

The anglicization of American personality tests: Panacea or problem?

OVER THE PAST few years concerns have been raised about whether it is appropriate to use personality tests for selection and assessment in Britain when the tests have been developed and standardized in the USA. In particular, it has been argued that such tests may lack validity when used on British applicants and possibly lead to problems of adverse impact (McHenry, 1997). While there may be a *prima facie* case in support of such arguments it is interesting to note that, until very recently, American personality tests (e.g. the CPI, 16PF and MMPI) were routinely marketed for use in the UK with little apparent concern about their potential lack of validity for British respondents.

Having recently become concerned about this issue, some test publishers have argued that American personality tests should only be used in the UK if they have been 're-standardized' on British samples. Unfortunately, what exactly is meant by 're-standardization' is not always clear. For example, re-standardization may simply involve anglicizing a few problematic test items and collecting new normative data on UK samples. Alternatively, it may involve reconstructing the test's scales on a UK sample, to ensure that the psychometric properties of the re-standardized measure are equivalent to those of the original American version of the test.

This latter, and more adequate, conceptualization of re-standardization involves demonstrating that the item characteristics (e.g. item homogeneities, variances etc.) and scale

Richard Budd asks whether American selection tests can be successfully re-standardized for British respondents.

characteristics (e.g. scale validities, reliabilities, distributions etc.), and the underlying factor structure, are equivalent across the British and American forms of the test. While such a procedure is the only adequate way to demonstrate that an anglicized test is both reliable and valid when used on a British sample, it is nonetheless a costly and time consuming process. This may explain why, for many test publishers, re-standardizing American personality tests for use in the UK simply means anglicizing problematic test items and collecting British norms. However, this simplistic approach to re-standardization does not necessarily solve many of the problems that are inherent in using psychometric tests on populations for which these tests were not initially designed. Moreover, anglicizing test items and collecting new normative data can, as will be shown below, create new problems with these scales.

The first problem that has been raised about using American personality tests on British respondents is that British and American populations show mean differences on a number of important personality traits. As Americans obtain, on average, higher scores on American

measures of extroversion (e.g. MBTI, 16PF-5) than do the British, it has been argued that such measures may underestimate how extroverted British applicants are (Pickard, 1996). Furthermore, it has been proposed that this problem can be solved simply by anglicizing such measures and collecting new normative data on British samples. However, collecting new normative data on a population for which a test was not initially designed can generate problems of restricted range, and skewed distributions, resulting from floor and ceiling effects.

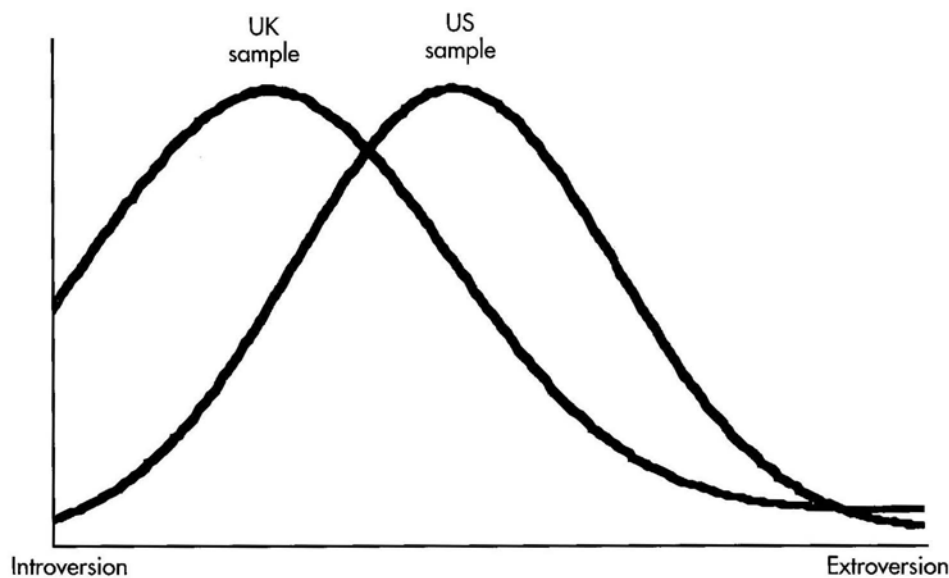
This potentially serious problem is represented graphically in Figure 1, which shows how, for a measure of extroversion, scores may be normally distributed for an American population but be severely skewed on a British sample. This results from the extroversion scores clustering around the low end of the scale for the British sample, due to British respondents failing to endorse extroversion items their American counterparts would have endorsed. Such a floor effect is thus potentially a direct consequence of the very problem that 're-standardizing' the test (on a British sample) was attempting to solve. In this way, collecting UK norms on an anglicized test

may simply create a new set of problems with the scale.

