perspective, especially given recent proposals to make such reliabilities the touchstone by which to compare tests' performances across samples (Thompson & Vacha-Haase, 2000). Doing so would be quite misleading, however, as Table 1 reveals serious problems with the 29-item formulation from a Rasch perspective. In particular, some of the items' infit and outfit statistics fall outside the acceptable range of 0.7 to 1.3. This,

together with the statistically significant age and gender DIF, implies that the 29 items do not define a scale with satisfactory properties of measurement.

The "Purified" Scale

The preceding sections imply that at least some of the items showing age or gender DIF will have to be eliminated to obtain an unbiased measure of transliminality. As pointed out earlier, this cannot simply be done by removing the biased items, as this is likely to introduce DIF or misfit in the remaining items. Also, it is neither possible nor desirable to consider all possible ways of dividing a given set of items into biased and unbiased subsets. Experience (Lange et al., 2000) indicates, however, that the search for non biased subsets can be speeded up by an iterative top-down "purification" approach (the term originates in Lord, 1980) in which biased items with the poorest Rasch fit are removed first. This procedure is repeated until an unbiased and scaleable subset of items is identified. Rejected items are then tentatively reintroduced to check whether the final selection remains stable.

Table 2 shows that this procedure excluded 12 items, leaving a subset of 17 purified transliminality items with an overall Rasch reliability of 0.82. We note that all items show acceptable infit (Column 3, range: 0.88–1.28) and outfit values (Column 4, range: 0.82–1.29), indicating that the items indeed define a Rasch measure. The contents of Column 2 are discussed later. Additionally, Fig. 4 shows that the item locations δ (as indicated by the numbers around $Y = 0$) cluster near the middle of the dimension. Accordingly, the measurement error SE_θ associated with respondents' transliminality estimates (θ) increases rapidly with extremity (solid line in top part of Figure 4). (For purposes of presentation the standard errors of measurement were standardized based on S_θ, i.e., the SE_θ represent z scores). Consistent with the above, the local reliability R_θ (as computed via Eq. (2), see dotted line in Fig. 4) reaches its maximum (0.90) near the middle of the dimension. However, the R_θ decrease with extremity, reaching a minimum of 0 at either extreme of the dimension (the next lowest local reliabilities are about 0.62). Thus, the very tails of the person distribution (which together contain about 4% of the cases) are not reliably measured. Finally, as is indicated by the horizontal dotted line, the overall score reliability (KR-20) as computed within the framework of classic test theory is 0.85. In other words, in addition to ignoring bias and misfit, the classic approach underestimates the reliability at intermediate levels, while grossly overestimating the reliability of extreme scores.

Differential Test Functioning. By construction, none of the 17 items show age and gender DIF. In particular, Columns 5 and 6 in Table 2 indicate that none of the 34 β-values are significant at $p < .05$; in fact, 29 of the 34 DIF tests have Type I error levels in excess of 0.10. Presumably, the removal of biased items should yield a measure without DTF as well. To verify this assumption, four separate Bigsteps runs were performed to determine the sum score to Rasch transformations, as well as the SE_θ, in the four (pairwise) age and gender groups (all in logits). Figure 5 plots the person measures in logits against the sum of the 17 purified items for each group, together with the person measure derived over the entire sample (circles). It can be seen that the four lines virtually coincide with each other as well as with the transformation derived over the entire sample. In other words, there is no evidence of any

