Reliance on accounting performance measures: dead end or new beginning?

David Otley*, Alexander Fakiolas

Department of Accounting and Finance, University of Lancaster Management School, Lancaster LA1 4YX, UK

Abstract

A considerable body of literature has developed following Hopwood (Hopwood, A. G. (1972). An empirical study of the role of accounting data in performance evaluation. Journal of Accounting Research Supplement 156–182) studying the consequences of evaluative style on managerial behaviour and performance. However, this literature also displays some confusion and ambiguity concerning both the conceptualization and measurement of evaluative style. The purpose of this paper is to clarify the different instruments that have been used to assess reliance on accounting performance measures and to set out proposals for future development.

Despite the opinion of Brownell and Dunk’s (1991) that “the continuing stream of work devoted to this issue constitutes, in our view, the only organized critical mass of work in management accounting at present”, the extant body of research contains a number of disjointed findings. By and large, there is little consistency in the reported results, perhaps partly explained by the variety of situations in which the topic has been studied, but also potentially accounted for by the variety of ways in which the central variable, evaluative style or RAPM, has been conceptualized and measured. The purpose of this paper is to trace the development of the concept of RAPM and its measurement. This will demonstrate that a variety of concepts and measures lie behind the same words and that apparently comparable results are in fact rather different from each other. It will conclude by suggesting some ways to develop what is still argued to be an important concept which has far-reaching practical implications.
1. Review

1.1. Hopwood’s original study

The development of the concept of RAPM derives from Hopwood’s (1972) work in which he identified three distinct evaluative styles\(^1\) used by senior managers in holding subordinates accountable for their performance. These styles depended upon both the extent and the manner in which budgetary information was used (Briers & Hirst, 1990). The three styles identified, in the context of cost centre managers, were as follows:

A Budget-Constrained [BC] style: Here budgetary information formed a central part of the evaluative process and it was used in a rigid manner, so that reasonable explanations of failure to meet (cost) budget targets were unlikely to be accepted. Thus success in meeting budget targets was the primary criterion for performance to be judged acceptable; conversely failure to meet such targets, for whatever reason, was likely to lead to adverse evaluative judgements.

A Profit-Conscious [PC] style: Here budgetary information was still an important indicator of good performance, but it was used in a more flexible manner and seen as just one indicator of a longer-term concern with costs and efficiency. Thus it would be possible to be positively evaluated despite showing a cost over-run against budgetary targets provided that there were reasonable explanations for such an outcome, and that action was being taken to ensure adequate long-term performance.

A Non-Accounting [NA] style: Here budgetary information was seen as being of secondary importance and performance was evaluated by reference to other information. Although this category was defined by Hopwood and a considerable proportion of managers in his study (44%) reported its use, it was not used in the further analysis he reported which concentrated solely on the difference between the BC and PC styles. The decision not to examine the consequences of the NA style was taken largely because it became apparent, on examining the characteristics of managers included within it, that it was not an homogeneous category, but rather a residual category into which an assortment of different styles were grouped.

Finally, a further style, the Budget-Profit [BP] style was also used in Hopwood’s analysis which represented an intermediate style between the pure BC and PC styles. This was essentially an artefact of the method of measurement used, and was reported by only a minority (10%) of managers surveyed.

Hopwood therefore reports results where managers perceive that budgetary information plays an important part in the evaluation of their performance, and distinguishes between a rigid, short-term orientation and a more flexible, longer-term use of budgetary information. In both cases, extensive use is made of budgetary information in performance evaluation; the two styles differ in the way in which that information is used in coming to an overall assessment of managerial performance.

1.1.1. Measurement of evaluative style

Hopwood measured evaluative style on the basis of a questionnaire completed by subordinate managers who reported on the evaluative style they perceived their superior to use. The development of this questionnaire was preceded by extensive interviewing of managers and incorporated two key phrases that were used within the particular organization studied (a single U.S. company in the steel industry) that were judged to encapsulate the two major styles to be studied. The first phrase was “meeting the budget” which was associated with a rigid style of performance evaluation which emphasized short-run results. The second was “my concern with costs” which was thought to indicate the more flexible style of evaluation which has a longer-run focus. These two items were incorporated with six other possible factors influencing a manager’s evaluation (e.g. “how well I get along with my boss”) and administered using a survey methodology. These questionnaire items, together with variants used by subsequent researchers, are presented in Fig. 1.

