Importance values for Importance–Performance Analysis: A formula for spreading out values derived from preference rankings

Javier Abalo a,⁎, Jesús Varela a, Vicente Manzano b

a University of Santiago de Compostela, Spain
b University of Sevilla, Spain

Abstract

Importance–Performance Analysis (IPA) is a simple and useful technique for identifying those attributes of a product or service that are most in need of improvement or that are candidates for possible cost-saving conditions without significant detriment to overall quality. To this end, a two-dimensional IPA grid displayed the results of the evaluation about importance and performance of each relevant attribute. This paper shows that ordinal preferences are better than metric measures of the importance dimension and proposes a formula to transform the ordinal measure into a new metric scale adapted to the IPA grid. This formula makes allowances for the total number of features considered, the number of rankings, and the reported orders of preference.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Importance–Performance Analysis; Importance measurement; Ordinal data; Multiattribute evaluation

1. Introduction

Since the 1950s, the majority of both decision-making (Churchman and Ackoff, 1954; Edwards, 1954; Tversky, 1969) and attitude models (Fishbein, 1963; Fishbein and Ajzen, 1975; Rosenberg, 1956) have included stimuli among the key attributes for study in order to obtain an overall evaluation of these topics. This approach is implicit in conjoint measurement techniques (Green and Rao, 1971; Luce and Tukey, 1964; Varela and Braña, 1996) and is explicit in complex models of consumer satisfaction (LaTour and Peat, 1979; Oliver, 1993; Spreng et al., 1996; Tse and Wilton, 1988; Wilkie and Pesse, 1973) and perceived service quality (Cronin and Taylor, 1992, 1994; Parasuraman et al., 1985, 1988).

Developing this idea, Martilla and James (1977) devised Importance–Performance Analysis (IPA) as a simple graphical tool to further the development of effective marketing strategies based on judgments of the importance and performance of each attribute. The key objective of IPA is diagnostic in nature: this technique aims to facilitate identification of attributes for which, given their importance, the product or service underperforms or overperforms. To this end, the importance measure represents the vertical axis, and the performance measure constitutes the horizontal axis of a two-dimensional graph. These two axes divide the IPA grid into four quadrants where every attribute shows up according to its mean rating on importance and performance scales. In the original version of IPA, the appearance of an attribute in the top left quadrant of the grid is indicative of underperformance, and its appearance in the bottom right quadrant is indicative of overperformance. Product or service improvement efforts should focus on attributes in the former situation, while attributes in the latter situation are candidates for possible cost-cutting strategies (Fig. 1). Therefore, Importance–Performance Analysis provides a useful and easily understandable guide for identifying the most crucial product or service attributes in terms of their need for managerial action, as a means to develop successful marketing programs to achieve advantage over competitors.

Originally devised with marketing uses in mind, the application of IPA extends to a wide range of fields, including health service provision (Abalo et al., 2006; Dolinsky, 1991;
Dolinsky and Caputo, 1991; Hawes and Rao, 1985; Yavas and Shemwell, 2001), education (Alberty and Mihalik, 1989; Ford et al., 1999; Nale et al., 2000; Ortinau et al., 1989), industry (Hansen and Bush, 1999; Matzler et al., 2004; Sampson and Showalter, 1999), internal marketing (Novatorov, 1997), service quality (Ennew et al., 1993; Matzler et al., 2003) or tourism (Duke and Persia, 1996; Evans and Chon, 1989; Hollenhorst et al., 1992; Hudson et al., 2004; Picón et al., 2001; Uysal et al., 1991; Zhang and Chow, 2004). However, difficulties in its application soon became apparent and have prompted IPA versions that generally differ from Martilla and James’ original version in one, or both, of the two ways: 1) the grid partition into areas with distinct significance for decision-makers; and 2) the measurement of both importance and performance, the former especially.

