EXPERT AND LAY OPINION IN THE EVALUATION OF HOUSING

Matthias BURISCH, Dr. phil., Lecturer in Psychology
University of Hamburg, Department of Psychology,
von-Melle-Park 5, Hamburg 13

Introduction
The design and construction of an apartment house is a venture of considerable moment for all parties involved, such as proprietor, architect, contractors, local authorities, and the like. The same is true for the tenant who eventually decides to move in; it is he or she who will have to live with the consequences of that decision. Although the evidence for close and causal relationships between aspects of the built environment on the one hand and measures of subjective well-being on the other is scarce (Mühlch et al., 1978), this may partly be due to inadequate research paradigms used.

That housing looms large not only in people's budget but also in their opinion is documented by Campbell et al. (1976) who found it at rank 4 among 13 domains of life satisfaction in a U.S. survey. Only "family life", "marriage", and "financial situation" were considered more important by their respondents.

In view of findings such as these and the widespread criticism today's architecture is facing, there seems to be some consensus, In West Germany at least, that residents should have a much stronger influence on the shape of their habitats than they have now. When put into practice, however, this proposal runs into trouble soon. In mass housing it is not normally known who will later inhabit the building being designed, nor can anybody predict how long they are going to stay. And this may not even be the key problem. An experimental housing project in Vienna (Kubinger, 1978) left the general lay-out of floor plans largely at the first tenants' decision; this group's satisfaction ratings later failed to significantly exceed those given by a matched comparison group. Similar results obtained in several other instances (Mühlch et al., 1978; Weise, 1976).

Studies such as these may be open to criticism from several angles. But
they are certainly apt to reinforce many architects' belief that "people don't know anyway what they want". And the burden of proof remains with those who recommend "participation" as a panacea.

Where laymen's success at managing their own affairs appears so unimpressive, "expert opinion" may be called for. Such has been supplied in the form of normative evaluation systems in countries such as Austria, the Netherlands, Sweden, Switzerland, and West Germany.

Viewed from the outside, a normative evaluation system is some kind of oracle: you put in a few apartment floor plans plus possibly some information on the entire buildings and their surroundings, out pops an array of scores that tell you, e.g., which is best. Fortunately the systems are fully transparent. Typically they decompose the global "quality" of an apartment into several dozen physical attributes which can be measured or rated more or less objectively. Utility functions are specified which determine the number of raw points corresponding to various levels along the criterion scales. The raw points are then multiplied by importance weights. The sum of these products provides the global evaluation of the apartment.

Obviously this is a very economical method of evaluating housing quality. Since the necessary information is usually available at the preconstruction stage, "bad" buildings need never be erected. There is only one hitch: the systems never ask who is to inhabit the premises; family size is all they want to know about the user. Question to the reader: do you believe that what is good for you is also good for all people with families as large as yours?

Clearly this is partly an empirical question but this small-scale study could only approximate it at best. Since "quality" is such an elusive characteristic, we had to deal with "attractiveness" instead. Do people "want" the same? Do they want what the systems say they should want? If not, is it because their judgments are unstable, i.e., because they do not know what they want? Do building professionals who should know for themselves agree more with the systems than other people do? Finally, do different systems agree? These are some of the questions this paper tries to come to grips with. To this end, two sets of apartment floor plans were evaluated by three normative systems. They were also ranked in order of preference three times by a sample of interviewees. These
different modes of evaluation could then be compared as to their
convergence.

**Normative Evaluation Systems**

Three commercially available systems were selected from a larger number
because they were most recent, comprehensive and operational. A more
detailed discussion of these systems can be found in Burlisch (1972) and
Dierschlag & Leicht (1978).

The "Wohnungsbewertungssystem" (Mollen et al., 1975; abbreviated WBSS)
was commissioned by a Swiss housing authority, mainly for the purpose
of channeling public funds to projects with a good cost-benefit ratio.
Of the WBSS's 75 criteria, 44 pertain to the apartment, 15 to the
building and its immediate surroundings, and 16 to its location within
the larger environment. As an example from the first category, a bath-
room must be at least 1.80 m wide to earn the maximum of 4 raw points
while for 1.60 m it will get only 2 points. The corresponding
importance weight here is 5. Top priority is given to criteria such as
"flexibility of spatial lay-out" (importance weight 38) or "orientation
of windows" (importance weight 32).

A predecessor of the WBSS is contained in a monograph entitled "Wohnung-
bewertung" (Frey et al., 1974; abbreviated WB). All of its 70 criteria
refer to the apartment as such.

A third system was published in Austria by Riccabona & Wachburger
(1971, 1976) under the title "Methoden der Bewertung geplanten/bestehender
Wohnungstypen" (abbreviated MBN). As with the WBSS, the MBN system's
41 criteria take into account attributes of the apartment, the building,
and the environment. The definition and grouping of these,
however, differs markedly from those of the two Swiss systems.