While the problems of restricted range and skewed distributions may, at first, appear to be little more than arcane psychometric issues, they do nonetheless create potentially serious practical problems for test users. This is perhaps best exemplified with reference to the 16PF-5. This recently revised version of the 16PF has a number of scales that have very poor distributions, with a total of 11 out of 15 sub-scales having truncated score ranges - most notably, Factors H (Social Boldness), M (Abstractedness) and Q2 (Self-Reliance) (Smith, 1994). Such significant distribution problems result in respondents only being able to obtain sten scores (i.e. scores scaled from 1 to 10) in the range of 3 to 8 on the Social Boldness scale, with this problem being even worse for some specific UK groups.

Moreover, what constitutes an adequate 're-standardization' of a US test is further confused by the fact that UK norms are sometimes collected using the anglicized version of the test, and on other occasions this new data is collected using the original American form of the test, as is the case with the UK norms for the 16PF-5. In

Figure 1: Illustrative comparison of scores on measure of Extroversion for UK and US samples



justifying this the UK distributors of the 16PF-5 state that 'the changes (anglicizations) are unlikely to influence item responses, since almost all of them affect only differences in punctuation, spelling and English usage' (Smith, 1994, p.9). However, this assertion runs counter to more recent pronouncements from the UK distributors of this test who note that, 'It is clear that when a British person completes the American version of the 16PF5, we cannot be confident that he or she will understand the item content in the same way that an American would' (Lord, 1997, p. 150). This clearly raises the question of whether American personality tests do, in fact, need to undergo major revision before they can be used in the UK, as some test publishers have recently begun to argue.

The answer to this question is probably, at least in part, dependent upon the personality test under consideration. For example, on occasion, anglicizing American personality test items may simply involve making minor changes to grammar and spelling. This may particularly be the case for those tests, such as the Jackson Personality Research Form (Jackson, 1967) and Hogan Personality Questionnaire (Hogan, 1984), that have been specifically designed to be easily translatable across different cultures. In such cases it is likely that few problems may result from making minor changes to item spelling and grammar. However, tests such as the California Psychological Inventory (Gough, 1987) contain many culturally specific items and may require major revision if they are to be used in the UK. Cultural differences between the UK and the USA can dramatically change the meaning of some American test items (see Table 1).

However, while some American test items may need to undergo major revision if they are to be used in the UK, these tests are often anglicized in a very arbitrary manner. Most importantly in this regard, there is little attempt to demonstrate systematically that the anglicized and original American test items have equivalent meanings in their respective cultural contexts. The arbitrary nature of this process is evident in the description of how the 16PF-5 items were anglicized. For example the UK publishers note that 'ASE

prepared the UK standardization version directly from an ASCII file of the text of the US version'. The agenda for the anglicization was to change as few items as possible while ensuring that all necessary changes were made to the content, punctuation, spelling and word order of items'. (Smith 1994, p.1)

A good example of the problems that can result from arbitrarily anglicizing American test items is to be found in the 16PF-5 item:

If I could I would rather exercise by:

- a) fencing or dancing
- b) ?
- c) wrestling or baseball,

with this item being designed to assess Factor I - Sensitivity. While baseball originated in Britain, it is not a game that many British people play for exercise. Moreover, wrestling, which is a popular (Olympic) sport in many American colleges, has a very different meaning in an American context to that attributed to this activity in the UK. Thus, if this item is to retain its intended meaning in a British context, it needs to be rewritten. For example, it might be appropriate to replace wrestling with judo and baseball with rounders. However, such choices are somewhat arbitrary in nature, as is indicated by the anglicized version of this item that appears in the British edition of the 16PF-5:

If I could I would rather exercise by:

- a) fencing or dancing
- b) ?
- c) wrestling or cricket

The inherent arbitrariness of anglicization is further highlighted by the following example. The most recent British revision of the California Psychological Inventory (Gough & Cook, 1995) contains items such as: 'I dread the thought of an earthquake' which, while meaningful in California, might not be so meaningful in a British context. This problem is further highlighted by Table 1, which lists some of the changes in item wording that have occurred in successive attempts at anglicizing the CPI.

This table not only demonstrates the arbitrary nature of anglicization, but also shows the dramatic revisions that some of these items have undergone. Most critically, such major changes in

Table 1: Original American version, original anglicized version and the most recent anglicization of some items from the CPI.

American item	Original anglicization	Current revision (309 item version)
I think Lincoln was greater than Washington	I think Queen Elizabeth was greater than Queen Victoria	In the long run, art, literature and music are more important than anything associated with political or economic events
Only a fool would try to change our American way of life	Only a fool would try to change our British way of life	The old ways of doing things are almost always the best
I prefer a shower to a bathtub	I prefer a shower to a bath	I enjoy vigorous, energetic, physical activities, even if there is a risk of pain
We ought to let Europe get out of its own mess; it made its bed, let it lie in it	We ought to leave the African countries to sort out their own problems; there is no reason for us to help them	We ought to let the poor countries of the world sort out their own problems; there is no reason for us to help them

item wording must raise concerns about whether any of these items have changed their meaning as a result of being anglicized. Clearly, the significance of this potentially serious problem depends upon how many test items have been anglicized. In the case of the 16PF-5, 17 out of 185 have been anglicized, and in the case of the CPI, 31 out of 309 items have been anglicized. With approximately 10 per cent of these items having been changed, test users can only wonder whether the American and British versions of these tests are in fact assessing the same personality traits.