---

\(^1\) Or, perhaps more properly, two evaluative styles and a residual category which includes a variety of potentially distinct approaches.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. How well I cooperate with colleagues</td>
<td>How well I cooperate with workers at my level in the organization [peers]</td>
<td>How well I cooperate with colleagues</td>
<td></td>
</tr>
<tr>
<td>b. My concern with costs</td>
<td>How efficiently I run my unit</td>
<td>My long-run concern with costs and revenues</td>
<td>My long-term concern with costs and revenues</td>
</tr>
<tr>
<td>c. How well I get along with my boss</td>
<td>How well I get on with Area [corporate] staff</td>
<td>How well I get along with my immediate boss</td>
<td>How well I get along with my superior [district manager]</td>
</tr>
<tr>
<td>d. How much effort I put into my job</td>
<td>The effort I put into my job</td>
<td>How much effort I put into the job</td>
<td>How much effort I put into my job</td>
</tr>
<tr>
<td>e. My concern with quality</td>
<td>My concern with quality</td>
<td>My concern with quality</td>
<td>My concern with quality of service</td>
</tr>
<tr>
<td>f. Meeting the budget</td>
<td>How well I meet my budget [financial targets, quotas etc.]</td>
<td>My ability to meet budgeted targets in the short-run</td>
<td>Meeting branch [unit] targets in the short-term</td>
</tr>
<tr>
<td>g. My attitude towards my work and company</td>
<td>My attitude towards my work</td>
<td>My attitude towards my work</td>
<td>My attitude towards my work</td>
</tr>
<tr>
<td>h.</td>
<td>My attitude towards my company</td>
<td>My attitude towards my company</td>
<td></td>
</tr>
<tr>
<td>i. My ability to handle my men [work force]</td>
<td>The relationships I have established with my staff and men</td>
<td>My ability to handle my men [subordinates]</td>
<td>My ability to handle my subordinates</td>
</tr>
<tr>
<td>j.</td>
<td>How much profit I make</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k.</td>
<td>How well I cooperate with individuals outside the firm (e.g. suppliers, customers)</td>
<td>How well I cooperate with individuals outside the branch (e.g. customers and clients)</td>
<td></td>
</tr>
<tr>
<td>l.</td>
<td>How well I cooperate with other [bank] managers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m.</td>
<td>How well I perform on my branch inspection</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. RAPM instrument items.
Respondents, who were responsible for the management of cost centres in a largely sequential production process, were firstly invited to attach importance scores to each of the eight questionnaire items using a 5-point Likert scale, each point being anchored by the phrases “very important”, “quite important”, “of some importance”, “of little importance” and “of no importance”. Then respondents were invited to look back over all the items so graded and rank the three most important factors in order of their importance by attaching the ranks “1”, “2” and “3”. This procedure was incorporated because of previous problems encountered by Hopwood’s supervisor in interpreting results based on absolute importance scales in Michigan-based psychological studies. In the event, this proved to have been an important precaution to have taken. Hopwood’s reports of his results are not totally explicit on this point, but it appears that analyses based on the importance scores failed to produce differences between the two groups at adequate levels of statistical significance to substantiate the prior hypotheses. However, if the rank information was used, it was possible to produce more marked contrasts, as had been anticipated by the experienced supervisor. To do this, the rank information was used as follows:

<table>
<thead>
<tr>
<th>Rank of ‘Concern with costs’</th>
<th>Rank of ‘Meeting the budget’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>PB</td>
</tr>
<tr>
<td>2</td>
<td>BP</td>
</tr>
<tr>
<td>3</td>
<td>BP</td>
</tr>
<tr>
<td>4</td>
<td>BC</td>
</tr>
<tr>
<td>5</td>
<td>BC</td>
</tr>
<tr>
<td>6</td>
<td>BC</td>
</tr>
<tr>
<td>7</td>
<td>BC</td>
</tr>
<tr>
<td>8</td>
<td>BC</td>
</tr>
</tbody>
</table>

If the item “meeting the budget” was ranked in the top three ranks given, but the item “my concern with costs” did not appear in these top three ranks, then the style was categorized as Budget-Constrained.

If the item “my concern with costs” was ranked in the top three ranks given, but the item “meeting the budget” did not appear in these top three ranks, then the style was categorized as Profit-Conscious.

If neither of the above two items appeared in the top three ranks given, then the style was categorized as Non-Accounting.

If both items appeared in the top three ranks given, the style was categorized as Budget–Profit.

Fig. 2 illustrates the way in which the ranking data was used to generate the style categories.