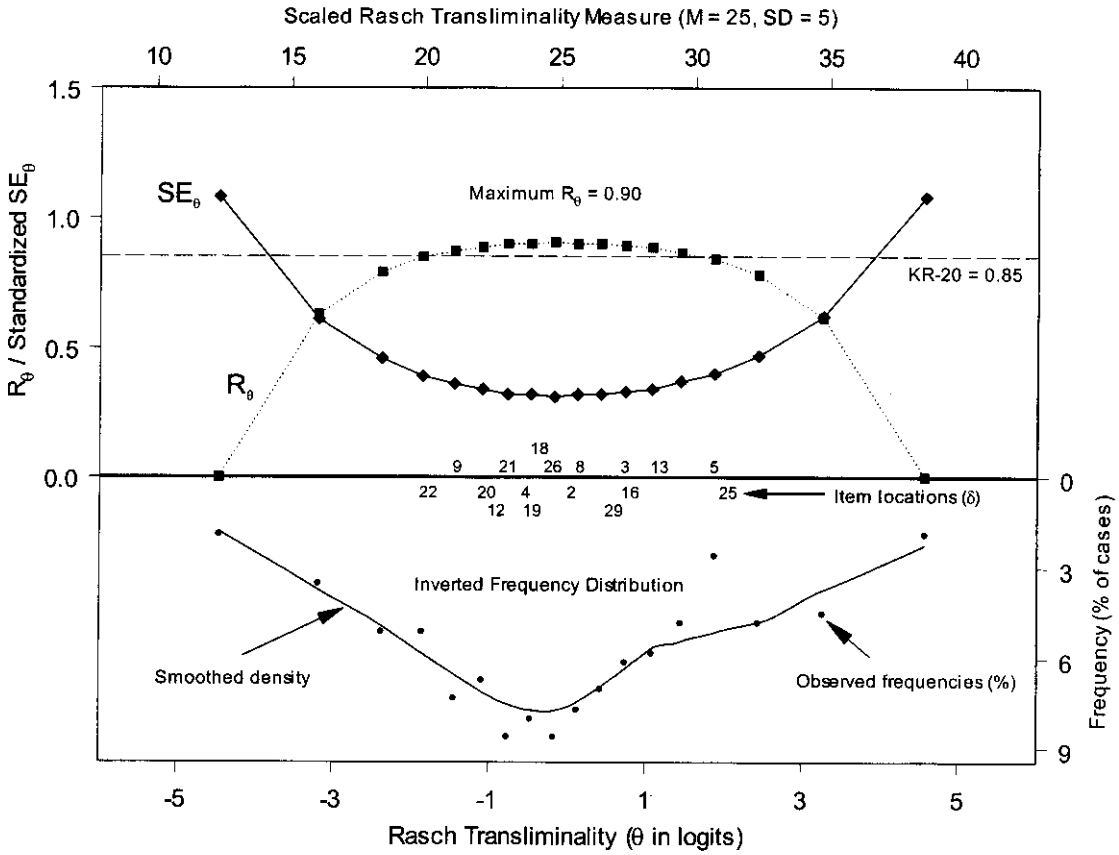

FIG. 4. Map of person and item locations, including normalized standard error of estimates (SE_θ) and local reliabilities (R_θ) (top) and inverted frequency distribution (bottom).

DTF due to age or gender. We further note that the SE_θ in the four groups are indistinguishable as well (not shown).

The preceding implies that the 17 items indeed define a useful measure of transliminality. For cosmetic reasons, it is desirable to avoid negative values and decimals. Therefore, the person logit measures were transformed to have a mean of 25 and a standard deviation of 5. The result is shown in Table 3, which lists the raw score to Rasch transliminality translation together with their SE_θ.

The "Meaning" of Transliminality. As pointed out earlier, the hierarchy implied by item locations defines the meaning of the transliminality construct. These locations are shown graphically in Fig. 6 by the tickmarks on the middle line, and their exact values are also listed in the second column of Table 2 (the standard error of each location is about 0.44 scaled units).

This figure reveals that low to intermediate levels of transliminality are characterized by ambiguous perceptions (Items 22 and 8) and evidence of fantasy proneness (Items 20, 4, and 29) often involving paranormal (Items 21, 2, and 8) or "New Age"

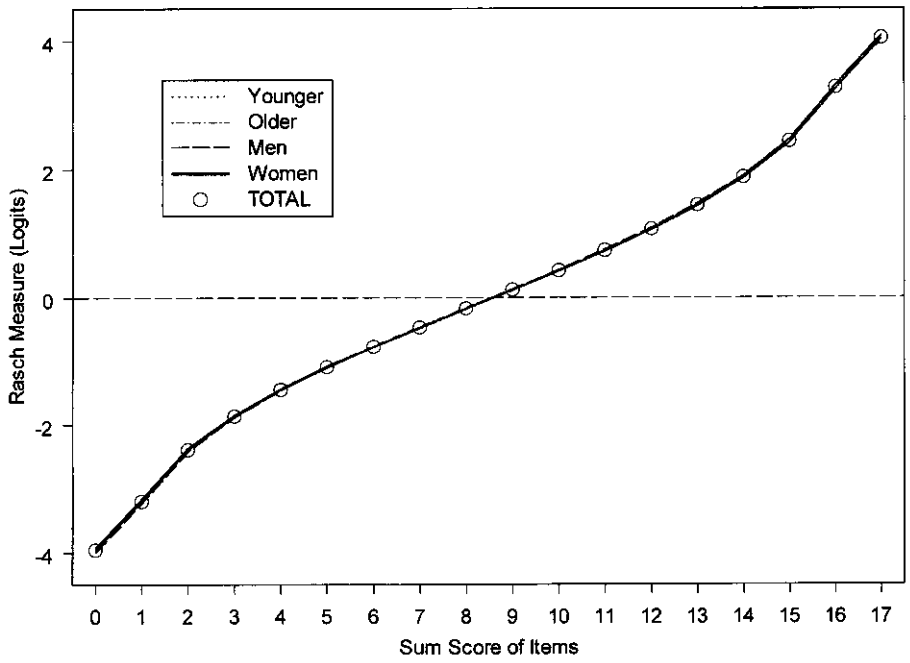

FIG. 5. Sum score to Rasch transformations for the purified measure for the total sample, younger vs older respondents, and men vs women separately.

(Items 12, 18, 26, and 3) themes (cf. Lange et al., 2000). While highly transliminal individuals show similar, but stronger, forms of these experiences (Items 3, 16, and especially 5), they also report very concrete sensory experiences such as being overwhelmed by smells (Item 19), being bothered by bright lights (Item 13), and an inability to shut out a heightened awareness of sights and sounds (Item 25). The occurrence of these items in the final scale is not accidental, as no similar items were excluded from the final version of the transliminality scale (see next section). These findings thus support the notion that transliminality manifests itself not only in people's subjective beliefs, but that it affects the thresholds of the perceptual system's interaction with the outside world as well (James, 1902/1982; Thalbourne et al., 1997). Rasch scaling adds the fact that the most suggestible sensory experience (smell) occurs at much lower levels of transliminality than do sights and sounds ($p < .001$).

