The Pictorial Housing Survey: A New Method of Measuring Housing Quality

By Leonard V. Zumpano and Edward R. Mansfield

Despite the significant increase in the quantity of housing data now available, problems of data comparability and the paucity of housing information at the local and neighborhood levels continue to hamper housing research efforts and community development planning. The purpose of this paper is to inform readers about a new method of measuring local housing conditions developed for the Texas Department of Community Affairs and, more importantly, to report on the reliability and applicability of the technique for other communities.

LIMITATIONS OF EXISTING SOURCES OF HOUSING DATA

For over 30 years, the decennial Census of Housing had been the only source of housing statistics. Consequently, little information was available with which to assess changes in the nation's housing stock or evaluate the effectiveness of on-going housing programs during the years between census publications. With the introduction of the Annual Housing Survey in 1973, however, this sizeable gap in our statistical knowledge has been substantially bridged. Through such publications as *Current Housing Reports* and *Construction Reports,* housing analysts and researchers now have more information at their disposal than ever before.[1] This is not to say, however, that data problems no longer persist. The lack of statistical continuity and definitional differences among these reports makes time

---

Leonard V. Zumpano received his Ph.D. in Economics from Pennsylvania State University. For the past three years he has been involved in teaching and research at the University of Alabama, concentrating on the economic effects of property taxation and government housing programs.

Edward R. Mansfield has been an assistant professor at the University of Alabama for the past three years in the statistics department. He conducted preliminary research on the *Texas Pictorial Housing Survey* at Southern Methodist University.

series or temporal comparisons not only difficult, but subject to serious misinterpretation.

Although the Bureau of the Census, the source for most of our housing data, presents various statistics for evaluating the condition or quality of the nation's housing, the census no longer defines quality. Past difficulties with attempts to arrive at universally acceptable definitions[2] (the 1950 and 1960 Census of Housing) have led to enumeration of those physical characteristics (i.e. plumbing facilities, number of rooms, types of heating systems, etc.) that can be measured with some degree of accuracy. Tabulations of neighborhood conditions, such as noise levels, adequacy of public services and the like, are included in the Annual Housing Survey, but there is significantly less respondent agreement on these aspects of housing quality (Bureau of the Census, 1976). Effectively then, it is left to the users of these reports to interpret these "broad indicators" as best they can in order to arrive at some understanding of the conditions that currently prevail in the housing market. Although this is not necessarily an insurmountable obstacle, revisions, modifications, and definitional differences among these various reports render much of the data non-comparable.[3] Consequently, conclusions drawn from such data could prove erroneous.

Another problem confronting users of Census Bureau reports is that not enough of the annually published data are sufficiently disaggregated to permit detailed investigation of local and neighborhood housing conditions. In light of the new focus of federal housing assistance programs which now require local officials to identify community housing problems and implement workable solutions, ready access to such local data takes on added significance. Up until quite recently, however, there were no satisfactory ways to ascertain local housing needs which did not entail expensive and time consuming on-site inspections by local housing officials.

Because of these inadequacies, the state of Texas initiated a research effort for the development of an inexpensive, quickly administered, and reliable method of assessing local housing conditions. The result of this research is the *Texas Pictorial Housing Survey,* a technique that is currently being employed successfully by the Dallas Department of Housing and Urban Rehabilitation in the preparation of annual applications for community development block grant funding and for targeting neighborhoods eligible for rehabilitation and low-cost home improvement loans (Schwabe, 1978).

DESCRIPTION OF THE PICTORIAL HOUSING SURVEY

The pictorial housing survey[4] represents an innovative departure from previous approaches to measuring housing conditions and is unique in two important respects. First, the enumerator is not required to make an overall quality judgment of the dwelling unit being surveyed. Rather, the enumerator is only required to observe and then individually rate on a scale of 1 to 7 ten separate characteristics of the housing unit. These individual characteristics are then weighted and summed to produce an

overall composite character rating, W, during the data processing stage.[5] Not only does this procedure minimize the tasks of the enumerator in the field, but it is also presumed to lead to more consistency among observers by obviating the necessity of summary value judgments. The weights are scaled such that the composite W scores range from 1 to 7, which coincide with the rating scale for the individual character components.