Hopwood’s reported results are based on these categorizations. Specifically, greater levels of job-related tension and dysfunctional behaviour were observed in managers reporting a BC style than in managers reporting a PC style. This was primarily because the budgetary information captured only part of the information necessary to make an overall assessment of managerial performance in the interdependent situation studied. However, it is important to note both the specific method used to categorize evaluative style and the fact that the
phrases used in the questionnaire were developed from interviews conducted within the specific organization studied. The central feature of the measurement techniques used was to construct a contrast between a rigid evaluation based on meeting budgetary targets in the short-term, and a more flexible evaluation based on longer-term factors associated with efficient operation. It is not entirely clear how important budgetary targets are in the latter style, but the measurement method requires that they be ranked below the third most important criteria (except in the intermediate Budget–Profit style).

1.2. Other studies—group 1

In this first group of other studies to be reviewed we consider those studies which use a very similar method to measure evaluative style to that used by Hopwood. The results of these studies can therefore be considered to be the closest replications of Hopwood’s original work.

The first study of this group was conducted by Otley (1978) in a single organization specifically selected to avoid the mismatch between accounting performance measures and overall performance. He thus chose an organization having independent profit centres as production units, namely the nationalized coal-mining industry in the UK. Because the operation of each coal mine was largely independent of other mines, and because the coal produced was sold to commercial customers in a marketplace, the profitability and other performance measures relating to each coal mine can be evaluated in isolation. Thus, if Hopwood’s results were driven by an inappropriate match between accounting data which assumed independence and an operating reality which was highly interdependent, Otley would not expect to replicate them in his study. As well as administering a pilot-tested questionnaire, he also conducted interviews within his selected organization, both with senior managers and also with his complete sample of coal mine managers who formed the focus of his research. He used an eight-item scale similar to that used by Hopwood, but with the following amendments, as shown in Fig. 1:

1. The item “how well I co-operate with colleagues” was omitted as co-operation with peers, as distinct from subordinates or superiors, was not a major factor in the activities of a manager of a coal mine which operated as an independent profit centre geographically separate from other coal mines.

2. The phrase “how efficiently I run my unit” was substituted for the item “my concern with costs” as this phrase was used by the managers concerned to indicate longer-run efficiency as distinct from meeting short-term budgetary targets.

3. The item “how much profit I make” was added, given that coal mines were treated as profit centres in accounting terms. Perhaps fortunately, this item did not prove important in the ranking data, and so had no effect on the categorization procedure. In retrospect, the item was problematic in that it failed to distinguish between absolute levels of profitability and the attainment of a budgeted profit target.

4. Some minor changes in wording were incorporated, so that the questionnaire items would be readily understandable by the recipients. It was not thought that these changes were significant.

The categorization procedure followed that of Hopwood exactly, except that the Budget–Profit category was split into two [BP and PB] depending upon whether the budget or efficiency item was ranked more highly when both appeared in the top two ranks. In contrast to Hopwood’s results, only one manager was categorized as perceiving a Non-Accounting style (see Table 1), and Otley’s analysis is therefore reasonably comparable with Hopwood’s in concentrating only on accounting-based styles of evaluation although the styles are represented by more of a continuum than a dichotomy. It is of interest to note that Otley’s managers reported to just three main superiors, but reported a range of evaluative styles; either superiors changed their style from manager to manager, or managers’ perceptions of the same style differed. Otley presents some evidence to
suggest that other antecedent factors (such as the expected profitability of the operating unit) appeared to affect the evaluative style used by the superior. In any event, Otley’s results did not replicate those of Hopwood, so he concluded that Hopwood’s results were driven by the technical inadequacies of the accounting system as a means of performance evaluation in the interdependent cost centres studied by Hopwood.

The next major study in this group was conducted by Brownell (1982). This was based on cost centres so Brownell used Hopwood’s original scale, and the ranking methodology in order to construct categories. However, he then merged the BC and BP styles into a single category he called “high budget emphasis” and the PC and NA styles into a category he named “low budget emphasis”. He partly justified this on the basis that the distribution of respondents across these two categories matched that found by Hopwood very closely (see Table 1). It should be noted, however, that although the questionnaire items and the ranking method was the same as Hopwood’s, the data now includes all respondents reporting the NA style, and the results cannot therefore be directly compared with those of Hopwood and Otley.

However, this is more than a difference of measurement as it represents a change in the underlying concept being measured. Brownell (1982) does not define a budget-constrained style in the same way as Hopwood, but rather develops a concept of evaluative style where “the primary emphasis in the evaluation of subordinates is on budget achievement”. This is clearly true of the high budget emphasis style, given that meeting the budget is ranked in the top three criteria of evaluation. In fact, the categorization process adopted reduces to whether the item “meeting the budget” appears in the top three items ranked (high budget emphasis) or not (low budget emphasis). The responses to the other questions, including “concern with costs” are irrelevant. Given a concept of budget emphasis is being developed, it clearly is consistent to think of this dimension as a continuum rather than a set of categories, although in this particular study it is actually analysed as a dichotomy (i.e. high and low budget emphasis). This may explain why subsequent work was more inclined to use continuous measures. Finally, the idea of relative budget emphasis was later developed by Harrison (1992), (1993).