As pointed out in the introduction, the item locations δ are defined such that an item j has a probability of 50% of being endorsed by respondents with transliminality level $\theta = \delta_j$. The "ruler" on the left-hand side of Fig. 6 shows how this probability changes as θ moves away from δ_j. The midpoint of this ruler (δ) can be superimposed on any of the marks on the middle line. Thus, while a person with $\theta = 25$ endorses Item 26 with a probability of about 50%, $\theta = 31$ is required to raise this probability to about 90%. Conversely, it can be inferred that someone with $\theta = 25$ will endorse Item 25 with a probability of about 10%. Finally, we note that Fig. 6 reveals a "gap" near the high end of the transliminality dimension. Accordingly, those interested in

TABLE 3
Raw Sum Score to Rasch Conversion Table for the Revised Transliminality Scale (M = 25, SD = 5) and the Local Standard Errors of Measurement

No. of endorsed items	Rasch transliminality measure[a]	Rasch standard error of estimate
0[b]	13.7	7.2
1	15.9	5.3
2	18.3	3.9
3	19.9	3.3
4	21.1	3.0
5	22.1	2.9
6	23.1	2.8
7	24.0	2.7
8	24.9	2.7
9	25.7	2.7
10	26.6	2.7
11	27.5	2.8
12	28.5	2.9
13	29.6	3.1
14	30.9	3.4
15	32.5	4.0
16	35.0	5.3
17[b]	37.3	7.3

[a] This transliminality measure was obtained by estimating the person measures (θ) in logits based on the item locations shown (δ_j) in Table 2 via Bigsteps software and applying the linear transformation $2.90 \times \theta + 25.30$. The values can be rounded to the nearest integer without introducing a noticeable loss of precision. To obtain θ in the presence of missing data, apply the algorithm in Wright and Stone (1979, pp. 142–144) using the δ_j in Table 2 or anchor the items at these locations before running Bigsteps. As none of the items is biased, θ estimates based on subsets of items can be assumed to be unbiased also.
[b] The transliminality measures in these rows, as well as their standard errors, are extrapolated values.

perfecting the scale might profitably focus on the construction of items indicative of high levels of transliminality (and especially items that focus on sensory experiences). Additional guidelines are provided in the following section.

Interpreting DIF. To facilitate interpreting the nature of the DIF effects obtained for the 12 excluded items (see Table 4), additional SIBTEST runs by gender and age were performed which contrasted each excluded item against the 17 items in the purified scale. The β-values obtained are listed in Columns 1 and 2 of Table 4, and the rows are organized according to the nature of the DIF exhibited by the items (i.e., gender DIF only, age DIF only, both age and gender DIF, and neither). As is indicated in Column 1, items 1, 10, 14, 23, 24, 27, and 28 show gender related DIF

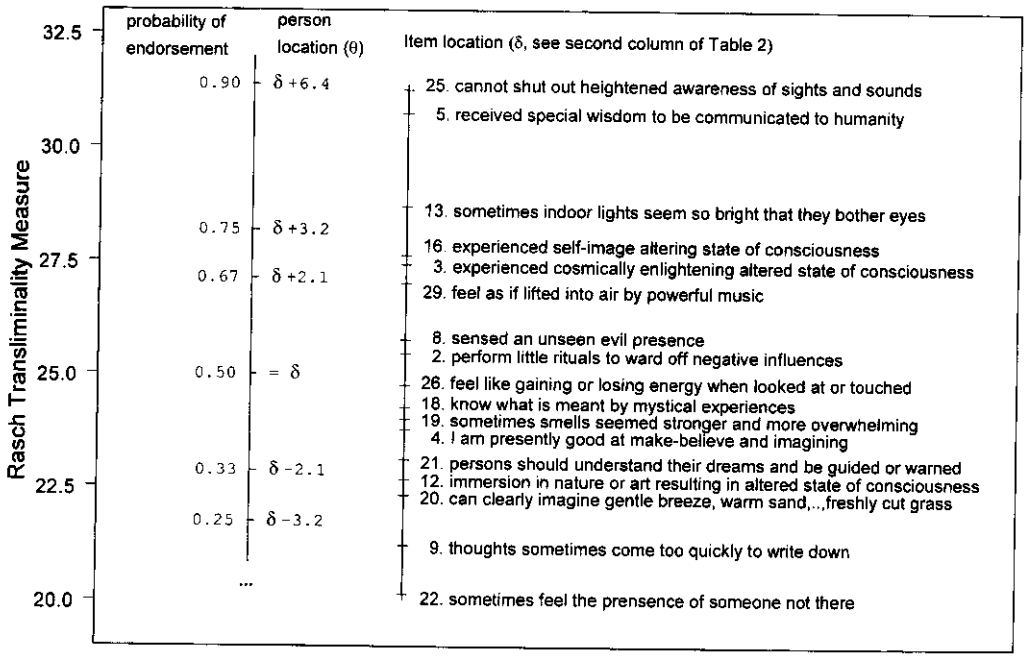

FIG. 6. Defining the meaning of the transliminality construct via the item map.

that revolves around the notion of internal vs external control. For instance, relative to women showing similar transliminality, men are more likely to endorse statements stressing internal control (Item 10 refers to being in control of one's individuality, and Item 28 refers to mastery of one self-image). By contrast, relative to comparable men, women are more likely to endorse statements related to external factors such as "horoscopes" (Item 1), "felt presence" (Item 23), mythical entities (Item 27), "telepathy" (Item 14), and anomalous perception (Item 24). The analysis of age-related DIF suggests that older respondents have a greater belief in paranormal issues involving "being psychic" (Item 15), "premonitions" (Item 17), "telepathy" (Item 14), and anomalous perception (Item 24). Conversely, relative to comparable older respondents, the younger ones appear to experience reality more strongly (Item 11).

