Measurement of Intangibles: Problems of Reliability and Validity

Santanu Roy and P S Nagpaul
National Institute of Science, Technology and Development Studies (CSIR), New Delhi 110 012, India

e-mail: rsan58@yahoo.co.uk Fax: 0091-11-5754640/6237401

Theorists in the social sciences often use concepts that are formulated at rather high levels of abstraction. These are different from the variables that are the stock-in-trade of empirical researchers in these fields. The problem of bridging the gap between theory and research is then seen as one of measurement errors. Measurement is a process involving both theoretical as well as empirical considerations. From an empirical standpoint, the focus is on the observable response – whether it takes the form of a mark on a self-administered questionnaire, the behaviour recorded in an observational study, or the answer given to an interviewer. Theoretically, interest lies in the underlying observable (and directly unmeasurable) concept that is represented by the response. Measurement focuses on the crucial relationship between the empirically grounded indicator(s) – that is, the observable response – and the underlying unobservable concept(s). Intangibles are often measured indirectly (referred to as quasi-quantitative measurement). Such a measurement has to encounter the problem of measurement error as well as face the twin problems of ensuring the reliability and validity of such measures. If the complexity of reality, in particular the qualitative nature of information that we possess about that reality, is to be captured in our theories that we develop, then it is necessary to construct methods of ensuring the reliability and validity of such measures. The paper discusses the issues of reliability and validity of such measures. The paper gives specific examples of measurement of intangibles in two different contexts – measurement by surrogate indicators and measurement by quasi-quantitative methods. The first example considers the case of tacit knowledge as measured by a surrogate indicator, that is, the way in which the scientific and technical manpower in the laboratories of the Council of Scientific and Industrial Research (CSIR), India are deployed across different functional areas like research and development, industrial liaison, planning and monitoring, infrastructure services, workshops including glass blowing and animal house, engineering services and in the pilot plants for long periods. The second example considers the case of measurement of effectiveness of research units in CSIR laboratories where quasi-quantitative measurement technique has been employed. A measurement model has been developed for measuring research unit effectiveness. The fit of the model, its convergent validity and its discriminant validity have been assessed.

1. Measurement and Measurement Errors

Measurement could be viewed as a process involving an explicit, organized plan for classifying (and often quantifying) the particular sense of data at hand – the indicants – in terms of the general concept in the researcher’s mind. Social science theories often involve concepts that are formulated at rather high levels of abstraction. These concepts are quite different from the variables that are the stock-in-trade of empirical researchers in these fields. The problem of bridging the gap between theory and research is then seen as one of measurement errors.

According to Carmines and Zeller, ‘Measurement is a process involving both theoretical as well as empirical considerations’. From an empirical standpoint, the focus is on the observable response – whether it takes the form of a mark on a self-administered questionnaire, the behaviour recorded in an observational study, or the answer given to an interviewer. Theoretically, interest lies in the underlying observable (and directly unmeasurable) concept that is represented by the response. Measurement focuses on the crucial relationship between the empirically grounded indicator(s) – that is, the observable response – and the underlying unobservable concept(s)’ (emphasis original). The question is how a social scientist can determine the extent to which a particular empirical indicator (or a set of empirical indicators) represents a given theoretical concept.

Subjective measures have not only weak measurement properties, they are also influenced by systematic and random measurement errors. Although much research has been conducted to deal with many of the problems associated with errors in variables, particularly in the area of parameter estimation or in the case of repeated time series measurements, most researchers do not spend much effort investigating the quality of the data, and often
ignore measurement errors which may be important enough to influence the conclusions of empirical work as typically conducted. Carmines and Zeller have observed that indicators always contain random error to a greater or lesser degree. They have also emphasized that the effects of random error are totally unsystematic in character. Hence, it is essential that the reliability and validity of quasi-quantitative measurement should be assessed before they are used in empirical studies.

The study reported in this paper looks into the question of the measurement of intangibles. Intangibles are not directly measurable. These are to be measured indirectly and this process is often subjective in nature. Any subjective measurement would imply tackling its associated problems of their reliability and validity, otherwise the acceptability of such measures would be put to question. The paper has been divided into different sections. These sections initially deal with aspects of the subjective measurement of the intangibles and examines in detail the question of reliability and validity of measurements. The latter sections give specific examples of measurement of intangibles in different contexts. This has been divided into two categories – measurement by surrogate indicators and measurement by quasi-quantitative methods. Measurement of tacit knowledge is the one cited under the first sub-category and the measurement of research unit effectiveness the one cited under the second sub-category. The paper concludes with an illustration of how the problems of reliability and validity have been taken care of while measuring research unit effectiveness.