Brownell and Hirst (1986) also mainly followed Hopwood’s questionnaire items, as they also studied cost centres, but made the modifications shown in Fig. 1. In particular, the words “short-run” and “long-run” were added to the two relevant items, and two additional items were added to the questionnaire making it ten items in length. Because of the additional length of the questionnaire, Brownell and Hirst decided to base the categorization into styles on the top four ranks, rather than the top three ranks. However, it is not clear why this is theoretically justified as additional items should not affect the ranking of the three most important. Finally, they pooled the categories in the same way as Brownell (1982) to produce the high and low budget emphasis categories. The same procedure was adopted by Hirst (1987), using the same ten-item questionnaire, but on this occasion he used only the top three ranks to perform the categorization.

Finally, Pope and Otley (1996) undertook a study based on branch bank managers in the U.K. which were treated as profit centres. A slightly amended version of the Brownell and Hirst (1986) instrument was used, with managers being categorized into styles using the top three ranks, enabling both Hopwood’s and Brownell and Hirst’s categories to be used. In addition, this study attempted to develop a further, more direct measure of evaluative style, which will be discussed later.

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>BC</th>
<th>BP</th>
<th>PB</th>
<th>PC</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopwood (1972)</td>
<td>167</td>
<td>20</td>
<td>10</td>
<td>26</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Otley (1978)</td>
<td>39</td>
<td>13</td>
<td>18</td>
<td>38</td>
<td>28</td>
<td>3</td>
</tr>
<tr>
<td>Brownell (1982)</td>
<td>38</td>
<td>21</td>
<td>24</td>
<td>24</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Brownell and Hirst (1986)</td>
<td>76</td>
<td>17</td>
<td>41</td>
<td>7</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Pope and Otley (1996)</td>
<td>216</td>
<td>72</td>
<td>20</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Hirst (1987)</td>
<td>44</td>
<td>32</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this group, the studies outlined have been closely based on the Hopwood questionnaire and categorization method. But even here, variations in wording and the different number of ranked items used may have affected the categorization process. More importantly, the amalgamation of the NA style and the PC style changes the specification of the comparator groups and prevents accurate comparisons of studies using these different schemas. It should also be noted that the relevance of the questionnaire items to the managers to whom they were administered was not explicitly reported upon in some cases, and there is therefore no evidence that the items had the same meanings attributed to them by the managers surveyed. Finally, although the questionnaire items and ranking procedures used here followed Hopwood’s original work quite closely, the concept being studied diverged significantly from it. This has caused considerable confusion in the later interpretation of the reported results.

1.3. Other studies—group 2

In this second group, the studies included use similar questionnaire items to those originally developed by Hopwood. However, the scoring method here does not use the ranking data as in the first group, but rather uses the Likert importance scores.

Brownell (1982) had noted that many of the respondents in his field study failed to complete the ranking question appropriately. To overcome this problem, Brownell (1985) decided to use only the importance scores. But rather than contrasting the importance ratings given to the two key items, he decided to aggregate them into a single overall importance score. That is, the importance scores for “meeting the budget” and “concern with costs and revenues” were added to give a single score in the range 2–10. He suggested that the legitimacy of this procedure depended on the existence of a significant positive correlation between the responses to the two items, which in his study was $r = 0.73 \,(p < 0.01)$. But this is more than the use of an alternative method of measurement for the same construct; rather, an entirely new concept is being advanced. As hinted at in Brownell (1982), Brownell (1985) is now concerned with the absolute importance being attached to accounting performance measures and the dysfunctional consequences of high levels of importance being attached to accounting information in inappropriate circumstances.

In his published paper Brownell (1985) does not state whether his units were cost or profit centres, but it appears that they were profit centres and that the item “how much profit I make” was used in place of “my concern with costs and revenues”. Whatever the case, the summation of the two items gives a measure of the absolute importance of accounting information in performance evaluation, rather than the contrast implied in the original Hopwood measure. The same probably applies to the Brownell (1987) study where it is reported that the same measure was used as in Brownell (1985).

This summation technique has been adopted in a number of subsequent studies. Brownell and Dunk (1991) and Dunk (1989a,b) used Hopwood’s (1972) wording and then followed Brownell (1985) in summing the importance scores of the two relevant items. Unfortunately, they failed to concern themselves with the legitimacy of such a procedure by not reporting the inter-correlation of the two items. In a further study, Dunk (1990) pursued the same approach whilst reporting a correlation of only $r = 0.325 \,(p < 0.01)$ to justify the addition.