The second unique aspect of the housing survey is the criteria by which individual housing components are judged. When rating an individual component of a house, such as a roof, the enumerator does not compare the roof in question to semantic concepts such as "sound" or "deteriorating." Rather the enumerator, equipped with a booklet of photographs, compares the roof under investigation to a series of pictures which depict a range of various roof conditions. Three sets of photographs are arranged in descending order from "2" (best pictorial condition), to "6" (worst pictorial condition). Interpolation between sets of pictures generates the seven-point value scale. In other words, the seven-point scale is referenced by sets of photographs at points 2, 4, and 6 on the scale.

If an observed characteristic of the house being rated looks better than the photographs corresponding to 4, but not as good as the photographs corresponding to 2, then the enumerator would score the component as a 3. Such interpolation is intentional and affords the field worker realistic latitude in rating the physical characteristics of a dwelling unit when the reference photographs do not exactly coincide with an observed characteristic. Sets of pictures, rather than just one photograph, are employed at each point to portray the same physical condition in order to make the pictorial survey inclusive enough to cover situations where housing styles, construction methods, and building materials differ.

The ten housing characteristics included in the pictorial survey are:

1) Neighborhood appearance
2) Appearance of property boundaries (i.e. sidewalks and curbs)
3) Appearance of lawn and shrubs
4) Condition of the roof
5) Condition of interior wall surfaces
6) Condition of porch (if any) and front entryway
7) Condition of doors and door trim
8) Condition of windows and window trim
9) Evidence of electricity
10) Evidence of plumbing

The last two characteristics, electricity and plumbing, are scored only as being present or absent and assigned a value of 1 or 7, respectively.

In order to assess the usefulness of the pictorial housing survey, two questions must be answered. First, is it reliable? Will replication of the survey procedure by different enumerators yield substantially the same results? Secondly, what is actually being measured by the numeric ratings and how should the composite score be interpreted? Preliminary testing of the pictorial housing survey directed at obtaining answers to these questions has been quite promising.

FINDINGS

Reliability of Measurement

One of the major problems the Bureau of the Census encountered using semantic survey techniques to evaluate the quality of housing was the inconsistency in enumerator quality ratings, especially with respect to the identification of substandard dwellings. Follow-up studies after the 1950 and 1960 Census indicated that of all the dwellings classified as dilapidated by post census enumerators, less than 50% (48% in 1950 and only 38% in 1960) had been similarly characterized by the original enumerators (Social and Economic Statistics Administration, Bureau of the Census, 1972).

In order to test the reliability of the pictorial housing survey, the study reported here (Schucany, Mansfield, Woodward, and Hess, 1978) administered the pictorial survey on a randomly-selected sample of dwellings in Dallas, Texas. Because the identification of substandard housing conditions is one of the major concerns of housing officials, the study incorporated a disproportionally large number of low-quality housing units by limiting the sample areas to old, low-income neighborhoods.

In order to include a sufficiently large number of houses and field workers, and still remain within the economic constraints of the study, a balanced, incomplete block design was used. The objective was to estimate the amount of variation in the scores given to a house by the population of field workers. The specific design used 105 houses and 21 field workers. Each house was evaluated by five field workers who each rated a total of 25 houses. This design has the property that each pair of enumerators would rate a common dwelling five times.

A particular evaluation was modeled as:

$$W_{ij} = \mu + \beta_i + \alpha_i + \epsilon_{ij}$$

where W_{ij} is the composite score given to the j_{th} house by the i_{th} field worker, μ represents the mean score for all dwellings, β_j is the effect of the particular house being rated and α_i gives the additive contribution to the score attributable to a particular field worker. The means of the populations of all β_j and all α_i are zero. The last term, ϵ_{ij}, represents the random error or unexplainable effect.

The estimates of the variance components of this random-effects model provide information about the reliability of the pictorial housing survey. The actual estimates of the variances of the three terms in the model are given below.

TABLE 1

ESTIMATED VARIANCE COMPONENTS FROM BIBD

Effect	Estimated Variances
Due to different houses	.4186
Due to different field workers	.0653
Due to random fluctuations	.1933

The relatively large value of the variance due to different houses is indicative of the substantial cross-section of the housing stock captured in the sample and reflects the different housing conditions encountered and identified by the enumerators. In contrast, the small estimated value of the variance due to field workers shows that enumerator eccentricities played a very small part in the determination of the final quality ratings of the dwelling units.

The best assessment of the validity of the pictorial housing scales is the standard deviation of the measurement of the condition of a particular housing unit. This value is a combination of the variation due to field workers, α_i, and the variation due to unexplained sources, ϵ_{ij}. The point value of this estimator is $\sqrt{.0653 + .1933} = .5086$ (a 95% confidence interval is .4610 to .5617). This suggests that one would expect a single measurement of the condition of a given dwelling to be within one unit of the real condition (i.e. $\mu + \beta_j$) of that unit.