2. Intangibles and the Issue of Subjective Measurement

Many concepts in the social sciences are difficult or impossible to measure objectively. This limitation forces a reliance on subjective measures that typically contain both systematic and random measurement errors. Intangibles, like those encountered while researching on innovation and the factors responsible for fostering an innovative climate, are often measured indirectly. This indirect measurement could take two forms. First, it could be measured by surrogate indicator(s). In this fashion tacit knowledge has been measured (an earlier work on this was carried out by Roy and Nagpaul). Second, two or more indicators (observed variables) might measure a certain latent variable or construct, like the way we have attempted to measure R&D effectiveness. This kind of measurement is referred to as quasi-quantitative measurement. Such a measurement has to encounter the problem of measurement error as well as face the twin problems of ensuring the reliability and validity of such measures. Other examples of intangibles of this kind could be quality maintainability, learning from market, information employed for competition, enterprise info-connectivity, skill in designing, and skill-shopfloor-RM (repair and maintenance) integration – all these indicators and a few others were developed as a part of study on Industrial Innovation Indicators carried out by a research team, of which the first author was a part.

From the above discussion, it is clear that if the complexity of reality, in particular the qualitative nature of information that we possess about that reality, is to be captured in the theories that we develop, then it is necessary to construct methods of analysis which take into account many variables whose inter-relations underline reality as well as any qualitative property of those variables. The theory-laden nature of data collection, the existence of spurious correlations, and the difficulty of defining latent variables which are not directly observable, are all examples where considerable potential for erroneous explanations exist if our theories and biases are not explicitly laid out and examined right from the start of a study. In this paper, examples of both the measurements of tacit knowledge and R&D effectiveness are given.

3. Reliability and Validity of Subjective Measurement

There is a body of literature that criticizes the approach to model human phenomena in time and space for their attempts to subject human relations to numerical analysis. According to these critics, knowledge of human beings involves the apprehension of qualities, which by their very nature escape net of numbers. Measurement is pointless at best, a hopeless distortion or obfuscation of what is really important. Notwithstanding these criticisms pointing to the limitations of measurement, however, there is increasing recognition that a qualitative approach need not eschew measurement. Social scientists have been more and more concerned with measuring qualities in order to grapple with complex configurations and the ambiguities inherent in human perceptions and behaviour. Bollen and Paxton have
demonstrated the feasibility of investigating biases in subjective measures under a broad range of research designs. Both systematic and random measurement errors can frustrate attempts to understand the concepts being studied. Hence, it is essential that their reliability and validity should be assessed before they are used in empirical studies.

Reliability refers to the ability to achieve identical or similar outputs from the work of different researchers and by the repeated use of the technical instruments for data collection. Reliability is a methodological issue. It concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials. The tendency towards consistency found in repeated measurements of the same phenomenon is referred to as reliability.

A reliability coefficient is always represented by a numerical value between zero and one which reflects the stability or consistency of the measurements obtained from a test. Regarding internal consistency reliability of tests, it estimates the reliability of the test on a single administration and indicates the extent to which the tests are homogeneous, i.e. measure a single trait.

Regarding the question, 'Is the reliability coefficient high enough?', Kelley has said that to make decisions about groups (such as attitudes of groups, general performance levels of groups, etc.), a reliability coefficient of at least 0.50 is required.

The reliability coefficient should never be misconstrued or misrepresented as the validity of the test. A perfectly reliable test may not measure anything of value. In a very general sense, any measuring device is valid if it does what it intends to do. An indicator of some abstract concept is valid to the extent that it measures what it purports to measure. Indeed, strictly speaking, one does not assess the validity of an indicator but rather the use to which it is being put.

Thus, while reliability focuses on a particular property of empirical indicators—the extent to which they provide consistent results across repeated measurements—validity concerns the crucial relationship between concept and indicator. Reliability is basically an empirical issue, focussing on the performance of empirical measures. Validity, in contrast, is usually more of a theoretically oriented issue because it inevitably raises the question, 'valid for what purpose?'. Validity is evidenced by the degree that a particular indicator measures what it is supposed to measure rather than reflecting some other phenomenon.