Dunk (1989a,b) was explicit about how he proceeded. He added the importance scores of the two relevant items (b and f in Fig. 1) to create a scale with values from 2 to 10. He dichotomized his scale at its mean to indicate accounting-based styles and non-accounting-based styles. However, he then referred to the two categories he had created as representing high and low reliance on

---

2 In a private communication to the author (28 May 1990) Brownell writes: “the item included in the (company) questionnaire ‘concern with profit’ was used because... the (company) people weren’t happy with the double-barrelled ‘costs and revenues’, indicating that, since both units studied had responsibility for each, I should use ‘profit’ instead. I suspect this is the ‘situation change’”. As for the ‘costs and revenues’ referred to in the published paper, Brownell suggests this was an editorial error that he failed to catch.
budgetary control. This is clearly a different usage of the term to that in the previous literature which refers to a budget-constrained style. To the extent that importance scores reflect ranks, the Brownell (1985) and Dunk (1989) logic does not map onto Hopwood’s schema. This is shown graphically in Fig. 3. Low values on both items represent a NA style and give a low overall score. High values on both items tend to represent a BP style and give a high overall score. In order to be categorized as a BC style requires one item to be high and the other lower; similarly a PC style requires the other item to be high and the former to be low. Both the BC and PC styles would generate an intermediate absolute score consisting of one high and one lower score. Thus, in Brownell’s (1985) measure the BP style falls at an extreme, and not between the BC and PC styles as expected. The only style of evaluation this instrument safely replicates is the NA style at the low end of the range.

The same comments apply to Harrison’s (1992, 1993) development of the approach, although it may have merit in its own right. Harrison followed Brownell (1985) in summing the importance scores of the two relevant scale items; however, he then divided this score by the sum of the importance scores of the other eight items on the scale. He also multiplied the result by four, to standardize the score, which has a value of unity when budgetary and non-budgetary criteria have the same average importance.

In summary, the studies presented in this group have used RAPM in a substantially different manner to that used in the studies presented in group 1. This “modification” of the original instrument constitutes a change of such an order that it represents not only a difference in measurement but also a difference in concept. The failure to note this difference has caused some confusion in the literature. It has caused the same terminology [RAPM] to be used for the different concepts, and has provided results which cannot be directly compared with the studies in group 1. Perhaps the problem with this body of work is not what has been done with a new concept, but with researchers uncritically using Hopwood’s original rationale to motivate their own work.

1.4. Other studies—group 3

Hirst (1983) and Hirst and Yetton (1984) argued that Hopwood’s (1972) instrument was specifically designed for manufacturing settings and was unsuitable for use in their study which encompassed non-manufacturing environments. They therefore developed a new five item scale, using a five-point Likert importance scaling, shown in Fig. 4. As can be seen from the scale items, this scale measures the degree to which performance is assessed using quantitative data. Although budgets are not explicitly mentioned, several of the scale items would obtain high scores if budgets were extensively used in performance evaluation. Both studies were based on the same data set and used a reported Cronbach alpha for the scale of 0.76 to argue that the instrument captures a single construct. They suggested that their new instrument measures a superior’s reliance on accounting information, and thus introduced the term “RAPM” into the literature, which has subsequently become used in a way that is interchangeable with “evaluative style”.

It is important to note, however, that this instrument differs substantially from both Hopwood’s original concept (a contrast between rigid and flexible use) and Brownell’s adaptation of it (a sum of two accounting-related items). Specifically, it focuses on quantitative information which may

<table>
<thead>
<tr>
<th>Score on Brownell scale</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likely style on Hopwood categorization</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;-</td>
<td>BC+PC</td>
<td>-&gt;</td>
<td>BP</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BP</td>
</tr>
</tbody>
</table>

Fig. 3. Schematic allocation of styles according to Brownell’s (1985) measure.
be both financial and non-financial in nature. Further, although Hirst suggests that Hopwood’s scale cannot be used in a non-manufacturing setting, this is not clear from examining the items included in it. Indeed, the extended scale developed in the Pope and Otley (1996) study was used with branch bank managers and uses basically identical items with two additions. Thus, although the Hirst instrument may have a value in determining the use made of quantitative performance measures and their effects, it is not a measure of reliance on accounting data specifically, and is therefore not directly comparable with the previous work.