Validity of Measurement

The validity of the pictorial housing survey was investigated by comparing the composite W scores with professional housing inspectors' reports on a sample of 566 single family and duplex housing units in Dallas, Texas (Schucany, Mansfield, Woodwood, and Hess, 1978). The housing inspections, filed by experienced housing inspectors of the Dallas Department of Housing and Urban Rehabilitation, characterized each sample unit as either in good condition (no repairs or only minor repairs needed) or in substandard condition (needing major repairs or dilapidated). Nonprofessional personnel, trained by the Dallas Department of Housing and Urban Rehabilitation, administered the pictorial survey.

A contingency table analysis was performed with the W scores categorized into seven groups and cross tabulated with the dwelling unit's condition, as reported by the housing inspectors. *Table 2* illustrates the ranges of the seven W categories as well as the cell frequencies and marginal totals. *Table 2* indicates that as the W score increases, the number of dwelling units rated in good condition by housing inspectors decreases. Only 3% of the dwelling units deemed in poor condition have a W score as low as 2.5. In contrast, almost 70% of the units with W scores of 5 or greater were rated

TABLE 2

COMPARISON OF PICTORIAL SCALE AND HOUSING INSPECTOR RATINGS

	W Score							
Condition of Housing Unit	1.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5	4.5-5.0	5.0-7.0	
Satisfactory	97	118	104	86	33	23	7	467
Substandard	3	10	20	26	12	12	15	98
	100	128	124	112	45	35	22	566

substandard. It can be easily seen that there is indeed a definite relationship between W scores and the condition of dwelling units sampled in Dallas, and that the pictorial survey is able to discriminate among dwellings on the basis of physical condition.

THE USES AND LIMITATIONS OF THE PICTORIAL HOUSING SURVEY

Time, Cost Advantages

Preliminary testing and actual field use in Dallas have shown the pictorial survey to be a reliable and valid method of ascertaining local housing conditions. Because it can be quickly and easily administered by non-professionals such as college students and local residents seeking part-time or temporary employment, the pictorial housing survey offers considerable cost saving advantages over traditional survey methods which require the services of highly-trained housing inspectors. Equally important, skilled manpower thus freed could be assigned more complex and demanding tasks where their skills could be employed more productively.

Program Implementation

The pictorial housing survey can be of invaluable assistance in the design and implementation of locally-initiated housing assistance programs. The numeric scale, which ranks dwelling units by their physical condition, would enable housing officials to not only identify neighborhoods with high concentrations of substandard housing but also determine the relative severity of these conditions. In this way, the pictorial survey can be used to formulate housing policy priorities and allocate funds to those neighborhoods where housing problems appear most severe.

Program Assessment

As part of a comprehensive community development program, the pictorial housing survey can help policymakers monitor annual changes in the condition of their community's housing stock and thereby provide important feedback with which to assess the operational effectiveness of ongoing housing assistance programs. In this regard, the city of Dallas administers the pictorial housing survey in conjunction with a citizen profile, which gathers socio-economic and demographic data on city residents who reside in surveyed dwellings. The citizen profile also samples citizen satisfaction with public services and their assessment of various housing policies. Survey results are then used to modify or re-design public policies found to be ineffective.

For example, when the results of a recent survey indicated that residents were extremely concerned about the overall appearance of their neighborhoods, the city responded by re-directing local and revenue sharing funds to up-grade housing code enforcement and develop a home repair training program.[6] When another survey showed greater citizen demand for street lighting in neighborhoods than on major thoroughfares, city officials de-

cided to reduce thoroughfare lighting, found to be redundant, and increase neighborhood lighting. The city, as a result, was able to reduce operating costs by $100,000. Equally important, the next year's citizen profile indicated a marked increase in citizen satisfaction with neighborhood lighting conditions (Schwabe, 1978). In fact, because of such successes the city now, as a matter of course, incorporates the results of both the pictorial housing survey and the citizen profile directly into the annual budgetary process in order to help determine future expenditure allocations (Schwabe, 1978).

