Scale validation issues in situations of minimal cultural difference
Heather J. Crawford and Gary D. Gregory – University of New South Wales

Abstract
There is a large and growing body of literature on methodology for ensuring that research conducted in one cultural context can be replicated in a dissimilar context. Most of this work focuses on situations of maximal cultural difference. This study investigates issues pertaining to the replication of research in a very similar cultural context, using the individual difference dimension of Need for Humour as its focus. Results show that even minimal cultural difference can introduce unwelcome complexities if equivalence is not explicitly assured.

Introduction
This paper is an attempt to highlight the issues involved in validating scales in conditions of minimal cultural difference. Most commonly, such processes are undertaken and reported in situations where substantial cultural differences are expected. A small body of research has addressed these issues using a selection of advertising rating scales, but the inclusion of personality differences as moderators in advertising research suggests a need to extend this work.

One individual difference that is posited to moderate the responses to humorous advertising is the need for humour (NFH). This scale was developed by Cline et al. (1997; 2003; 2007) but has yet to be validated for use outside the United States.

Background
Humour is frequently used in advertising, but its effects and effectiveness are not well understood. With over US$114 billion globally being spent annually on ads with humorous appeals (ZenithOptimedia 2008), it is critical that advertisers achieve the highest possible effectiveness. Television advertising remains the largest contributor at almost 38% of adspend globally. In Australia, adspend in 2007 was A$13.2 billion (Lee 2008) with over 41% on television. With a minimum of 24% of all advertising using humorous appeals (Weinberger and Spotts 1989), this equates to almost A$3 billion annually.

One major element in the research into humour is the process by which humour has an effect on an individual. This will be influenced by individual factors such as personality, values, demographics and psychographic factors (Gulas and Weinberger 2006). A growing body of research has been published that investigates individual differences and their impact on responses to advertising and in particular humorous appeals. Individual-level cultural differences (Rustogi et al. 1996), personality factors including need for cognition (NFC) (Geuens and De Pelsmacker 2002; Zhang 1996), need for humour (NFH) (Cline et al. 2003; Cline and Kellaris 2007) and self-monitoring (Lammers 1991), gender (Fugate et al. 2000; Martin et al. 2003) and political ideology have all been shown to influence responses to humorous advertising.
The need for humour (NFH) is an individual difference variable examined by Cline, Machleit and Kellaris (2003; 2007; 1999), based on a subset of the Need for Levity (NFL) scale developed by Cline during his doctoral research (1997). According to Cline et al, need for levity describes “an individual's tendency to engage in (to seek-out) levity” (Cline et al. 1999:155). Levity was seen as comprising both humour (amusement and wit) and whimsy (caprice, spontaneity, and playfulness) (Cline et al. 1999).

The NFL scale was developed with the four dimensions of internal/external whimsy and internal/external humour (Cline et al. 1999). A pool of items was constructed and subjected to both exploratory and confirmatory factor analysis to purify the scale. Convergent validity and reliability were assessed following recommended procedures (Gerbing and Anderson 1988; Nunnally and Bernstein 1994). Discriminant validity was tested using procedures recommended by Dillon and Goldstein (1984).

In later work examining humour in advertising, the scale was revised to include only the two humour dimensions and renamed the need for humour (NFH) scale (Cline et al. 2003). The scale items are shown in Table 1. The investigation included three separate studies and offered some evidence that an individual's need for humour may play a significant role in moderating attitudes. NFH appears to act as a motivator and high NFH individuals respond more favourably to humorous ads. Results also indicated that subjects high in NFH form more favourable attitudes based on humorous ads, and show less favourable attitudes to ads with lower humour content.

Although this scale has been used in a number of studies in a U.S. context, it has not yet been used elsewhere. In order to use the scale to test the potential moderating effects of NFH in other cultural contexts, it is necessary to first undertake cross-cultural validation.

Cultural differences between Australia and the United States are routinely classified as minimal when compared to the level of difference to be expected between other cultural contexts (Hofstede 2001; Schwartz and Bilsky 1990). However, these differences, although small in absolute terms, must still be recognised and taken into account when attempting to replicate U.S. research in other markets (Frazer et al. 2002).

Research Methodology

A sample of 195 undergraduate subjects was drawn from volunteers at a large university in New South Wales, Australia. Complete responses were obtained from all but three of the subjects for a final usable sample of 192. The sampling frame was intended to duplicate that of the original research in the United States (Cline 1997; Cline et al. 2003). Demographic analysis of the sample populations of the two studies did highlight some differences. The sample drawn from the University of Cincinnati in 1997 and 2003 was approximately 87% Ohio residents, with little apparent cultural diversity (University of Cincinnati 2008). The sample in Australia comes from a region where 37% of people are not born in Australia, and more than 50% have at least one parent born in another country (ABS 2007). In addition, the university sources over 13% of its students from countries outside Australia (UWS 2006).