We note that the items showing age or gender DIF tend to address issues that are rather more specific than those mentioned in the text of the purified items. For instance, the entries in Table 2 refer to generalized feelings (Items 2, 4, 5, 20, 22, and 26) and sensations (Items 8, 16, 18, 19, and 25) as well as to a variety of abstract concepts. By contrast, Table 4 names specific paranormal mechanisms, including horoscopes, elves, witches, fairies, being psychic, and telepathy (Items 1, 14, 15, 27), in addition to indicators of mania (Items 6 and 28) or possible psychopathology (Items 11 and 23). We hypothesize therefore that while men and women, as well as older and younger respondents, recognize and express abstract signs of transliminality in a similar fashion, DIF occurs since their modes of expression differ with respect to specific instances. This interpretation assumes that DIF was the main source of item

TABLE 4
Significance Tests of β-Values in DIF Tests by Gender and Age for the 12 Excluded Items

Items	Gender β^a	Age β^b
Gender DIF		
1. Horoscopes are right too often for it to be coincidence	0.18***	−0.04
10. If I could not pretend or make-believe anymore, I wouldn't be me—I wouldn't be the same person	−0.17**	−0.06
23. At times I somehow feel the presence of someone who is not physically there	0.13*	−0.10
27. Now that I am grown up, I still in some ways believe in such beings as elves, witches, leprechauns, fairies, etc.	0.11*	−0.04
28. Sometimes people think I'm a bit weird because my ideas are so novel	−0.17**	0.05
Age DIF		
11. Sometimes I experience things as if they were doubly real	−0.02	−0.11*
15. I am convinced that I am psychic	0.04	0.12**
17. I am convinced that I have had a premonition about the future that came true and which (I believe) was not just a coincidence	0.06	0.13*
Both Gender and Age DIF		
14. I am convinced that I have had at least one experience of telepathy between myself and another person	0.13*	0.16**
24. I am convinced that it is possible to gain information about the thoughts, feelings or circumstances of another person, in a way that does not depend on rational prediction or normal sensory channels	0.17**	0.16**
Neither Gender nor Age DIF		
6. I have sometimes behaved in a much more impulsive or uninhibited way than is usual for me	−0.02	−0.02
7. I am fascinated by new ideas, whether or not they have a practical value	−0.03	−0.00

[a] A positive value indicates that women's likelihood of endorsing the item is greater than that of comparable men. A negative β-value indicates the opposite effect.

[b] A positive value indicates that younger respondent's likelihood of endorsing the item is greater than that of comparable older ones. A negative β-value indicates the opposite effect.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

misfit, but that it did not simultaneously also significantly increase the dimensionality of the total item set. As is shown in a later section, this proved indeed to be the case.

It may seem surprising that Items 6 and 7 were excluded since neither item showed gender DIF or age-related DIF relative to the 17 purified items. However, whereas these items show no DIF in isolation, their reintroduction into the purified scales proved counterproductive since doing so causes DIF or misfit in at least two of the purified items listed in Table 2. We note that SIBTEST almost exclusively flagged items that lie outside the range defined by the top and bottom dotted lines in Figs. 2 and 3. These lines correspond to item location differences of ±0.5 logits between the various gender and age groups. Except for Items 6 and 7, these SIBTEST results

thus support Wright and Douglas' (1975, pp. 35–39) rule of thumb that only location differences in excess of 0.5 logits are a cause for serious concern.

Ancillary Analyses

While the purified transliminality test is considerably shorter than the original 29-item version, the reliability of the scale suffered little. For instance, Table 1 indicates that the overall Rasch reliability of the original 29-item scale is 0.88, whereas the reliability of the 17-item version is still 0.82. This small decrease strongly suggests that the purification process succeeded in eliminating those items that contributed disproportionately to the (systematic) errors of measurement as associated with our respondents' age and gender. A number of questions remain, however. For example, is the purified scale is indeed unidimensional? If so, does it measure the same latent variable as the original 29-item version? Additionally, we will investigate how the measures derived from the two versions differ when age and gender are taken into account.

Dimensionality. The unidimensionality of the purified scale can be tested by treating all 17 items as a single Rasch factor and determining its fit relative to a suitable multidimensional Rasch model. As suggested by one of the reviewers, a suitable choice for this purpose is to define the items with "lowest" item locations (i.e., 4, 9, 12, 18, 19, 20, 21, and 22) as one factor and to assign those with the "highest" locations (i.e., 2, 3, 5, 8, 13, 16, 25, 26, and 29) to a second factor. (The items' locations are listed in Column 1 of Table 2). Both the one-factor model and the two-factor model as defined above were fitted using the ConQuest software (Wu et al., 1998). Although the two-factor formulation proved superior to the one-factor formulation [$\chi^2(3) = 8.62$, $p < .05$], the "low" and the "high" factors in the two-factor model showed an extremely high correlation ($r = 0.91$). Accordingly, we conclude that the items are essentially unidimensional and that a single factor model captures a sufficient portion of the latent variable to remain viable (for a discussion of related issues, see, e.g., Hattie, 1985; Linacre, 1998a; Nandakumar & Stout, 1993).

Next, to determine whether the purified scale addresses a materially different latent dimension than the original 29 items, we treated the 12 rejected items (i.e., 1, 6, 7, 10, 11, 14, 15, 17, 23, 24, 27, and 28) and the 17 purified items as two separate Rasch factors. Perhaps not surprising given the DIF in the rejected items (cf. Stout, 1990; Lange et al., 2000), this two-factor model showed a superior fit [$\chi^2(3) = 21.23$, $p < .001$] relative to a model that treated all 29 items as a single factor. However, the correlation between the two factors was extremely high ($r = 0.94$). Thus, again there is little reason to adopt a two-factor model. We conclude therefore that the purified 17-item scale addresses approximately the same latent dimension as the original 29 items.