The apparent arbitrariness of item anglicization raises concerns about the reliability and validity of these anglicized items (and scales) when used in the UK. Therefore, when evaluating an anglicized version of an American personality test it is particularly important to review British data on the item characteristics of those items that have (or have not!) been anglicized. Moreover, it is also important to review data on the validity of the anglicized scales, when used in a British context. However, this is often difficult as many test manuals rely heavily on validity data collected on American samples (using the non-anglicized form of the test). For example, in the UK manual for the

CPI, of the 40 criterion validity studies reported, only three are based on UK samples.

The question of the cross-cultural validity of American personality tests may be particularly critical for criterion-referenced tests, such as the California Psychological Inventory. Unlike traditional methods of test construction (e.g. item analysis and factor analysis), which aim to construct scales that measure discrete dimensions of personality, criterion referenced tests aim to construct scales that are keyed to 'real world' criteria. It has recently been argued (McHenry, 1997) that criterion referenced scales may be particularly useful in an occupational context as they measure broadly defined characteristics, rather than measuring statistically distinct, and tightly defined, personality traits (as do tests constructed via traditional procedures). However, if the 'real world' criteria that have been used to define these scales are poor, then it will be difficult to know exactly what personality characteristics these scales assess. Moreover, if the criteria that have been used to construct these scales are very culture specific, then it may be difficult to define what these scales measure in different cultural contexts.

For example, the criterion reference groups

that were used to define the Dominance scale in the CPI were constructed from college students' ratings of the most and least dominant members of their fraternity or sorority. Similarly, the CPI Socialization scale was constructed with reference to groups of juvenile offenders and 'high school disciplinary cases' versus 'normal high school students' (Gough, 1957). This must clearly raise concerns about the validity of these scales when completed by British adults. This problem is further compounded by the fact that most of the validity data that is reported in the CPI manual has been collected on American samples using the original (American) version of this test. Thus if cultural differences are sufficiently significant to require revision of American tests before they can be meaningfully used in the UK, then this must raise serious concerns about the applicability of such validity data to British samples.

Thus, simply anglicizing American personality tests and collecting British norms is insufficient to ensure that these 're-standardized' tests will be as valid and reliable when used on British respondents, as are the original versions of these tests when used on American respondents. Moreover, under certain circumstances, the anglicized versions of American personality tests may be as likely (or even more likely) to generate problems of adverse impact, as are the original American forms of these tests.

While some of these issues above may appear complex, the solution to these problems is in fact quite simple: anglicized versions of American personality tests should be judged against the same benchmarks that would be used to evaluate British-constructed personality tests. Thus, it is only appropriate to use such tests if data which has been collected on appropriate British samples, using the anglicized form of the test, demonstrate that the test is a reliable and valid measure of job relevant personality characteristics. American personality tests have historically been used in the UK because of the lack of appropriate British instruments. Given the

number of 'home grown' personality tests that are now available, for which there is good reliability and validity data, assessors must now seriously question the utility of American personality measures, where such data is missing, whether or not these tests have purportedly been anglicized.

References

- Gough, H.G. (1957) **Manual for the California Psychological Inventory**. Palo Alto, California: Consulting Psychologists Press
- Gough, H.G. (1987) **Administrator's Guide for the California Psychological Inventory**. Palo Alto, California: Consulting Psychologists Press
- Gough, H.G. & Cook, M. (1995) **Manual for the 309-item form of the California Psychological Inventory**. Oxford: Oxford Psychologists Press
- Hogan, R. (1986) **Hogan Personality Inventory Manual**. Minneapolis: National Computer Systems Inc
- Jackson, D. (1967) **Personality Research Form Manual**. Goshen, New York: Research Psychologists Press
- Lord, W. (1997) **Personality in Practice**. Windsor: NFER-Nelson
- McHenry, R. (1997) Presentation to 2nd Test User Conference, Stratford-upon-Avon
- Pickard, J. (1996) Misuse of tests leads to unfair recruitment. **People Management**, 19 December, p.7
- Smith, P. (1994) **The UK Standardization of the 16PF5: A Supplement of Norms and Technical Data**. Windsor: NFER-Nelson

Dr Richard Budd is in the Department of Clinical Psychology, Whitburch Hospital, Cardiff CF4 7XB and is an associate of Psytech International Ltd.

CPI™ and MBTI™ are Trade Marks of Consulting Psychologists Press Inc.