The scale was administered to the participants during a break in a lecture on a strictly voluntary basis. It was explained that they were assisting in research to assess personality differences between consumers of their age group and education level across three countries.
Analysis and Results

Analysis of the data followed an iterative process of exploratory factor analysis (EFA), item purification, qualitative research, further item purification, EFA, confirmatory factor analysis (CFA) and reliability testing.

In order to test the NFH scale (Table 1), it was necessary to use factor analysis to assess if the 6-item subscales each linked to the theorised underlying factors. The Likert scales were treated as interval level data as the wording of response levels clearly implies symmetry of response levels about a middle category. The sample size has a 16:1 ratio of observations to variables which falls well within acceptable limits (Hair et al. 2006:112).

To assess factorability of the scales, examination of the bivariate correlations of variables in each scale showed that 50% were greater than .30 (significant at .01) (Coakes et al. 2008:127), and that the Bartlett’s Test of Sphericity value was significant at the .0001 level. The overall measure of sampling adequacy (MSA) was significant (KMO=.86), and the MSA value of individual items exceeded the threshold value (.70) (Hair et al. 2006). These results suggest that the necessary preconditions for factor analysis exist for these scales. Assumptions of normality and linearity were tested with no serious departures obtained. No overly influential observations were found.

Table 1. Need for Humor scale

Section A INTERNAL
1a. People expect me to say amusing things.
2a. I can crack people up with the things I say.
3a. I often come-up with witty comments.
4a. I am good at thinking-up jokes or funny stories.
5a. People tell me that I am quick-witted.
6a. I often feel the need to make other people laugh.

Section B EXTERNAL
1b. I am a connoisseur of humor.
2b. I prefer situations where people are free to express their senses of humor.
3b. I enjoy being with people who tell jokes or funny stories.
4b. I often read jokes and funny stories.
5b. I enjoy being around quick-witted people.
6b. I need to be with people who have a sense of humor.

Following Cline et al. (2003) we used principal components analysis with varimax rotation to conduct exploratory factor analysis (EFA). The 12 items loaded cleanly on the posited two factors (eigenvalues 4.98 and 1.84) with good reliability (Cronbach α = .866). However, the loading and communality of one item (4b) was below acceptable standards (.497 and .275) so it was removed and the analysis was repeated. Subsequent iterations were required to remove two more items (5b and 6a) which also had unacceptable results (loadings .636 and .697, communalities .445 and .541). Direct oblimin rotation showed almost identical results, so despite the preference for non-orthogonal rotations where factors are correlated, the varimax results are reported to allow direct comparison with the earlier research.
The 9-item solution had strong clear loadings (> .7) and communalities (> .5) for all items. Reliability was only marginally lower than the full 12-item scale (Cronbach $\alpha = .851$). The two factors were labelled Internal (6 items) and External (3 items).

One unexpected result from this EFA was that one item theorised to contribute to the External factor (1b) instead loaded strongly (.785) on the Internal factor. In order to understand this complication, a qualitative study was undertaken with a sample of 30 graduate research students and academics from the School of Marketing at a major Sydney university. The 12-item NFH scale was completed, with comments added to highlight any confusion by the subjects surrounding specific words or scale items.

This suggested that one item (1b) was not easily understood by a number of subjects (27%), a result supported by the fact that missing data from the earlier study occurred only on that item. In addition, concerns were expressed about the Read Jokes item (4b), suggesting the word ‘often’ severely reduced their scores on that item. The term ‘Quick-witted’ (5a and 5b) was also seen as an issue, with interpretations of ‘smart, clever’ which made the item appear out of place in a humour scale.

CFA was then conducted to test the proposed 9-item scale and an 8-item scale with the removal of the problematic ‘Connoisseur’ item (1b). The analysis showed that the two factor model demonstrated excellent fit on a range of goodness-of-fit measures with improvement indicated from the 9-item to the 8-item scale (Table 1). The correlation between the internal and external humour dimensions was positive and significant ($\Phi_8 = .35$, $t = 3.79$, $p < .001$). Hence, for the purposes of the experiment, the dimensions were summed and averaged into a single "global" measure of NFH with $\alpha$ reliability of .82. High and low groups were formed via median split (median = 5.0) and found to differ statistically in terms of NFH scores ($t = 19.71$, $p < .001$).

**Table 2: Goodness of Fit and Reliability values**

<table>
<thead>
<tr>
<th></th>
<th>GFI</th>
<th>AGFI</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA</th>
<th>Cronbach $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-item model</td>
<td>0.847</td>
<td>0.735</td>
<td>0.71</td>
<td>0.791</td>
<td>0.174</td>
<td>.851</td>
</tr>
<tr>
<td>8-item model</td>
<td>0.95</td>
<td>0.905</td>
<td>0.946</td>
<td>0.963</td>
<td>0.075</td>
<td>.821</td>
</tr>
</tbody>
</table>

**Figure 1: 8-item model with path coefficients**
Discussion

This research is important to the literature for three reasons. First, it highlights equivalence issues that must be explored when replicating research in a new context with even minimal cultural difference from the original source. Second, the present study demonstrates the impact that lack of equivalence can have on the results from a scale transferred cross-culturally without appropriate validation. Third, despite issues with specific items, the overall structure of the NFH scale has been supported by the Australian sample. Thus, opportunities abound for further research using this individual difference variable to explore responses to advertising in Australia.