1.5. Other studies—group 4

Another approach to the measurement of RAPM was developed by Govindarajan (1984) and Govindarajan and Gupta (1985). They used a single item instrument to measure the relative reliance on objective versus subjective approaches to performance appraisal and the determination of an incentive bonus. Each respondent was required to indicate one of the following choices:

(a) whether his superior used exclusively subjective criteria in evaluating his performance and in deciding his bonus; or
(b) whether his superior evaluated him solely on meeting targeted levels of financial performance; or
(c) whether his bonus was decided partly in a subjective manner and partly using a formula-based procedure; if so, a percentage breakdown of the importance of each factor.

If the objective, formula-based approach were to use budgetary and other accounting data, as is likely often to be the case, then there is a similarity between this approach and that taken by Brownell. Choice (a) would indicate a budgetary-based style of evaluation; choice (b) would indicate a non-accounting style; and choice (c) would place managers on the continuum between the two extremes. However, the instrument is focused upon bonus determination, which may be central in some managerial cultures, but is completely absent in others. Also, being a single item question there is no means of assessing its internal reliability.
2. Discussion

It is clear from the above discussion that the measures used in each group are rather different in their scope and intention, and that they are in no sense interchangeable with each other. Early studies tended to choose between the Hopwood and Otley variants of the original scale depending upon whether they were to be used in a cost or profit centre environment. But it needs to be emphasized that both Hopwood and Otley designed their scales to reflect the terminology in use within the particular companies they studied, rather than just cost or profit centres in general. This aspect of validation appears to have become neglected in their subsequent use.

This point is of particular importance in using the Otley (1978) variant. Firstly, the phrase "how efficiently I run my unit" is quite ambiguous. Although it may refer to a long-run concern with costs and revenues (and appeared to have this connotation in Otley’s organization) it could also be interpreted in other ways. Secondly, the item "how much profit I make" was not an important response in Otley's study, but could easily be interpreted as either a PC response or a BC response in different circumstances. In the case where it was used, it was seen as a non-budget-based response indicating a concern with absolute profitability which was of relatively little importance compared with relative performance against pre-determined budget targets. Indeed, it is precisely this point which appears to have influenced Brownell (1985) to replace "long-run concern with costs and revenues" with "concern with profit".

Brownell and Hirst (1986) made a serious effort to prepare an instrument which could be used for both cost and profit centres, and enriched the wording of the two central Hopwood items by including reference to short-run and long-run concerns. This is consistent with Hopwood’s original ideas, for he makes several references to this time factor difference (e.g. Hopwood, 1972, p. 171 and 1974, p. 110). Their modification can be used to contrast a short-run concern with meeting the budget [BC style] with a longer-run concern with costs and revenues [PC style], as indeed they used it in their own study. However, in common with Brownell (1982) and Hirst (1987), they collapsed the Hopwood categories into just two: "high budget emphasis" and "low budget emphasis". This procedure may have provided roughly equal numbers of managers in the two categories for an efficient comparison, but no theoretical justification of the procedure was made. Again this is in contrast to Hopwood’s original work, where his major comparison was between the BC style (sometimes with the BP style included within it) and the PC style; the NA style was not studied in any detail, a decision which in retrospect he regretted.

The method used by the authors in group 2, beginning with Brownell (1985) also has little explicit theoretical grounding, except insofar as the two items aggregated involve the use of accounting information (costs and budgets respectively). This may be justified where there is a high correlation between the two items, but again this correlation is not always reported in later studies. The most interesting use of this procedure is that used by Harrison (1992, 1993) in comparing the relative importance of accounting versus non-accounting information in performance evaluation. But it is clear that, although the same questionnaire items may be used, this method produces data relating to a very different underlying concept.

One quirk that deserves some attention concerns the internal reliability statistics for the complete set of items in the Hopwood questionnaire. Although these items are not intended to provide an additive scale, nevertheless some authors have provided Cronbach alphas for the whole set of items, and we have been able to compute others from data supplied by other authors who have used the scale, with the results shown in Table 2. The most notable feature of this table is the relatively high Cronbach alphas obtained in all cases (with the possible exception of Brownell, 1985). Given that the scale is expecting to find different relative importances for each of the items included, it is worrying to find such a high internal consistency. Is it that respondents are tending to answer high (low) across all items; or is because the Cronbach alpha statistic appears to be insensitive to such variations? If the items do represent
a single underlying construct, what is it? It may be this feature of the “scale” that caused Hopwood to utilize only the ranking data, having failed to find adequate discrimination using the importance scores. The schematic allocation of Hopwood’s styles to Brownell’s absolute scale shown in Fig. 3 is justified by the data shown in Table 3 where Brownell’s (1985) data is categorized in both ways. This shows that the highest scores are associated with the BP and PB styles, and the lowest with the NA and BC styles, indicating that the two scales measure substantially different concepts. A similar conclusion can be drawn from the modest value of Kendall’s tau for the correlation between the two scales.