2. Alternative IPA representations

Among the alternative partitions of the IPA grid (reviewed by Abalo et al., 2006; Bacon, 2003; or Oh, 2001), the most interesting are those that highlight the difference between importance and performance ratings by means of an upward diagonal line that represents points where ratings of importance and performance are exactly equal; this iso-rating line divides the graph in two great areas (Abalo et al., 2006; Bacon, 2003; Hawes and Rao, 1985; Nale et al., 2000; Picón et al., 2001; Sampson and Showalter, 1999; Slack, 1994). This paper follows Abalo et al. (2006) in using a partition that combines the quadrant and diagonal-based schemes, enlarging the top left quadrant of the original Martilla–James partition so the new area occupies the whole of the zone above the diagonal on which importance is equal to performance (Fig. 2). With this partition, any attribute with an importance rating greater than the corresponding performance rating is a candidate for improvement efforts. Moreover, the greater discrepancy between the importance and the performance for an attribute, the greater the need for remedial action, because of the assumption that a larger discrepancy causes greater dissatisfaction (Sethna, 1982). The interpretation of the areas below the diagonal is the same as in the original Martilla–James diagram.

3. Importance measures

The second major difficulty facing IPA, and the focus of this paper, is the measurement of importance and performance. In this regard, performance has generally been less controversial than importance: the usual measurement procedure has been to take the mean of the performance ratings obtained from an appropriate group of people by means of a metric or Likert scale. However, a variety of means exist to perform importance measurement. In particular, two quite different kinds of measure are common in IPA applications: 1) direct measures based on Likert scale, k-point scale or metric ratings obtained in the same way as for performance; and 2) indirect measures obtained from the performance ratings, either by multivariate regression of an overall product or service rating on the ratings given to the individual attributes (Danaher and Mattsson, 1994; Dolinsky, 1991; Neslin, 1981; Taylor, 1997; Wittink and Bayer, 1994) or by means of conjoint analysis techniques (Danaher, 1997; DeSarbo et al., 1994; Ostrom and Iacobucci, 1995).

Although a recent review of these methods (Bacon, 2003) supports earlier studies (e.g. Alpert, 1971; Heeler et al., 1979) in finding that direct measures capture the importance of attributes better than indirect measures, the former have serious drawbacks. In particular, Bacon (2003) himself emphasizes that direct importance assessment is often misleading because ratings are uniformly high. The main source of this problem is inherent to Martilla and James’ procedure, in which the first step should be to identify the most salient attributes of a product or service by qualitative studies (focus groups and/or unstructured personal interviews) or by reviewing previous research. This procedure has a natural tendency to record high importance ratings on a metric or Likert scale for all the attributes selected for evaluation, with the result that they all crowd together at the top of the IPA grid. This defeats the purpose of the exercise, since the objective of the IPA diagram is not to act simply as a record of absolute importance and performance values, but to discriminate among attributes as
targets for improvement action. Other factors that several studies pointed to as contributory causes of the crowding phenomenon include respondents’ lack of involvement (Bacon, 2003) and their possible lack of expertise regarding the product or service assessed (Sambonmatsu et al., 2003). However, the main cause is the use of measures of absolute rather than relative (competitive) importance.

The development of indirect measures of importance helps to circumvent the above problems. Among these indirect measures, standardized regression coefficients obtained by multivariate regression of the attributes performance ratings over the overall performance rating given to the product or service. This is intrinsically a measure of relative importance in that each regression coefficient depends on data for all the attributes. This is, in fact, a more realistic conception of human reasoning and psychology of choice (Edwards, 1954; Tversky and Kahneman, 1981). This method has the further advantage of reducing the demands on the rater’s attention (since only the performance scale is available, not importance), which may favor increased rater involvement. However, this approach also has at least one major weakness: the possibility of collinearity (Danaher, 1997; Bacon, 2003). Collinearity among the attribute performances when used as predictors of overall performance can lead to the precision of the regression coefficients being so poor that they fail to discriminate reliably among the attributes (and can even take on negative values).

A second criticism of linear regression coefficients as importance measures is that the relationship between the overall performance of a product and its performance with respect to its individual attributes may well be nonlinear (Danaher, 1997). Specifically, some research reports that the negative effect of below-average attribute performance on the overall product rating is greater than the positive effect of above-average attribute performance (Mittal et al., 1998; Sethna, 1982). Generalization of the “regression coefficient approach” to this situation would mean fitting a response surface to the performance data and using partial derivatives as “local” measures of importance. Although in some situations this procedure might provide the decision-maker with useful information, this procedure seems unlikely to be generally applicable.

Another major group of indirect importance measures comprises those derived by conjoint analysis. This approach has the advantage over multiple regression of using orthogonal designs, which excludes the possibility of collinearity between attribute partial utilities. Furthermore, the use of several levels for each attribute lessens the risk of difficulties due to non-linear dependence of overall performance on attribute performances. However, conjoint analysis requires quite an onerous data collection process and becomes unfeasible when involving more than a very few attributes, both of which are especially serious drawbacks for the application of IPA to services.