There are both similarities and differences in the results gained in this study compared to those obtained by Cline (1997) and Cline et al. (1999) in the U.S. The EFA clearly indicated that the items loaded on the expected two factors of internal-humour and external-humour. However, a number of items failed to achieve loadings that supported their retention in this model, despite the theoretical considerations that may have suggested their inclusion. Thus a truncated model was developed for which empirical support could be gained from the Australian data.

The qualitative research provided some potential explanation for these differences. Although Australia and the U.S. are seen as culturally very similar (Frazer et al. 2002; Schwartz 1992), lexical and idiomatic equivalence tests (Hui and Triandis 1985) indicate that for these samples at least, there are significant dissimilarities. Respondents reported difficulty in understanding some terms used in the original NFH scale and further probing revealed diverse interpretation of items.

The differences in the sample populations go beyond mere nationality. The multicultural nature of the Australian population compared to the predominantly homogenous sample in the United States appears to have had a significant impact on the responses to the scale items. The age group and education level are comparable, but ethnicity and country of permanent residence vary widely between the two contexts. The number of first- and second-generation immigrants, as well as a sizable proportion of international students, means that the proportion of those with English as a second language is much higher in the Australian sample. Beyond this, familiarity with some of the idiomatic terms used in the original NFH scale may be low, as they are ‘Americanisms’ (eg ‘crack people up’).

Interpretation of the items may also be influenced by cultural context at both individual and national level (Schwartz 1992). Relatively minor differences between U.S. and Australian residents have been reported on values such as Mastery and Affective Autonomy (Schwartz 1994). These values could impact on the subjects’ willingness and desire to alter their social environment and their emphasis on enjoyment of stimulating activity. Consequently, their scores on individual NFH items could be skewed sufficiently to show the results reported above.

The Australia sample was selected to be as comparable to the original U.S. sample as possible. However, due to the extremely multicultural nature of the Australian population, and its possible overrepresentation in the student body at a major Sydney university, the results may not be generalisable to the broader student population in Australia, nor to Australia as a whole.
Future research

From the evidence obtained in this study, it is obvious that the NFH scale needs to be refined and tested further before it can be used in subsequent research in Australia or other more disparate cultural contexts. One method being explored for this scale refinement is the use of Rasch modelling. This methodology is seen as a powerful tool for discriminating between people, better capturing all levels of a construct, highlighting redundant items and providing insight as to why an item is mis-performing. It has been recognised as a strong method for scale development due to the breadth it assigns to the construct and the superior information it can provide the researcher throughout the refinement process (Salzberger 2000).

Construct validity of the scale (convergent, discriminant and nomological) must be assessed using the Gerbing and Anderson (1988) paradigm to ensure that the NFH scale exceeds the minimum requirements (Bagozzi 1984; Bagozzi and Yi 1988). Once the scale is refined to resolve issues that have arisen during its testing in Australia, it has a much greater chance of being useful in non-English speaking contexts. The recommended double-blind back-translation process has a far higher probability of achieving conceptual and translational equivalence (Hui and Triandis 1985) if it is working from a basis of unambiguously worded items.

References


Cline, T. W., 1997. The role of expectancy and relevancy in humorous ad executions: An individual difference perspective, Ph.D., University of Cincinnati.


Salzberger, T., 2000. An alternative way of establishing measurement in marketing research-its implications for scale development and validity in Australia and New Zealand Marketing Academy Conference (ANZMAC). Gold Coast, Queensland, Australia.


Review comments addressed

Reviewer 1649
To prepare the paper for publication in the Conference Proceedings it should be properly formatted, and Table 1 should simply list the items in the scale (cut out the scale numbers) and be properly placed in the text. If the authors could explain why a Varimax rotation was used and make some comment about the American scale validation process, it would add value to the piece. I would really have liked to have seen a table showing the factor loadings, too; I suppose that space precludes this possibility.

Reviewer 1525
I don’t understand why table 2 appears before Table 1 in the paper. On bottom of page 2, you report that Bartlett’s Test was significant at .0001 level and also report the p value. This is tautological.

• Formatting was altered to comply with Australasian Marketing Journal guidelines.
• Table 1 was moved to before Table 2, and the scale numbers removed.
• The original scale validation process was described.
• Direct oblimin rotation, more appropriate for correlated factors, was explained as having almost identical results to the varimax rotation. The reporting of the varimax rotation was explained as being consonant with the original scale development and testing process.
• The p value for Bartlett’s Test was removed.