Since the WBSS appears to be the most developed of the three, data
analyses are mainly focused on it.

**Floor Plans**

For interviews 1 and 2 a set of 8 apartments with 3 or 4 rooms, plus
kitchen, bathroom, and at least one balcony were chosen from published
sources. Net sizes varied from 32 to 66 m². For the third round of
interviews an additional 8 floor plans, somewhat larger than the first
Fig. 1. Two Floor Plans from Set 2

ones were produced. All plans, on a 1:100 scale, indicated room sizes, various subtotals, and total size. Two specimens from the second set are shown in Fig. 1. Plan C is the "best" one according to WBS (as well as to WBF and MNW), while plan H ranks second (assuming a 4-person household) or third (assuming a 3-person household).

By accident for the first set the WBS rank order is identical to that of net size. For the second set the rank-order correlation (Spearman's rho) of WBS total and size is only .48 (4-person households) or .19 (3-person households).

In order to restrict our comparison of evaluation modes to more or less "functional" attributes and to eliminate idiosyncratic preferences for certain city districts, orientations, or external facilities, all information as to these was omitted. Consequently, only 34 apartment criteria of the WBS, 18 of the MNW, but all 70 of the WBF entered the normative evaluation.

Respondents
Respondents in Interviews 1 and 2 were 38 couples, each with one or two

I am indebted to Erika Wieland and Peter Eggers for drawing the plans.
children, making up a total of 76 interviewees. The sample included 14 building professionals (architects, construction engineers, building researchers), among whom 1 couple, the professionals' 12 wives, and 50 male and female "laymen". Most were 25 to 40 years old and all could be termed "middle class".

For the third wave of interviews 11 professionals, 9 wives, and 38 lay persons could be followed up.

**Interviews**

Interviews with husbands and wives were conducted separately but concurrently by two interviewers. Each respondent was asked to imagine he or she had to move out and when looking for a new home there were only the 8 apartments available. These plans were then examined in some detail. Frame conditions such as location, age, and--most importantly--rent were assumed to be identical.

Interviewees were then to rank the floor plans in order of preference. In session 2, 4 weeks later, this was followed by a consensus ranking, i.e., each couple had to agree on a common rank order.

The third round took place 1 year after the first one. Individual rank orders were obtained for both the first and second set of plans. Again, a consensus ranking was asked for, but only for set 1. Analyses of all consensus rankings are omitted here; see Burisch (1979).

**Results**

It is time now to return to the questions posed in the introduction. Wherever two rank orders are compared, Spearman's rho is the measure of association. Where agreement within a group is to be assessed, Kendall's coefficient of concordance (see Hays, 1963) is used.

Do people want the same? Table 1 contains coefficients of concordance for the total sample and also for various subgroups. Figures are given for the first and last interviews, respectively, and separately for the two plan sets.

Although all coefficients are highly significant statistically, their numerical value is not very impressive, as was expected. It should be pointed out that even the figure of .59 in the bottom row corresponds to an average interrater rho of only .57. A substantial factor that
Tab. 1. Agreement of Respondents with Each Other

<table>
<thead>
<tr>
<th></th>
<th>Interview 1</th>
<th>Interview 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st Set</td>
<td>1st Set</td>
</tr>
<tr>
<td>All 76 Respondents</td>
<td>.29</td>
<td>---</td>
</tr>
<tr>
<td>All 58 Respondents of Interview 3</td>
<td>.31</td>
<td>.42</td>
</tr>
<tr>
<td>All 24 Respondents of Interview 3 (1 Child)</td>
<td>.32</td>
<td>.43</td>
</tr>
<tr>
<td>All 34 Respondents of Interview 3 (2 Children)</td>
<td>.35</td>
<td>.48</td>
</tr>
</tbody>
</table>

will have served to inflate some of the coefficients is a broad consensus among parents with two children that a 3-room apartment was totally unacceptable to them. There were two such plans in set 2 and no less than 23 interviewees (68% of this subgroup) put them at the bottom of their rank orders. Note that agreement is consistently highest among those with 2 children. (With the first plan set the situation was somewhat more complex as two of the four 3-room apartments made it easy to partition one room.)

Do people want what they “should” want? If, as we have seen, the old saying “one man’s meat is another’s poison” seems to be more true than false with respect to housing, then expert opinion, crystallized in a standard evaluation system, cannot achieve much agreement with individual laymen’s preferences.