Strictly speaking, one validates, not a test, but an interpretation of data arising from a specified procedure. The distinction is central to validation because it is quite possible for a measuring instrument to be relatively valid for measuring one kind of phenomenon but entirely invalid for assessing other phenomena.

Construct validity is the extent to which an observation measures the concept, it purports to measure. The test user wishes to infer the degree to which the individual possesses some hypothetical trait or quality (construct) presume to be reflected in the test performance. Construct validity deals with validation of theory. A test has construct validity if relationship between scores obtained on it and various other measures entering into the theoretical formulation turn out to be significant and in the predicted direction.

Campbell and Fiske have proposed two broad criteria for construct validation: convergent validity and discriminant validity. Convergent validity refers to the extent of which multiple measures of a construct agree with each other. If two or more measures are true indicators of a concept, then they should necessarily be highly correlated. This assumption is consistent with the 'reflective measurement model'. Failure to find high covariation among multiple measures of a construct would imply that either the measures are poor and/or the construct and the measures do not correspond with each other. Discriminant validity is the degree to which measures of different constructs are distinct from each other. This means that measures of different constructs should share little common variance (in a relative sense).

Cronbach and Meehl have observed that construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured. Construct validity is woven into the theoretical fabric of the social sciences, and is thus central to the measurement of abstract theoretical concepts. Fundamentally, construct validity is concerned with the extent to which a particular measure relates to other measures consistent with theoretically
derived hypotheses concerning the concepts (or constructs) that are being measured.

The above section highlights the importance of ascertaining the reliability and validity of any measurement, particularly of the subjective variety, before these measures are used in any further analysis. These concerns, clearly, assume criticality while attempts are made to measure intangibles. The following sections give specific examples of measurement of intangibles like tacit knowledge and effectiveness of research units, and how the reliability and validity of such measures have been assessed.

4. Measurement of Intangibles

4.1 Measurement by Surrogate Indicator(s)

The surrogate indicators are supposed to measure an intangible variable which otherwise cannot be measured directly. For example, in the study carried out on Industrial Innovation Indicators mentioned earlier, 'Skill in Designing' was developed as an indicator. It had been defined as the product of two variables. These variables were supposed to act as surrogate indicators measuring this aspect. The first variable was defined as the ratio between number of design employees having qualifications above Industrial Training Institutes (ITI) and the total number of employees. The second variable was defined as the ratio between the number of design employees having design experience of five years and above and the total number of employees. How another intangible variable, tacit knowledge, had been measured is the subject matter of the following section.

4.1.1 Measurement of Tacit Knowledge

Tacit knowledge was measured by a surrogate indicator, that is, the way in which the scientific and technical manpower in the laboratories of the Council of Scientific and Industrial Research (CSIR), India are deployed across different functional areas like research and development, industrial liaison, planning and monitoring, infrastructure services, workshops including glass blowing and animal house, engineering services and in the pilot plants for long periods. This work had been initiated by the present authors which has since been further improved and has been brought out in an as yet unpublished paper.

CSIR laboratories generate knowledge – both along the theoretical as well as empirical dimensions. In this knowledge generation process, we have inputs of basic research, applied research, pilot plants, etc. These also operate at different levels of qualifications and experience and expertise among the scientific and technical personnel of the laboratories. Does there exist a mismatch among these different dimensions and levels? How are these different components structured? These questions are vital in order to put into a proper perspective the functional dimensions scientific and technical personnel working in these laboratories.

The scientific and technical personnel of CSIR are strategic assets for the organization, more so because of the tacit knowledge they possess as a result of actively pursuing R&D activities in different functional areas over a long period of time. Tacit knowledge has been recognized as a major input to any technological innovation effort. The strategic technological agenda is linked to the organization’s technical and managerial knowledge and assumptions. This is largely experiential, cumulative and often tacit. Much of this tacit knowledge is held in decentralized units and structures, often non-disseminated and immune to external challenge. Studies of innovation, technology transfer and technology diffusion identify tacit knowledge as an important component of the knowledge used in innovation. Tacit knowledge is a source of competitive advantage. Dutta and Weiss have argued that the protection of tacit technological knowledge from potential opportunism is of importance to technologically innovative organizations. Tacit know-how has become recognized as playing a key role in organizational growth and economic competitiveness. It forms an important element in an organization’s knowledge base and has a central role in organizational learning. The generation of tacit knowledge is an inevitable adjunct to advances in science and technology, and organizations acquire such knowledge to support innovation in a purposive manner. Therefore, when we talk of strategic deployment of scientific and technical manpower across different functional areas in CSIR laboratories, we also refer to the component of tacit knowledge inherent in such a deployment.