Partial correlation or logistic regression methods have difficulties similar to those of the above approaches to importance measurement, or they have equally negative consequences (Crompton and Duray, 1985; Danaher and Mattsson, 1994; Heeler et al., 1979; Jaccard et al., 1986). Moreover, the various measures proposed generally disagree with each other (Alpert, 1971; Bacon, 2003; Crompton and Duray, 1985; Heeler et al., 1979; Jaccard et al., 1986; Matzler et al., 2003; Oh, 2001), so the interpretation of IPA differs according to the chosen importance measure. Thus, the decision-maker has little way of knowing a priori which measure, if any, may be the most reliable at incorporating IPA as a means of designing appropriate marketing strategies.

4. An alternative assessment of importance

The starting point for this paper is the idea that, for the purposes of IPA, asking raters to rank attributes in order of importance (with no ties allowed) rather than assigning them absolute importance values offers the possibility of better measurement of importance. Clearly, this procedure avoids the problems that beset the indirect measures described above, and, for each rater, avoids the problem of crowding, since ranking $k$ attributes automatically spreads their “rank scores” evenly between 1 and $k$. Furthermore, as the evaluation task forces the rater to consider attributes jointly rather than one by one, this procedure has two great advantages by: 1) obtaining a relative or competitive measure of importance for each attribute; and 2) favoring rater commitment and involvement in the evaluation task, especially by asking raters only to rank their top few $k$ preferences among the $s$ attributes under consideration, which reduces risk of fatigue. However, before the application of this method, the researcher must settle the question of how multiple raters’ ranking reports should aggregate in the IPA grid.

If every rater ranks all the $s$ attributes under consideration, with rank 1 being low and rank $s$ high, the mean rank assigned to an attribute may be a reasonable measure of its importance. However, this first procedure has three major drawbacks: 1) asking for only the top few $k$ preferences can leave many attributes with low aggregate rank scores, resulting in crowding at the bottom of the IPA grid; this effect will increase with greater numbers of attributes; 2) this procedure omits the reported orders of preference listed by each rater; 3) comparison with performance mean values would be difficult in both quadrant and diagonal approaches to IPA.

Previous authors adopted a ranking approach in their IPA research (Matzler et al., 2003; Mersha and Adlakha, 1992; Sampson and Showalter, 1999), but they ignored the whole number of attributes that compose the evaluation task, and/or the rankings given by raters, using as their measure of importance the absolute frequency for each attribute reported by all raters or the mean rank of the ordered $k$ attributes.

The measure proposed in this paper gets round those problems by subjecting appropriately scaled mean ranks to a transformation that depends on the proportion of attributes that the rater must rank. The remainder of the paper describes the scaling and transformation operations and the use of the new measure for importance-performance analysis of a primary health care service and then compares the results of this analysis with those of an earlier study based on a direct measure of importance.
5. Method

5.1. Importance measurement

Suppose \( n \) raters presented with \( s \) attributes rank their top \( k \) preferences using natural numbers from 1 (most preferred) to \( k \) (least preferred), with no ties allowed. The problem is to use these rankings to assign each attribute \( j \) an importance value \( p_i \), lying in some specified interval that, for convenience, is here \([0,1]\); the value of \( p_i \) should increase with the importance of attribute \( i \).

Denoting by \( g_{ij} \) the rank assigned to the \( i \)-th attribute by the \( j \)-th rater, the first thing to do is to recode the \( g_{ij} \) as ranking scores \( h_{ij} \) that lie in the desired interval, increase with degree of preference, rather than decrease, and assign the value 0 to all attributes not mentioned by rater \( j \):

\[
h_{ij} = \begin{cases} \frac{(k-g_{ij}+1)}{k} & \text{if } g_{ij} \not= 0 \\ 0 & \text{otherwise} \end{cases}
\]

(1)

Table 1 lists \( h_{ij} \) for values of \( k \) and \( g_{ij} \) up to 9.