A more appropriate methodology would be to use factor analysis to determine whether scale items in a multi-item instrument form a uni-dimensional or a multi-dimensional scale. It may be that respondents tend to respond to all items on a scale in a similar way (the so-called “halo” effect). Halo error will manifest itself in high inter-item correlations (and thus a high Cronbach alpha). Without an analysis of the factor structure of an instrument, a high Cronbach alpha does not necessarily indicate the uni-dimensional nature of a scale. It is significant that such factor analyses have rarely been performed in the RAPM literature. Indeed, the literature tends to lack any systematic attention to the assessment of construct validity; neither is there much coherent attempt to develop better measurement instruments. It would be useful if researchers could support the development of their measurement instruments with field studies and extensive pilot testing of instruments. This would not only ensure that the instrument is relevant to the setting, but would also help clarify ambiguities in wording and enable the collection of rich, qualitative data that could then help explain unexpected results.3

The conclusions to be drawn at this stage are as follows. Firstly, the different measures used in the four groups of study reviewed produce quite different outcomes. Even those using the same instrument differ depending upon whether ranking or importance scores are used, and whether an aggregate measure or a contrast is computed. Secondly, not only are the measures different, they also represent distinct underlying concepts representing different aspects of the use made of budgetary (or financial or quantitative) information by a superior. Thirdly, the original Hopwood measure was carefully developed to capture a phenomenon observed in a specific organization; its later use has tended to apply it uncritically to a range of organizations where its terminology may not capture the same phenomena. Finally, it is not clear that this instrument is the most suitable for gathering the information required to establish the manner in which various types of information are used in an organization. Much more attention needs to be paid to the development of appropriate measurement instruments for the concepts being studied.

### Table 2
Cronbach alpha statistics for Hopwood’s instrument

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Cronbach alpha</th>
<th>Number of items included in instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otley (1978)</td>
<td>40</td>
<td>0.70</td>
<td>8</td>
</tr>
<tr>
<td>Brownell (1985)</td>
<td>60</td>
<td>0.62</td>
<td>8</td>
</tr>
<tr>
<td>Brownell and Hirst (1986)</td>
<td>76</td>
<td>0.77</td>
<td>10</td>
</tr>
<tr>
<td>Pope and Otley (1996)</td>
<td>213</td>
<td>0.85</td>
<td>11</td>
</tr>
<tr>
<td>Fakiolas (1992)</td>
<td>167</td>
<td>0.83</td>
<td>11</td>
</tr>
</tbody>
</table>

### Table 3
Mean scores of Brownell’s budget emphasis instrument in each evaluative style

<table>
<thead>
<tr>
<th>Style of evaluation (Otley’s categories)</th>
<th>BC</th>
<th>BP</th>
<th>PB</th>
<th>PC</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kendall tau correlation with evaluative style</td>
<td>0.054 (p = 0.21)</td>
<td>n = 43</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Budget emphasis</td>
<td>Brownell (1985)</td>
<td>Mean</td>
<td>7.19</td>
<td>8.00</td>
<td>9.00</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(1.40)</td>
<td>(0.82)</td>
<td>(1.00)</td>
<td>(1.67)</td>
</tr>
</tbody>
</table>

---
3 I am indebted to Margaret Abernethy for these suggestions.
3. Development

The former conclusions may seem damning, but the stream of work that has resulted from Hopwood's original study does form a useful basis on which to develop an improved conceptualization of the concept of budget emphasis and reliance on accounting measures of performance. However, a number of distinct dimensions to evaluative style are becoming apparent, and effort should be devoted to constructing better measurement instruments to detect these.

Some of the dimensions of evaluative style that have emerged are as follows:

(a) Hopwood's rigid use of budgets vs a more flexible use.
(b) Hopwood's short-run vs long-run emphasis (which may be a sub-dimension of (a) above).
(c) As above, but referring to any quantitative targets, not just financial budgets.
(d) An emphasis on absolute (quantitative) measures of performance rather than comparisons with pre-set targets.
(e) The relative importance of objective and subjective criteria of evaluation.

There have also been some measurement variants of the above, including ranking, absolute importance ratings and relative importance ratings.

