

Paul A Whitelaw, Victoria University, Melbourne ¹

RELIABILITY AND VALIDITY: The Terrible Twins of Good Research

Much comment is made about the reliability and validity of the MLQ. However, whilst reliability and validity are important in statistical terms, they need to be seen in the broader research context. At best, reliability and validity are tactical or operational level issues. More importantly, are the higher level strategic issues of good research, present? Namely, has the research been scientifically developed and executed?

One of the key strengths of the MLQ is its scientific genesis. It has emerged from a thorough and rigorous research process over the past fifteen years. According to renowned research academic, Uma Sekaran, there are Eight Hallmarks to good scientific research (Sekaran, Research Methods for Business: A skill building approach. 3rd. ed. John Wiley & Son. New York. pp.10-14).

The Eight Hallmarks are:

- 1 Purposiveness
- 2 Rigour
- 3 Testability
- 4 Replicability
- 5 Precision and confidence
- 6 Objectivity
- 7 Generalisability
- 8 Parsimony

Purposiveness means that the research has been done to serve a worthwhile and meaningful purpose. The development of the MLQ has certainly met the needs of many individuals and organisations.

Rigour implies that the researchers have been painstaking and thorough in their research. A quick review of Bass and Avolio's work on the MLQ suggests a very rigorous approach to the development of the instrument.

Testable research must be about measuring some form of performance. By using the three performance measures of staff satisfaction, staff performance and staff extra effort, the MLQ instrument provides a high level of testability.

Replicability means that the research produces similar results in similar settings. This consistency of results of the MLQ across a wide variety of organisations and institutions confirms a high level of replicability.

Precision refers to how close the results of the sample are to "the wider reality" whilst **confidence** refers to the probability that the results are correct. Precision and confidence can only be assured by comprehensive sampling. Given that the MLQ has now been conducted repeatedly for over 15 years to more than xxx individuals in xxx organisations, one can be increasingly confident in and precise with its findings.

The **objectivity** of the researchers is critical to good research. Whilst they can be passionate about the search for the answer, they must not be "wedded" to any particular answer and must be prepared to subject their work to the intense scrutiny of other, independent

¹ Paul A Whitelaw is a Lecturer in the School of Hospitality, Tourism and Marketing and the Victoria Business School at Victoria University where he lectures on statistics and research methods. He is also a member of the MLQ Network.

researchers. The widespread usage of the MLQ by other academics and the extensive debate about the MLQ in the research literature confirms that the MLQ has been scrutinised and validated by objective, disinterested parties.

Generalisability refers to the applicability of the findings in a variety of settings. The generalisability of the MLQ is one of its great strengths. Findings from research in large industrial settings, are consistent with findings in government departments and small not for profit organisations.

Finally, good research must be **parsimonious**, that is it must be comprehensive enough to cover the key issues, yet be small enough so that people can understand it. The nine dimensions of the MLQ are sufficient to cover the richness of the full range of leadership issues, yet nine factors are not too extensive to become incomprehensible and unmanageable.

From this high level, strategic perspective, we can develop key operational or tactical requirements, namely reliability and validity.

Reliability and **validity** are the statistical criteria used to assess whether the research provides a good measure. In the case of the MLQ we use reliability and validity to assess whether the MLQ really does measure leadership. "Validity tests how well an instrument measures the *particular concept* it is supposed to measure. Reliability tests *how consistently* an instrument measures that concept" (edited from Sekaran, p. 171).

Reliability and validity is best conceptualised by using the chocolate cake recipe as an example. Reliability suggests that any person, provided that they follow the recipe, will produce a reasonable chocolate cake, or at least something that you can identify as a chocolate cake. Validity suggests that if the recipe includes chocolate, then the cake will look like a chocolate cake, smell like a chocolate cake and taste like a chocolate. That is, the "proof is in the pudding".

When you think about it, it becomes apparent that an instrument must firstly be reliable before it can be valid. That is, you have to be confident that what you get is consistently reproducible (ie. the recipe consistently produces something like a cake). Once you are confident in the consistency of the output, then you can scrutinise it to assess whether it is what is purports to be (ie. a chocolate cake).

Therefore, does the MLQ accurately measure leadership (validity)? And, does it do this consistently in a variety of situations (reliability)? The comparative studies and replication studies confirm that the MLQ can be considered a reliable and valid instrument.