In order to analyze different functions carried out by the scientific/technical personnel of CSIR, these were grouped into six categories. Information was sought from
Correspondence analysis is an exploratory statistical study which displays the rows and columns of a rectangular data matrix as points in a scatterplot, often called a 'map'. It is a powerful graphical tool in many situations involving categorical data. The method is particularly helpful in analyzing cross-tabular data in the form of numerical frequencies, and results in an elegant but simple graphical display which permits a rapid interpretation and understanding of the data. Correspondence analysis allows the representation of column and row elements of the data matrix in low dimensional (usually two-dimensional) subspaces. This representation can be used to reveal the structure and pattern hidden in the data. The two dimension factorial map reveals the main feature of the multi-dimensional data. The third factorial axis (and perhaps the fourth axis as well which together with the first three axes might yield the most parsimonious representation of the data) represents complementary data for further analysis. In this study, correspondence analysis was carried out using SIMCA-2 software.

This study has provided the authors with a map with groupings of laboratories possessing an in-built strength in basic research or an in-built strength in engineering services or in the working of pilot plants or in other R&D thrust areas. What we see in such correspondence analysis maps are displays and profiles of such tacit knowledge in different functional areas.

An example of such correspondence analysis maps are given in Figures 1 and 2. Figure 1 presents the two-dimensional map constituted by the first two factorial axes for the CSIR laboratory points and Figure 2 presents the same for the various function points. Both these figures represent the case of the scientific manpower of CSIR, which should be read and interpreted simultaneously. The representation of functions and laboratories in different graphics was done to avoid cluttering of the points in the same graphic. However, these two graphics are super-imposable.

The first two axes, accounting for 83.6% of the total information, represent the main features of the multi-dimensional data.
On the cloud of functions, the first factorial axis represents a polarity (bi-polar) between function 4 - pilot plants, experimental field stations, etc. and function 1 - R&D work. Function 1 is projected on this axis with negative coordinates, whereas function 4 is projected on this axis with positive coordinates. This implies that laboratories which emphasize R&D work for their scientific personnel and deploy their scientific manpower in this area tend to de-emphasize their work related to pilot plants, etc., and vice-versa.

The laboratories projected on this axis can be classified into two clusters, depending upon whether they are projected with positive coordinates (correlated with function 4) or negative coordinates (correlated with function 1).

Cluster 1 (positive coordinates): CIMAP, Lucknow; CFRI, Dhanbad; and NCL, Pune.

Cluster 2 (negative coordinates): RRL, Jorhat; NGRI, Hyderabad; IIP, Dehradun; CEERI, Pilani; CSMCRI, Bhavnagar; NBRI, Lucknow; IICB, Calcutta; NEERI, Nagpur; SERC, Chennai; SERC, Ghaziabad; NISCOM, New Delhi; CSIR, Complex, Palampur; NIO, Goa; NISTADS, New Delhi; NAL, Bangalore; and ITRC, Lucknow.

On the cloud of functions, the second factorial axis is unipolar - both function 2 - S&T services including testing, data processing, field work, planning and coordination, etc. and function 6 - research support functions, are projected on the axis with positive coordinates. This implies that the laboratories which are projected on this axis with positive coordinates emphasize the function of working in the areas of S&T services and research support functions for their scientific personnel and deploy scientific manpower in these areas whereas laboratories, which are projected with negative coordinates on this axis, de-emphasize these roles for their scientific manpower.

The laboratories projected on this axis can be classified into two clusters, depending upon whether they are projected on this axis with positive coordinates (correlated with both function 2 and function 6) or negative coordinates (anti-correlated with both function 2 and function 6).

Cluster 1 (positive coordinates): CMERI, Durgapur; NML, Jamshedpur; CFTRI, Mysore; IMT, Chandigarh; and CBRI, Roorkee.
Cluster 2 (negative coordinates): CECRI, Karaikudi; CEERI, Pilani; SERC, Ghaziabad; NISCOM, New Delhi; RRL, Jammu; NCL, Pune; NIO, Goa; ITRC, Lucknow; and NISTADS, New Delhi.