DTF. Since over 40% of the items were excluded mainly due to DIF, we expected that the Rasch measures produced by the original 29-item version of the transliminality scale should show bias at the test level (DTF) relative to the unbiased 17-item version. To test this prediction, we performed a Gender (Men vs Women) × Age (Young vs Old) × Test Form (29 vs 17 Items) analysis of variance with repeated measures over the last factor. To arrive at commensurate measures, both Rasch measures were transformed to have $M = 25$ and $SD = 5$.

As expected, we found a significant Gender × Test Form interaction effect [$F(1, 311) = 6.13, p < .05$]. This interaction reflects that men score about 0.2 *SD* lower on the original 29-item test than women [$M = 24.5$ vs 25.4; $t(313) = -1.71, p < .05$], whereas the 17 item version shows no such gender effect [$M = 24.7$ vs 25.3; $t(313) = -1.05, p > .10$]. In other words, the removal of DIF produced a net decrease in the gender effect. Additionally, transliminality decreased with age, and this decrease is stronger for the 17-item form ($M_{\text{Young}} = 25.3$ vs $M_{\text{Old}} = 24.6$) than for the 29-item form ($M_{\text{Young}} = 25.1$ vs $M_{\text{Old}} = 24.9$). Whereas none of the pairwise age comparisons are significant (all $p > .10$), the Age × Test Form interaction is highly significant [$F(1, 311) = 7.59, p < .01$]. Thus, in contrast to gender, the removal of age-related DIF actually slightly *accentuated* the average transliminality differences between the two levels of this independent variable. None of the other effects in the ANOVA reached statistical significance (all $p > .17$).

Relation to Previous Research

Recently, Storm and Thalbourne (1998–1999) examined the relationship between transliminality and the Cattell 16PF (Russell & Karol, 1994), and aspects of the results of this study were reanalyzed to determine the effects of the purification process. In particular, the first column in Table 5 shows the correlations with the original 29-item transliminality scale, and the correlations for the purified 17-item version are shown in the second column. Since the 16PF does not yield interval measures, the results are reported as Spearman rank (rho) correlations (cf. Linacre, 1998b).

As is shown in Table 5, the rank correlations of the two transliminality measures with each of the 21 variables are highly similar, and each variable that correlates significantly with the 29-item version also shows a significant correlation with the 17-item version. While the correlation coefficients vary somewhat, only three pairwise differences reach statistical significance ($p < .05$). Thus, although over 40% of the items were removed from the original set of items, most correlations are unaffected and we conclude therefore that the shorter 17-item version yields essentially equivalent results. Undoubtedly, the similar patterns of correlations are partly attributable to the relatively low magnitude of the correlations in Table 5. However, if future research should reveal the existence of clear-cut differences in the performance of the two scale versions, we suggest that the results for the 17-item scale are preferable because they are free of age and gender DIF, whereas the original scale yields slightly biased results at an ordinal level of measurement.

Predictive Validity

As a second step toward testing the validity of the Transliminality Scale, we investigated how it related to dreaming experiences. Specifically, we replicated a study by Hicks, Bautista, and Hicks (1999) in which these authors administered Hartmann's (1991) Boundary Questionnaire and the Spadafora and Hunt (1990) Dream Scale. This dream scale purportedly measures the level of experience with seven dream types, i.e., *lucid dreams* (''vivid dreams in which you realize that you are dreaming while you are still in the dream, and you may then find that you can control the dream while it continues''), *archetypal dreams* (''dreams that carry a sense of awe

TABLE 5
Rank Correlations (Rho) in Previous Research Using Both Versions of the Transliminality Scale (N = 93)

	Test version		
Cattell's 16PF	29-item	17-item	29 vs 17
A (warmth)	0.22*	0.20*	
B (reasoning)	−0.08	−0.02	+
C (emotional stability)	−0.09	−0.05	
E (dominance)	0.02	0.02	
F (liveliness)	0.06	0.03	
G (rule-consciousness)	−0.24*	−0.20*	
H (social boldness)	0.06	0.06	
I (sensitivity)	0.10	0.10	
L (vigilance)	0.03	−0.02	
M (abstractedness)	0.40*	0.36*	+
N (privateness)	−0.02	−0.00	
O (apprehension)	0.15	0.13	
Q1 (openness to change)	0.19*	0.22*	
Q2 (self reliance)	0.15	0.18	
Q3 (perfectionism)	−0.09	−0.09	
Q4 (tension)	−0.05	−0.06	
EX (extraversion)	0.04	0.01	
AX (anxiety)	0.01	−0.03	
TM (tough-mindedness)	−0.29*	−0.32*	
IN (independence)	0.09	0.09	
SC (self-control)	−0.27*	−0.21*	+

* $p < .05$ (correlation with row variable).
+ $p < .05$ (difference between correlations using Bonferoni correction).

and fascination and/or include encounters with strange and unusual beings, perhaps reminiscent of mythology and fairy tales''), *fantastic nightmares* (''very vivid, upsetting dreams that you remember in detail upon awakening and can involve a wide range of negative emotions''), *prelucid dreams* (''where one questions whether one is dreaming but cannot decide''), *control dreams* (''where control not possible in waking life is deliberately exercised in the dream, with or without lucid awareness of dreaming''), *posttraumatic nightmares* (''dreams repeating an actual past trauma''), and *night terrors* (''terrifying awakenings without any recall of dream content'').

A Likert-style response format is used with five categories that are labeled *''Never (or almost never),'' ''Seldom,'' ''Occasionally,'' ''Frequently,''* and *''Always (or almost always).''* These categories are assigned the values 1 through 5, respectively, and the rating for each type of dream is treated as a separate variable. Consistent with their predictions, Hicks et al. (1999) found that those with ''thin'' boundaries reported significantly greater frequency of dreaming than those with ''thick'' boundaries as determined by Hartmann's Boundary Questionnaire. As pointed out in the introduction, the transliminality concept resembles Hartmann's notion of boundaries.