As an example of the development of suitable measurement instruments, the instrument used in the Pope and Otley (1996) study is given in Appendix A. This is not put forward as an ideal type, as it contains some problems of its own, but merely to illustrate a possible way forward. It was developed for use in a study of bank managers and related to quantitative targets of all types (although these were mainly financial) not just budgets. The wording of each scale item was chosen to capture an aspect of Hopwood’s three styles and pilot tested to ensure that it was understood by managers at the level it would be used. In retrospect, the wording “the most important factor in my evaluation” is potentially problematic, as it would be possible to disagree with these items if, for example, targets were a very important but not the most important factor involved. A factor analysis of the above scale clearly indicated a three factor solution, as shown. This was more or less as expected and yielded the following three dimensions:

(a) A flexible and long-run approach to using comparisons of actual performance with targets.
(b) A more single-minded emphasis on meeting targets, with item 10 (“my superior tends to use target variances as a pressure device to emphasise the need to “meet targets”) indicating a possible over-emphasis.
(c) Targets are relatively unimportant, but absolute performance is still important. This may indicate the existence of implicit targets (e.g. positive profitability).

The major unexpected outcome was that item 6 (“my superior is more concerned with actions which produce good short-term results than with long-run effectiveness”) did not load onto the second factor, but rather negatively onto the first. This may indicate a difference between the rigid/flexible dimension and the long-run/short-run dimension. However, it should also be noted that the sample of managers to whom these questions were administered showed a tendency to bunch heavily at the Budget-Constrained end of the spectrum when the Hopwood instrument was used (see Table 1).

Finally, a comment needs to be made on the changes in organizational control practice that have emerged in the last decade that cast doubt on the current role of accounting-based performance measurement techniques. Although the Anglo-Saxon world has traditionally emphasized accounting-based indicators of overall organizational and managerial performance, there are indications that such measures are playing a reduced role (see Otley, 1994 for one account of such changes). Budgeting is being increasingly regarded as a problematic technique with some organizations relegating its use to a financial planning role. The control of manufacturing operations appears to becoming less dependent on budget-based performance indicators. The development of management accounting is also becoming increasingly externally and market oriented. Many new developments come under the
general banner of “strategic management accounting” and a focus on customers, competitors and other external factors is replacing a solely internal focus. Even internally, the most prevalent new development of the past 15 years, activity-based costing, is more concerned with product costing than with internal control. However, the major driver of its development has been the reduced level of direct labour costs relative to overhead costs; this driver also changes the focus of cost control from real-time operations to product planning and design. All of these changes suggest that the roles of budget-based and accounting-based control techniques are likely to be undergoing substantial change in response to these pressures. At the very least, the design of studies of the impact of accounting techniques in the performance measurement and management function should ensure that they are able to pick up such changing emphases rather than expecting the world to behave in the same way as it did in the early 1970s.

But performance measurement and the use of performance measures in performance evaluation are still key management issues. The use of the Balanced Scorecard technique (Kaplan and Norton, 1996) is proving very popular and adopts a wider framework than that of just accounting and financial performance measures. Future studies of the impact of performance measurement and its use by superior managers will need to be sensitive to discovering the frameworks used for performance measurement in target organizations, as was originally done in the Hopwood and Otley studies, rather than just assuming that budgeting and accounting data play a central role in this process. Nevertheless, the methods used in studies on RAPM and evaluative style appear to be generalizable into this arena with only relatively minor adaptation. The door is open for researchers to build upon this foundation and make a central contribution to the management literature.

Appendix A:

[Pope and Otley (1996) evaluative style instrument]

1. My superior emphasizes mostly target-related information in my evaluation.
2. Targets are used with flexibility in my evaluation.
3. How cost efficiently I run my unit is the most important factor in my evaluation.
4. How well I meet my targets is the most important factor in my evaluation.
5. My superior expects me to operate in line with my targets, but allows me to explain reasons for any unfavourable variance.
6. My superior is more concerned with actions which produce good short-term results than with long-run effectiveness.
7. My superior believes that target-related information must be supplemented with other types of information to assess my performance.
8. Target-related information plays a relatively unimportant role in my evaluation.
9. How much profit I make is the most important factor in my evaluation.
10. My superior tends to use target variances as a pressure device to emphasize the need to “meet targets”.

These items were scored on a seven-point Likert scale from Strongly Agree to Strongly Disagree.

A factor analysis (with orthogonal rotation) gave a three factor solution (with eigenvalues > 1 and explaining 55% of the total variance) as follows, with items in bold indicating the highest factor loadings:

Factor 1: Q2, Q5, Q6 (reverse scored), Q7—flexible and long-run use of targets.
Factor 2: Q1, Q4, Q10—rigid use of targets.
Factor 3: Q3, Q8, Q9—non-use of targets, but stress on absolute efficiency.

References