4.2 Measurement by Quasi-Quantitative Methods

The second kind of measurement involves measuring a latent variable or a construct by means of several observed variables or indicators. This is usually accomplished with the help of a questionnaire study, sometimes using semantic differential scales like the five point Likert scale, commonly employed in organizational behaviour and psychological studies. This kind of measurement involves subjectivity and is accompanied by systematic and random measurement errors. Therefore, the question of reliability and validity of such measurement assumes criticality. Following is an example of such an approach, that is of the measurement of research unit effectiveness involving four latent variables or effectiveness dimensions. The discussion also details out how the problems of assessing reliability, convergent and discriminant validity have been addressed for such quasi-quantitative and subjective measurement of intangibles.

4.2.1 Measurement of Research Unit Effectiveness

This is an example of the measurement of the other kind where two or more indicators (observed variables) measure a certain latent variable or construct. This kind of measurement is referred to as quasi-quantitative measurement. This example is based upon the work carried out by Roy, Nagpaul and Mohapatra. Developing a model to measure the effectiveness of the Research Units (RU) working in different laboratories of CSIR was the subject matter of this work. In this study, a research unit (RU) was operationally defined as a unit that has the following characteristics: (i) it has at least one project in the unit; (ii) it has a total expected life span of at least one year; and (iii) it is comprised of at least three core members, among whom there is one scientist who is the head of the unit (a core member is an individual researcher or a technician who devotes at least eight hours per week to the work of the research unit and who has direct or indirect communication with the head of the unit at least once in a month).

For the purpose of the study, usable data were obtained from 236 research units.

In our model, there are four dimensions of research unit effectiveness. These are measured in terms or four constructs or latent variables operationalized by thirteen indicators, measured on 5-point semantic differential scales. The effectiveness measures are: (1) R&D effectiveness; (2) recognition; (3) user-oriented effectiveness; and (4) administrative effectiveness. The latent variables and their manifest indicators are listed in Table 1.

The performance of research units was evaluated by the following strata: (i) head of the unit; (ii) core members of the unit: a minimum of three and maximum of six scientists and technicians selected randomly within the unit; (iii) one to three external scientific evaluators; and (iv) one to three external administrative evaluators familiar with the research activities of the unit.

For each item, unit level scores were constructed by adopting the procedure suggested by Van de Ven, Delbeg and Koenig (Jr). RU head’s score was added to the mean value of the scores given by the core members in the unit and the mean values of the scores given by the external evaluators and the sum was divided by the number of strata involved in the evaluation.

A reflective measurement model was hypothesized for subjective evaluation of research performance at the micro-level of research unit. In Table 1, \( \xi_1, \xi_2, \xi_3, \) and \( \xi_4 \) represent the latent variables of R&D effectiveness, recognition, user-oriented effectiveness and administrative effectiveness, respectively. The \( x \) variables represent the indicators of these latent variables.

LISREL technique has been adopted as the primary methodology for the study. The model incorporates unobserved (latent variables), the relation between these and observed variables and an allowance for errors of measurement in the independent and dependent latent variables, and a causal model linking the latent variables together. LISREL produces a full information maximum likelihood solution (FIML), which makes use of all information in the data about each parameter in generating its estimates. LISREL 7.16 program was used in the study reported in this paper.

Measurement Errors, Reliability and Validity

The development of confirmatory factor analysis and the more general model for the analysis of covariance
structures has been largely instrumental in providing a statistical procedure to tackle the problem of measurement errors. Bollen as well as Miller and Blalock have suggested that analysis of experimental data could benefit from structural equations that allow measurement error, multiple indicators and test for confounding variables.

Assessment of Model's Fit, Reliability and Validity

The probability level associated with a given $\chi^2$ statistic indicates the probability of obtaining a larger $\chi^2$ value, given that the hypothesized model is supported. The higher the value of $p$, the better is the fit, and as a rule of thumb, values of $p > 0.10$ are considered as indication of satisfactory fit. However, the $\chi^2$ statistic is sensitive to sample size and has been criticized for several reasons. In large samples, the $\chi^2$ test is too merciless; even models that approximate the sample covariance matrix are usually rejected. In small samples, the $\chi^2$ test lacks power as it is too forgiving of important misspecifications in the model. Browne and Cudeck have succinctly pointed out that statistical goodness of fit tests are often more a reflection of the sample size than of the adequacy of the model.