TABLE 6
Rank (Rho) and Partial Correlations (Removing Age and Gender) between Seven Dream Types Identified by Spadafora and Hunt (1990) and the Revised Transliminality Scale (N = 57)

Type of dream	Rho	Partial correlation
Archetypal	0.39**	0.38**
Control	0.43**	0.39**
Fantastic nightmare	0.24*	0.24*
Posttraumatic nightmare	0.48**	0.38**
Lucid	0.42**	0.55**
Prelucid	0.35**	0.34**
Night terror	0.46**	0.32**

* $p < .05$.
** $p < .01$ (one-tailed).

Hence, we predicted that our 17-item transliminality measure should show similar correlations with Spadafora and Hunt's Dream Scale.

Both questionnaires were administered in counterbalanced order to a convenience sample of 57 participants (M_{age} = 39.2 yrs., SD = 13.8, range = 21–72 years, 61% women). Table 6 shows the Spearman rank order correlations between the transliminality scale and the seven types of dreams. As predicted, the transliminality measure showed positive associations with all seven dream types and all correlations were significant at $p < .05$. Although our results conceptually replicate those of Hicks et al. (1999), it would be premature to discuss the findings in greater detail. However, it seems unlikely that the correlations are due to any confounding effects of age and gender because their magnitudes remain largely the same when age and gender are partialed out (based on the standard product moment correlation, see Table 6). As such, the findings add to the accumulating evidence that transliminality is a useful predictor of certain mental phenomena.

Caveats

We note that the variables listed in the rows of Tables 5 and 6 did not include any DIF analyses. As it seems likely that at least some of these are biased, the correlations in Tables 5 and 6 may be erroneous to some degree. The extent of such biases cannot be ascertained, however, without first purifying all measures involved.

DISCUSSION

We opened this article by quoting William James. Perhaps then it would be appropriate to close it with a final quote from him, looking at the subliminal aspects of transliminality:

> If the word 'subliminal' is offensive to any of you . . . call it by any other name you please, to distinguish it from the level of full sunlit consciousness. Call this latter the A-region of personality, if you care to, and call the other the B-region. The B-region, then, is obviously

TABLE 7
Rank Correlations (Rho) between Transliminaty and Selected Mental Phenomena Discussed by William James[a]

Variable	Rho	Reference
Three-item dream recall scale	0.17*	Thalbourne and Delin (1999)
Lucid dreaming (Australians)	0.39**	Thalbourne and Houran (2000), further analysis
Lucid dreaming (Americans)	0.32**	Thalbourne and Houran (2000), further analysis
Hood's Mysticism Scale	0.66**	Thalbourne and Delin (1999)
Mystical Experience Ratings	0.51**	Thalbourne and Delin (1999)
Absorption (corrected)	0.72**	Thalbourne (1998)
Paranormal experiences (Australians)	0.80**	Thalbourne and Houran (2000)
Paranormal experiences (Americans)	0.76**	Thalbourne and Houran (2000)
Religiosity	0.48**	Thalbourne and Delin (1999)

[a] Pearson correlations were reported in the original sources.
* $p = .06$ (two-tailed).
** $p < .01$.

> the larger part of each of us, for it is the abode of everything that is latent and the reservoir of everything that passes unrecorded or unobserved. It contains, for example, such things as all our momentarily inactive memories, and it harbors the springs of all our obscurely motivated passions, impulses, likes, dislikes, and prejudices. Our intuitions, hypotheses, fancies, superstitions, persuasions, convictions, and in general all our non-rational operations come from it. It is the source of our dreams, and apparently they may return to it. In it arise whatever mystical experiences we may have, and our automatisms, sensory or motor; our life in hypnotic and 'hypnoid' conditions, if we are subjects to such conditions; our delusions, fixed ideas, and hysterical accidents, if we are hysterical subjects; our supra-normal cognitions, if such there be, and if we are telepathic subjects. It is also the fountainhead of much that feeds our religion. (James, 1902/1982, pp. 483–484)

Although we have made only a beginning at this stage, we can say that transliminality correlates with at least some of the phenomena James mentioned above (see Table 7) in addition to those already contained in Tables 5 and 6. In addition, the Rasch methodology allowed us to measure the transliminality construct in a superior fashion. Specifically, top-down purification of Thalbourne's (1998) original 29-item transliminality scale yielded a 17-item subset that we now call the Revised Transliminality Scale (RTS). This subset consists of items that span magical ideation, mystical experience, absorption, hyperaesthesia, manic experience, dream interpretation, and fantasy proneness, but it is important to note that these items in fact form a unidimensional hierarchy. In other words, as argued by James (1902/1982) and later by Thalbourne (1991; Thalbourne & Delin, 1994), these groups of cognitive and emotional phenomena share a common core. This conclusion cannot be attributed to obvious confounding variables, as the RTS contains no detectable age or gender bias, while providing measures at an interval level with acceptable reliability and scaling properties. We further note that the hierarchical nature of the RTS supports the notion that transliminality entails a number of thresholds.