For the reasons noted at the end of the Introduction, the mean of the \( h_{ij} \) over raters, \( m_i \), tends to crowd attributes together at the bottom of the importance scale. This tendency is even greater, the smaller the proportion of attributes ranked by each rater, \( k/s \). Thus, this value is not attractive as a measure of importance. However, this difficulty can be overcome by subjecting the \( m_i \) to a monotonic transformation that increases small \( m_i \) while leaving large \( m_i \) relatively unaltered. In view of the relationship between the likelihood of crowding and \( k/s \), a particularly suitable transformation of this kind is \( m_i \rightarrow (m_i)^{k/s} \). This procedure increases discrimination between attributes and augments the diagnostic and strategic properties of IPA. The transformed importance measure proposed in this paper is therefore:

\[
P_i = \left( n^{-1} \sum_j h_{ij} \right)^{k/s}
\]

(2)

6. Application: IPA of a primary health care service

6.1. Attributes

In a previous study of the public primary health care service in Galicia, Spain, Abalo et al. (2006) identified eight essential components of the service on the basis of the existing literature and focus groups with health care staff and patients. The present study includes a further factor excluded in the earlier study: health center opening hours, making nine in all (Table 2).

6.2. Data collection procedure

The data for application in this research came from 1767 patients recruited at 75 primary health care centers from among all patients aged 18 years or older, respecting proportionality as regards age groups and sex. Eight trained interviewers collected the data for a month. For the purposes of the evaluation of service performance, every rater scored from 0 (very bad) to 10 (very good) all of the nine attributes considered, and the mean of the raters gave aggregate data. On the other hand, the aggregate importance values of the nine attributes came from the new method proposed in this paper. To this end, raters ranked what they considered to be the three most important attributes in order of importance, from 1 (most important) to 3 (least important). To minimize the likelihood of bias due to the order of presentation (Ares, 2003; Tversky and Kahneman, 1982, 1983), each interviewer presented the nine attributes in a different order. Both the performance evaluation and the importance ranking questionnaires formed part of a larger questionnaire on the perceived quality of primary health care in Galicia.

7. Results and discussion

Table 3 lists the number of ranks 1, 2 and 3 assigned by the raters for each attribute, and Table 4 lists the aggregate...
importance and performance value of each attribute together with the difference between the two, where the determination of aggregate importance is Eq. (2) multiplied by 10, in order to achieve the same rank scale as the performance values. For comparison, Table 5 lists the performance and importance values obtained in the earlier study (Abalo et al., 2006), in which the importance measure of the eight attributes considered are the averages of ratings on a scale of 0–10 from 1601 subjects, obtained in the same way as the performance ratings.

In the earlier study, aggregate importance values squeezed into a rank of only 1.66 units (from 7.34 to 9.00). Although some groups of attributes are different from the others (because the wide sample and subsequent little standard error, Table 5 and Fig. 3), the uncertainty in the values is large enough to be able to use them for effective discrimination among the attributes within these groups. By contrast, the importance measure used in the present study successfully spread the aggregate importance scores over a range of more than 6 units, from 2.86 to 9.05.

Fig. 4 presents the aggregate importance values obtained in the present study plotted against the corresponding aggregate performance values on an IPA grid of the type described in the Introduction. Fig. 5 shows the corresponding diagram for the earlier study (Abalo et al., 2006), where aggregate importance and performance both clustered around high values, crowding the attributes into the upper right corner of the IPA grid. In fact, since in all cases importance exceeds performance and all the attributes lie above the diagonal, a strict interpretation of the diagram would suggest that all the attributes are good candidates for improvement measures. This is somewhat counterintuitive, given that the average performance value is 7.23.

The IPA grid obtained in the present study (Fig. 4) clearly discriminates better among attributes than that of the earlier study (Fig. 5). However, although its importance values show a good spread, its performance values, like those of the earlier study, crowd into a much narrower range (less than 3 units, from 5.66 to 8.58, in the present study). This is in keeping with the usual tendency observed in studies of user satisfaction with health services that have employed questionnaires soliciting quantitative ratings (Fitzpatrick, 1991; Hawes and Rao, 1985). In other words, direct performance ratings in the health service