Another problem with $\chi^2$ statistic is that a test of statistical significance does not provide information regarding the degree of fit. Therefore, rather than trying to ask whether a model is correct or fits the population covariance matrix exactly, it is sensible to assess the degree of lack of the model fit. A number of fit indices have been proposed in the literature for assessing the discrepancy between the predicted and observed covariances. These indices can be classified into two types. One type of indices directly assesses how well the covariances predicted from the parameter estimates

<table>
<thead>
<tr>
<th>Table 1 — Latent variables and indicators*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Strata</td>
</tr>
<tr>
<td>RU Head Core Member</td>
</tr>
<tr>
<td>External evaluators</td>
</tr>
<tr>
<td>Scientific Administrative</td>
</tr>
<tr>
<td>R&amp;D effectiveness</td>
</tr>
<tr>
<td>X Innovativeness</td>
</tr>
<tr>
<td>X Productiveness</td>
</tr>
<tr>
<td>X Meeting Institute's R&amp;D objectives</td>
</tr>
<tr>
<td>X Contribution to science &amp; technology</td>
</tr>
<tr>
<td>X Meeting quality requirements</td>
</tr>
<tr>
<td>Recognition</td>
</tr>
<tr>
<td>X National reputation of the unit's work</td>
</tr>
<tr>
<td>X Intern'l reputation of the unit's work</td>
</tr>
<tr>
<td>X Demand for publications</td>
</tr>
<tr>
<td>User-oriented effectiveness</td>
</tr>
<tr>
<td>X Social value of output</td>
</tr>
<tr>
<td>X Usefulness in solving societal problems</td>
</tr>
<tr>
<td>X Use of R&amp;D results</td>
</tr>
<tr>
<td>Administrative effectiveness</td>
</tr>
<tr>
<td>X Success in meeting time schedule</td>
</tr>
<tr>
<td>X Success in staying within the budget</td>
</tr>
</tbody>
</table>

*Latent variables $\xi_i$, indicator variables $X_i$
Table 2 — Matrix of intercorrelations among effectiveness measures

<table>
<thead>
<tr>
<th></th>
<th>RD Eff</th>
<th>Recognition</th>
<th>User Eff</th>
<th>Adm Eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD Eff</td>
<td>0.601</td>
<td>0.784</td>
<td>0.813</td>
<td>0.514</td>
</tr>
<tr>
<td>Recognition</td>
<td>0.601</td>
<td>0.784</td>
<td>0.813</td>
<td>0.514</td>
</tr>
<tr>
<td>User Eff</td>
<td>0.630</td>
<td>0.402</td>
<td>0.448</td>
<td>0.665</td>
</tr>
<tr>
<td>Adm Eff</td>
<td>0.366</td>
<td>0.324</td>
<td>0.448</td>
<td>0.518</td>
</tr>
</tbody>
</table>

(Disattenuated Correlations)

approximate the sample covariances. Goodness-of-Fit index (GFI) is an example of such an index. Another type of indices assesses the fit by the degree of which the model accounts for sample covariances related to a more restricted nested model, usually null model, in which all indicators are specified as uncorrelated. Bentler and Bonnet's normed-fit-index (NFI) is an example of such an index. The null model serves as the statistical baseline of comparison for the evaluation of the fit. NFI can be expressed in terms of \( \chi^2 \) values:

\[
NFI = \frac{(\chi^0 - \chi^t)}{\chi^0}
\]

where the suffix 0 refers to the null model and the suffix t refers to the target model. However, as noted by Bollen, NFI is also affected by the sample size. To rectify this problem, he has proposed an incremental fit index \( \Delta_2 \) which is defined as follows:

\[
\Delta_2 = \frac{(\chi^2 - \chi^2)}{(\chi^0 - df)}
\]

where \( df \) is the degree of freedom of the target model. The rationale underlying \( \Delta_2 \) is that if the model is correctly specified then the denominator of \( \Delta_2 \) is the expected value of its numerator.

In the case of the measurement model of R&D effectiveness under consideration, Bollen's incremental fit index was computed as follows:

\[
\Delta_2 = \frac{(\chi^2 - \chi^2)}{(\chi^0 - df)} = \frac{(1021.91 - 90.86)}{(1021.91 - 53)} = 0.961
\]

The value of \( \Delta_2 \) being quite close to unity supports the hypothesized model.