Initial support for the construct validity of the RTS is provided by the reanalysis of Storm and Thalbourne's (1998–1999) data as many of the personality characteristics measured by the 16PF, as well as those included in our dream study, showed associa-

tions with transliminality that support our theory of transliminality. For example, we define transliminality as the tendency for psychological material to cross thresholds into or out of conscious awareness. Inherent to this definition is the assumption that high transliminals possess a greater degree of connection among a broad array of mental processes than do low transliminals. Assuming that these connections are physiologically based, transliminality would seem to facilitate the spontaneous elicitation of emotions, cognitions, and behavior due to a lack of inhibition or separateness among mental processes. We further predict that transliminality should show positive associations with abstractedness, openness to experience (confirmed by Thalbourne, 2000), and frequency of dreaming, while negative associations with tough-mindedness and self-control are to be expected. In this context we note that the original 29-item Transliminality Scale also addressed topics related to control, anomalous perception, and intuitive modes of knowing. Although these items were eliminated due to age and/or gender DIF, they did not form a second, distinct Rasch factor. We hypothesize that these items showed bias because they addressed specific issues that were interpreted in an age- and gender-dependent fashion. If correct, this implies that researchers interested in extending the RTS should formulate items that are general and abstract and avoid formulations that explicitly name particular indicators of transliminality.[5]

Clearly, many of the issues touched upon in the preceding paragraph require further research, including laboratory investigations concerning the relation between transliminality and sensory processes as well as questionnaire-based research on empathy, intuition, and so on. With respect to the latter, we note that top-down purification provides a powerful new tool for the development and revision of other instruments in various areas of assessment (cf. Lange et al., 2000; Lange, Thalbourne, Houran, & Lester, 2000). Its specific advantages are that research findings based on purified instruments cannot be attributed to item or test bias; moreover, Rasch scaling ensures the scalability of items and respondents alike, while providing error estimates that are more realistic than those obtained within a classic test theory framework. These psychometric properties are especially important in applications that make strong data assumptions, including structural modeling (e.g., Lange & Houran, 1998, 1999b) and nonlinear dynamics (Lange & Houran, 2000). Additionally, it has been shown (Lange et al., 2000) that item bias affected the results obtained from familiar statistical techniques such as factor analysis and correlational analysis. Accordingly, the available research on top-down purification indicates that many established findings in psychology and psychiatry that are based on questionnaires may need to be revisited.

ACKNOWLEDGMENTS

We thank Dr. William Banks and two anonymous reviewers for their helpful comments on an earlier draft of this article. M. Thalbourne and L. Storm participated in this research through a grant from the Bial Foundation (Portugal).

[5] The context in which items occur sometimes affects their interpretation (Mellenbergh, 1983, p. 294). Therefore, administering the 17-item scale without the 12 rejected items might introduce DIF or DTF. The presence of such effects cannot be determined without first administering the 17-item scale (i.e., without the rejected 12 items) to a new sample of respondents.

REFERENCES

Ackermann, T. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional IRT perspective. *Journal of Educational Measurement,* **29,** 67–91.

Ackermann, T. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement,* **20,** 311–329.

Baars, B. J. (1988). *A cognitive theory of consciousness.* New York: Cambridge Univ. Press.

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease,* **174,** 727–734.

Boring, E. G. (1963). *History, psychology, and science: Selected papers.* New York: Wiley.

Byrne, D., Barry, J., & Nelson, D. (1963). Relation of the revised repression-sensitization scale to measures of self-description. *Psychological Reports,* **13,** 323–334.

Claridge, G., & Broks, P. (1984). Schizotypy and hemisphere function. I. Theoretical considerations and the measurement of schizotypy. *Personality and Individual Differences,* **8,** 633–648.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice,* **17,** 31–44.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology,* **46,** 648–659.

Daniel, M. H. (1999). Behind the scenes: Using new measurement methods on DAS and KAIT. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 37–64). Mahwah, NJ: Erlbaum.

Embretson, S. E. (1995). The new rules of measurement. *Psychological Assessment,* **8,** 341–349.

Embretson, S. E., & Herschberger, S. (1999). *The new rules of measurement: What every psychologist and educator should know.* Mahwah, NJ: Erlbaum.

Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck personality scales (EPS Adult).* London: Hodder & Stoughton.

Hartmann, E. (1991). *Boundaries in the mind: A new psychology of personality.* New York: Basic Books.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement,* **9,** 139–164.

Herbart, J. F. (1961). Psychology as a science, newly founded upon experience, metaphysics and mathematics. In T. Shipley (Ed.), *Classics in psychology* (pp. 22–50). New York: Philosophical Library. (Original published in 1824).

Hicks, R. A., Baustista, J., & Hicks, G. J., (1999). Boundaries and level of experience with six dream types. *Perceptual and Motor Skills,* **89,** 760–762.

James, W. (1982). *The varieties of religious experience.* Harmondsworth, Middlesex, England: Penguin. (Original published in 1902).

Kumar, V. K., & Pekala, R. (1992). *The Mental Experience Inventory.* Unpublished Psychological Test. West Chester, PA: West Chester Univ.

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. (1999a). Scaling MacDonald's AT-20 using item-response theory. *Personality and Individual Differences,* **26,** 467–475.

Lange, R., & Houran, J. (1999b). 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., 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., Houran, J., & Lester, M. (2000). Depressive response sets due to gender

and culture based differential item functioning: A case study using Thalbourne's Manic Depressiveness Scale. Manuscript submitted for publication.

Launay, G., & Slade, P. D. (1981). The measurement of hallucinatory predisposition in male and female prisoners. *Personality and Individual Differences,* **2,** 221–234.

Linacre, J. M. (1998a). Rank-ordered raw scores imply the Rasch model. *Rasch Measurement Transactions,* **12**(2), 637–638.

Linacre, J. M. (1998b). Do correlations prove scores linear? *Rasch Measurement Transactions,* **12**(1), 605–606.