Initial Screening Tool

The pictorial housing survey will not substitute for detailed housing inspections directed at housing code enforcement or the detection of code violations. However, as a preliminary screening device, the pictorial housing survey can serve as an initial cost-saving step in the inspection and code enforcement process. By identifying those housing units most likely to fail city housing standards, it would obviate the need for a more detailed and time-consuming 100% inspection of all the dwelling units within neighborhoods designated for concentrated code enforcement. A suggested procedure would be to administer the pictorial survey on all the housing units within a subject neighborhood. Next, select a threshold W value and inspect only those dwellings units whose pictorial scores exceed this cutoff value. The threshold W score chosen would depend upon the degree of detection desired relative to the amount of time and money available for detailed inspections.

As an example, we can use the sample of 566 dwelling units in Dallas to illustrate the procedure. If a threshold score of 3 were chosen, only 60% of the dwelling units would have been inspected (those units with composite W scores of 3 or greater) but almost 90% of the units rated as substandard by housing inspectors would have been detected. A complete inspection of all units would have done little to improve the degree of detection, but would have involved significantly greater cost.

Multi-Family Unit Use

The applicability of the pictorial survey to multi-family dwellings remains indeterminate. Although the original Dallas housing sample included multi-family units, their number was too small to allow any meaningful statistical analysis. While the pictorial survey may validly discern the condition of apartment units, it is also not unreasonable to conjecture that the exterior appearance of these dwellings would not be as indicative of interior condition as would be the case with single-family homes. Alternatively, it is also possible that exterior physical characteristics, other than those included in the photographic portfolio, may be superior discriminators of multi-family housing conditions. Additional field testing and analysis, however, is needed before definitive answers to these questions can be attained. Certainly the encouraging test results so far achieved, and the successful operational use of the pictorial housing survey in Dallas, justifies continued research in these areas.

REFERENCES

1. For a concise discussion of the various types and sources of housing data currently available, see Phillip E. Kidd, "Housing Statistics: Some Sources and Some Limitations," *American Real Estate and Urban Economics Association Journal* 5 (Fall 1977), pp. 337-354.
2. For a detailed analysis of the problems associated with defining satisfactory semantic measures of housing quality, see Robert Moore Fisher, "Housing Production, Consumption Statistics and Policy," *American Real Estate and Urban Economics Association Journal* 4 (Winter 1976), p. 7-18, and U.S. Department of Commerce, Bureau of the Census, *Measuring the Quality of Housing: An Appraisal of Census Statistics*, Working Paper no. 25 (Washington, D.C.: U.S. Government Printing Office, 1967).
3. Some of the statistics in the *Annual Housing Survey* and the *Decennial Census of Housing* are not comparable because they are based on different or restricted universes. For example, statistics on sewage disposal in the former report are limited to units occupied three months or longer, whereas census data on sewage disposal are shown for all units, including units which were occupied less than three months.
4. The pictorial housing scale was developed by Curtiss C. Grove, Grove and Associates, Inc., Dallas, Texas.
5. The composite character rating, W, is the first principal component which consists of the linear combination of the ten individual rating characteristics, which account for more of the variance in the data than any other linear combination.
6. This information was obtained from a case study of Dallas, Texas, prepared by Charles J. Schwabe. The case study was developed through a contract with the Office of Policy Development and Research, U.S. Department of Housing and Urban Development, and conducted jointly by the International City Management Association and the American Institute of Planners.

BIBLIOGRAPHY

Fisher, Robert M., "Housing Production, Consumption Statistics and Policy," *American Real Estate and Urban Economics Association Journal, vol. 4, no. 3 (Winter 1976), pp. 7-18.*

Kidd, Phillip E., "Housing Statistics: Some Sources and Some Limitations," *American Real Estate and Urban Economics Association Journal,* vol. 5, no. 3 (Fall 1977), pp. 337-354.

Schucany, William R., Mansfield, Edward R., Woodwood, Wayne A., and Hess, James L., "An Analysis of the Reliability of a New Scale For Housing Quality," *Journal of Statistical Planning and Inference* (forthcoming).

Schwabe, Charles J., *"701" Planning and Management. Dallas, Texas: City Profile Survey, A Program Designed to Increase Citizen Participation through Citizen Attitude Surveys.* International City Management Association, Washington, D.C., 1978.

U.S. Department of Commerce, Bureau of the Census, *Annual Housing Reports: 1974.* Washington, D.C.: U.S. Government Printing Office, 1976.

_____. *Measuring the Quality of Housing: An Appraisal of Census Statistics,* Working Paper No. 25. Washington, D.C.: U.S. Government Printing Office, 1967.

_____. Social and Economic Statistics Administration, *Delineation of Problem Housing Areas, Data Access Description Number 28.* Washington, D.C.: U.S. Government Printing Office, May 1972.