Both convergent validity and discriminant validity of the model were assessed. For assessing convergent validity, correlations among the latent variables (disattenuated correlations) and among the composite measures (attenuated correlations) are evaluated and compared. For comparison, the correlations between the latent variables (disattenuated correlations) and the composite measures (attenuated correlations) are presented together in Table 2. Random measurement error always attenuates simple correlations. When two non-reliable measures are correlated, the correlation coefficient is said to be attenuated. Attenuation means that the correlation is reduced because of uncorrelated errors of measurement of the imperfect variables. In Table 2, it is observed that each of the disattenuated correlations among the latent variables is greater than the attenuated correlations among the composite measures. This effect is greater for more unreliable measures. Therefore, the observed score correlation involving administrative effectiveness (the least reliable measure) tends to be most strongly attenuated.

The discriminant validity results are presented in Table 3. Discriminant validity is achieved when measures of each dimension converge on their corresponding true scores. In other words, it is the degree to which a theoretical dimension in a theoretical system differs from other dimensions in the same system. This can be tested by requiring that the correlations between the pairs of dimensions are significantly different form...
Table 3 — Assessment of discriminant validity

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²</th>
<th>Degrees of freedom</th>
<th>Difference in χ²</th>
<th>Level of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Model</td>
<td>90.86</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constrained Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>φ 12=1</td>
<td>115.93</td>
<td>54</td>
<td>25.07</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>φ 13=1</td>
<td>110.82</td>
<td>54</td>
<td>19.96</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>φ 14=1</td>
<td>110.55</td>
<td>54</td>
<td>19.96</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>φ 23=1</td>
<td>143.97</td>
<td>54</td>
<td>53.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>φ 24=1</td>
<td>111.36</td>
<td>54</td>
<td>20.50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>φ 34=1</td>
<td>102.69</td>
<td>54</td>
<td>11.83</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

This requires the comparison of a model with the correlation constrained to equal one with the unconstrained model. This kind of χ² significance test of the difference between two models was advocated by James et al. A significantly lower χ² value for the unconstrained model compared with the χ² value for the constrained model provides support for discriminate validity. A χ² difference with associated p value less than 0.05 supports the discriminant validity criterion. The data presented in Table 3 support the discriminant validity of the hypothesized model.

Taken together, the various statistics computed by the program and goodness-of-fit tests provide a strong support for unidimensionality and convergent validity of the hypothesized model. The value of 0.961 for Bollen's incremental fit index indicates acceptable fit of the model. The hypothesized model also satisfies the criterion of discriminant validity.

5. Conclusion
Intangibles, by the very nature of these variables, are often difficult or at times almost impossible to precisely define and therefore quantify. How to measure the values of such variables has become a moot research question. As it turns out, there is no single method to go about it. It depends upon the specificity of the situation and the larger research question of the purpose of such a measurement and the context in which such measures would be used that would greatly determine the actual methodology followed. The importance of this problem of measurement of intangibles is highlighted. Often the intangibles are measured subjectively, and such an approach is accompanied by systematic and random measurement errors. Therefore, the question of measurement errors should also be considered under such situations. Thus, the paper also attempts to bring out the significance of ascertaining the reliability and validity of such measurement before these measures are used in any further analysis. Examples of measurement of intangibles by means of surrogate indicators and by quasi-quantitative methods are given to illustrate the way intangibles could be measured in different situations and specificities and the way the reliability and validity of the measures could be assessed.

References


42 Saris W E & Stronkhorst L, Causal Modelling in Non-Experimental Research: An Introduction to the LISREL Approach (Scientometric Research Foundation, Amsterdam, the Netherlands) 1984.


44 Miller A D, Logic of Causal Analysis: From Experimental to Non-Experimental Designs, in Causal Models in Social Sci-


About the authors

Dr Santanu Roy, Scientist, NISTADS, New Delhi, obtained his PhD in Industrial Engineering Management, from IIT, Kharagpur and his M.S. Physics from IIT, Delhi. He is the member of several professional societies, including Systems Dynamics Society. He has travelled extensively and has presented his papers at several international fora. He has published in several international journals and edited volumes. His research interests include systems dynamics, modelling, and measurement.

P.S.Nagpaul, formerly Emeritus Scientist, NISTADS, New Delhi has been actively engaged in research in the areas of Scientometrics, measurement, organizational behaviour, scientific productivity, etc. for the last two decades. He has cooperated in several international cooperative research projects, guided, edited and written books and contributed to several international journals and volumes. He has guided several PhD students.