Linacre, J. M., & Wright, B. D. (1997). *A user's guide to BIGSTEPS rasch-model computer program.* Chicago, IL: MESA Press.

Lord, F. M. (1980). *Application of item response theory to practical problems.* Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison–Wesley.

Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement,* **55,** 967–975.

MacDonald, A. P., Jr. (1970). Revised scale for ambiguity tolerance: Reliability and validity. *Psychological Reports,* **26,** 791–798.

MacKinnon, D. W. (1971). Creativity and transliminal experience. *Journal of Creative Behavior,* **5,** 227–241.

Mellenbergh, G. J. (1983). *Conditional item bias methods.* New York: Plenum.

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Erlbaum.

Myers, F. W. H. (1903). *Human personality and its survival of bodily death.* London: Longmans, Green.

Myers, S. A. (1983). The Wilson-Barber Inventory of Childhood Memories and Imaginings: Children's form and norms for 1337 children and adolescents. *Journal of Mental Imagery,* **7,** 83–94.

Nandakumar, R., & Stout, W. F. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics,* **18,** 41–68.

Panter, A. T., Swygert, K. A., Dahlstrom, G. W., & Tanaka, J. S. (1997). Factor analytic approaches to personality item-level data. *Journal of Personality Assessment,* **68,** 561–589.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: MESA Press.

Raven, J. C. (1965). *Advanced progressive matrices: Plan and use of the scale.* London: H.K. Lewis.

Riley, K. C. (1988). Measurement of dissociation. *Journal of Nervous and Mental Disease,* **176,** 449–450.

Rugg, H. (1963). *Imagination.* New York, Evanston, and London: Harper & Row.

Russell, M. T., & Karol, D. L. (1994). *The 16PF fifth edition administrator's manual.* Champaign, IL: Institute for Personality and Ability Testing, Inc.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/dif from group ability differences and detects test bias/dtf as well as item bias. *Psychometrika,* **58,** 159–194.

Spadafora, A., & Hunt, H. T. (1990). The multiplicity of dreams: Cognitive affective correlates of lucid, archetypal and nightmare dreaming. *Perceptual and Motor Skills,* **71,** 627–644.

Storm, L. C., & Thalbourne, M. A. (1998–1999). The transliminal connection between paranormal effects and personality in an experiment with the *I Ching. European Journal of Parapsychology,* **14,** 100–124.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality and ability estimation. *Psychometrika,* **55,** 293–325.

Stout, W., & Roussos, L. (1996). *SIBTEST manual.* Statistical Laboratory for Educational and Psychological Measurement, Univ. of Illinois at Urbana–Champaign.

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology,* **83,** 268–277.

Thalbourne, M. A. (1991). The psychology of mystical experience. *Exceptional Human Experience,* **9,** 168–186.

Thalbourne, M. A. (1996). An attempt to predict precognition scores using transliminality-relevant variables. *Journal of the Society for Psychical Research,* **61,** 129–140.

Thalbourn, M. A. (2000). Relation between transliminality and openness to experience. *Psychological Reports,* **86,** 909–910.

Thalbourne, M. A. (1998). Transliminality: Further correlates and a short measure. *Journal of the American Society for Psychical Research,* **92,** 402–419.

Thalbourne, M. A., Bartemucci, L., Delin, P. S., Fox, B., & Nofi, O. (1997). Transliminality: Its nature and correlates. *Journal of the American Society for Psychical Research,* **91,** 305–331.

Thalbourne, M. A., & Delin, P. S. (1994). A common thread underlying belief in the paranormal, creative personality, mystical experience and psychopathology. *Journal of Parapsychology,* **58,** 3–38.

Thalbourne, M. A., & Delin, P. S. (1995). *Correlates of belief in the paranormal: A partial replication.* Unpublished manuscript, Department of Psychology, Univ. of Adelaide, Adelaide, South Australia.

Thalbourne, M. A., & Delin, P. S. (1999). Transliminality: Its relation to dream-life, religiosity and mystical experience. *International Journal for the Psychology of Religion,* **9,** 45–61.

Thalbourne, M. A., & Houran, J. (2000). Transliminality, the Mental Experience Inventory, and tolerance of ambiguity. *Personality and Individual Differences,* **28,** 853–863.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics *is* datametrics: The test is not reliable. *Educational and Psychological Measurement,* **60,** 174–195.

Torrance, E. P. (1966/1974). *Torrance tests of creative thinking: Directions manual and scoring guide: Verbal test booklet A.* Lexington, MA: Personnel Press.

Usher, F. L., & Burt, F. P. (1909). Thought transference. (Some experiments in long distance thought-transference.) *Annals of Psychical Science,* **8,** 561–600.

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Virtanen, L. (1990). In J. Atkinson & T. Dubois (Transl.), *That must have been ESP!* Bloomington, IN: Indiana Univ. Press.

Waller, N. G., Thompson, J. S., & Wenk, E. (2000). Using IRT to separate measurement bias from true group differences on homogeneous *and* heterogeneous scales: An illustration with the MMPI. *Psychological Methods,* **5,** 125–146.

Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement,* pp. 65–104. Mahwah, NJ: Erlbaum.

Wright, B. D., & Douglas, G. A. (1975). Best test design and self-tailored testing. In *Research Memorandum No. 19.* Chicago: Statistical Laboratory, Department of Education, University of Chicago.

Wright, B. D., & Mok, M. (2000). Understanding Rasch measurement: Rasch models overview. *Journal of Applied Measurement,* **1,** 83–106.

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

Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *ConQuest: Generalized item response modelling software.* Australian Council for Educational Research (ACER).

Received September 3, 1999