### Table 4

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Performance</th>
<th>Importance</th>
<th>Discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical staff</td>
<td>8.58</td>
<td>9.05</td>
<td>−0.47</td>
</tr>
<tr>
<td>Nursing staff</td>
<td>8.40</td>
<td>5.82</td>
<td>2.58</td>
</tr>
<tr>
<td>Auxiliary staff</td>
<td>7.56</td>
<td>4.07</td>
<td>3.49</td>
</tr>
<tr>
<td>Ease of appointment</td>
<td>6.80</td>
<td>5.91</td>
<td>0.89</td>
</tr>
<tr>
<td>Time in waiting room</td>
<td>5.74</td>
<td>5.64</td>
<td>0.10</td>
</tr>
<tr>
<td>Opening hours</td>
<td>7.17</td>
<td>4.95</td>
<td>2.22</td>
</tr>
<tr>
<td>State of building</td>
<td>7.17</td>
<td>4.54</td>
<td>2.63</td>
</tr>
<tr>
<td>Signage</td>
<td>7.41</td>
<td>2.86</td>
<td>4.55</td>
</tr>
<tr>
<td>Equipment</td>
<td>5.66</td>
<td>7.18</td>
<td>−1.52</td>
</tr>
<tr>
<td>Mean</td>
<td>7.16</td>
<td>5.56</td>
<td></td>
</tr>
</tbody>
</table>

* a Scaled up to occupy the same range as performance values.
  b Performance value minus importance value.

![Fig. 3. Aggregate importance values of health service attributes according to Abalo et al. (2006). Bars show 95% confidence intervals, and frames surround values that do not differ significantly at the \( \alpha = .05 \) level.](image)

![Fig. 4. IPA diagram of a primary health care service constructed using the performance and importance measures and data of this study.](image)
area in general and these two studies in particular exhibit the same crowding tendency discussed in the Introduction in relation to direct importance ratings. Likewise, this phenomenon extends to other fields like the motor industry, in which the performance of a service recently rated from 1 to 10 with respect to six attributes achieved a spread of just 0.61 rating points between the best and the worst mean rating (Matzler et al., 2004).

The result of spreading out importance values while failing to spread performance values, which cluster at the high end of the scale, is that the great majority of attributes fall below the diagonal of the IPA diagram (negative discrepancies: greater performance than importance mean values for each attribute).

Only attributes to which users assign very great importance (in this study, the medical staff and equipment) can appear as candidates for improvement, which often is somewhat counterintuitive, especially when those attributes achieve the best performance mean values (e.g., medical staff).

Therefore, for the purposes of IPA, preference ranking and Eqs. (1) and (2) may often be useful to achieve a good spread among both performance values and importance values. Although the mean of the ratings given directly by users may well be a better performance measure of individual attributes in absolute terms, IPA requires the construction of a diagram that highlights the relative importance and relative performance of the attributes considered. To this end, the transformed rankings approach described in this paper appears to be most appropriate. Unfortunately, the requirements of the sponsor of the present study on primary health care services prevented application of the new approach to performance as well as importance, so the reconstruction of raters’ preferences from their performance ratings was impossible, as ties among attributes is the rule rather than the exception.

8. Conclusion

This paper presents a new method for measuring the importance of product or service attributes for the purposes of IPA. Instead of aiming for exact and “absolute” quantification of raters’ perceptions of the importance of the individual attributes, the new method, defined by Eqs. (1) and (2), summarizes their perception of the “relative” importance of what they consider to be the most important attributes. Moreover, this procedure spreads the attributes over the importance dimension of the IPA grid, thereby improving the readability and utility of this tool for successful marketing planning. The new method thus overcomes a major difficulty faced by IPA based on absolute rather than relative measures of importance, namely, the tendency for all attributes to crowd into the upper part of the diagram because of equally great importance among them. Empirical results obtained by Abalo et al. (2006) using direct importance ratings illustrate this tendency, and the empirical results of the present study show that this problem is indeed surmountable with the new method. By forcing raters to select what they consider to be the most important attributes of the product, and to rank them by order of importance, the new method may also increase raters’ involvement with the evaluation task. Previous attempts to base importance measures on preferences (Matzler et al., 2003; Mersha and Adlakha, 1992; Sampson and Showalter, 1999) have failed to achieve the desired value spread because they have failed to take into account the proportion of ranked attributes and/or the reported orders of preference. Finally, in view of the empirical results of this study, in which performance values obtained by direct rating clustered at the high end of the performance dimension, the new method may successfully measure not only importance but also performance.

References


Taylor SA. Assessing regression-based importance weights for quality perceptions and satisfaction judgments in the presence of higher order and/or interaction effects. J Retail 1997;73(1):135–59.