Table 2 exhibits mean correlations between individual rankings on the one hand and WBS rank order on the other. Interpreting these results

Tab. 2. Mean Correlations with WBS Ranking

<table>
<thead>
<tr>
<th></th>
<th>Interview 1</th>
<th>Interview 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st Set</td>
<td>1st Set</td>
</tr>
<tr>
<td>All 76 Respondents</td>
<td>.58</td>
<td>---</td>
</tr>
<tr>
<td>All 58 Respondents of Interview 3</td>
<td>.59</td>
<td>.59</td>
</tr>
<tr>
<td>All 14 Building Professionals</td>
<td>.55</td>
<td>---</td>
</tr>
<tr>
<td>All 11 Building Professionals of Interview 3</td>
<td>.52</td>
<td>.61</td>
</tr>
</tbody>
</table>

one should keep in mind that for the first plan set there was a perfect correlation between size and WBS ranking (which, by the way, was the
same for both household sizes). For the second set, a simple ordering according to size would have produced more predictive accuracy than the WBS did; size correlated on the average at .57 with individual rankings (versus .45 for the WBS). In the case of the 11 professionals the figures were .52 (size) versus .49 (WBS).

How stable is laymen's judgment? A simple explanation for the lack of agreement between all these evaluations would be instability of personal preferences. People do change their mind and if they do so excessively any attempt to predict their opinion must necessarily fail. To test this hypothesis, rank orders for the first vs. second, and of the first vs. third interviews were correlated for each individual. Mean rank correlation was .77 (N=76) over the 1-month interval and .68 (N=58) for 1 year. (Incidentally, professionals were only slightly more stable in their judgment-.81, 1 month; .72, 1 year—than the rest of the sample.) This provides little support for the instability hypothesis.

Do experts agree more with "expert opinion" than laymen do? As Table 2 shows, our building professionals, who certainly had considerable experience in reading and evaluating floor plans, did not consistently agree more with the WBS than others did. This may be taken as an indication that even among well-informed and competent evaluators there is ample room for disagreement on what makes a "good" apartment.

Do different systems agree among themselves? If normative evaluation systems are to be taken seriously as guidelines to better housing, then there must be a high degree of agreement, if not unanimity, among different experts' systems. The correlations of the WBS, WBF, and MBW across the floor plans would therefore be of crucial importance. Assuming a 4-person household throughout the WBS and the WBF, two closely related systems, correlate at .86 (set 1) and .48 (set 2). The MBW seems to be virtually unrelated to the two Swiss systems: the rho's with the WBS and the WBF are .26 and .22 (set 1) and .10 and .26 (set 2), respectively. This outcome might raise some doubts about the feasibility of normative evaluation in general.

* The last two figures are erroneously given a negative sign in Table 5 of Burlisch (1975).
Discussion

The limits of this small study are too obvious to require much elaboration. The number of apartments, the use of simple floor plans to represent them, the size of our interview sample, and the comparatively short duration of the interviews certainly introduced quite a bit of chance into all figures reported here. There is no plausible reason, however, why these shortcomings should have systematically biased our results against the systems.

Returning to our starting point, what should be done to better adapt buildings to their residents' needs? There is no hard and fast answer to this question. If individualism does indeed prevail to the degree found here, then the outlook for single-standard normative systems seems dim. It should be noted that the construction of the present three systems was accomplished by teams of experienced architects and economists after years of extensive work. But if users disagree, it is impossible to agree with all of them.

One possible avenue of research would aim at a typology of evaluators. This could lead to a limited number of descriptive evaluation systems. We still do not know just how people arrive at their judgments. In this study time restraints precluded any systematic probing into our respondents' manner of evaluating, but simple questioning would not have helped much anyway. The "think aloud" method as used by Svenson (1974; see also Montgomery & Svenson, 1976; Payne, 1976) might be worth trying more systematically.

What can be done in the meantime? Reports of participative projects (for a review, see Mihlich et al., 1978, ch. 8) should be analyzed carefully for possible reasons of disappointing results. Planning a housing project is a demanding task and where so much is at stake, competent help may be needed to prevent communication and cooperation breakdown. From an optimal pooling of both laymen's and architects' resources there should emerge residences that match their inhabitants' needs better than standard public housing. Postconstruction follow-up would also seem highly desirable. Psychologists might have something to offer along these lines. It ought to be expert and lay opinion, not expert or lay opinion.
Abstract

Sets of apartment floor plans were evaluated by means of two Swiss and one Austrian system, all based on "expert opinion". The same floor plans were also ranked in order of preference by a total of 76 interviewees, including 14 building professionals. The ranking was repeated after 4 weeks and, by part of the sample, after 1 year. Results showed respondents' rankings to be remarkably stable even over the 1-year interval. Average agreement with even the best of the systems was rather modest, interrater-agreement being already slight. Building professionals—that is, experts in themselves—did not consistently agree more with the systems than "lay" persons did. A particularly telling result was that agreement among the three systems varied from less than perfect to zero.

It is tentatively concluded that expert opinion has very little to offer when standardized in a standard evaluation system. Alternative approaches are briefly discussed.

References


