

# Searching for the Good Plan

## A Meta-Analysis of Plan Quality Studies

Philip Berke

David Godschalk

*University of North Carolina*

Despite the centrality of comprehensive plans to the profession of city and regional planning, there is a gap in knowledge about the quality of plans as they are not routinely evaluated against best practice standards. We discuss plan quality evaluation, an emerging methodology for assessing the quality of plans. We review the *evolution* of the concept, the *dimensions* covered, and the *principles and criteria* used. We then provide a comparative meta-analytic analysis of the *findings* from published plan quality evaluations, ranging from research studies in the United States to national planning policy applications in Holland and New Zealand, and offer recommendations about *future directions* in improving content analysis of plan documents and research design in plan quality evaluation.

**Keywords:** *land use plan evaluation concepts; plan evaluation methodology; plan quality principles and criteria; meta-analysis of plan quality research; future directions in plan evaluation*

Local comprehensive plans play a pivotal role in guiding and regulating urban development (Kaiser and Godschalk 1995). The growing power of local planning is evidenced by state government experimentation with decentralized approaches to growth management and land use. Since the mid-1990s, twelve states have adopted smart growth legislation that requires or encourages local governments to adopt comprehensive plans to alter development practices dominated by low-density sprawl (Meck 2002). An increasing number of local governments are experimenting on their own with comprehensive land use plans that strive to create balanced alternatives to unbridled growth (Beatley 2005). Countries such as Holland and New Zealand have adopted national legislation mandating that local plans and implementation actions integrate key principles for sustainable land use (Beatley 1999; Ericksen, Berke, and Crawford 2004).

Adopted plans have wide-ranging powers to influence environmental justice, quality of life, economic opportunity, disaster resistance, transportation efficiency, infrastructure costs, and many other important aspects of community life (see, e.g.: Berke, Godschalk, and Kaiser 2006; Bullard 2007; Burby and Dalton 1994; Burby et al. 1999; Knaap, Ding, and Hopkins 2001; Meck 2002). Preparation of these

powerful instruments encourages democratic determination of visions and support of public engagement to foster integration of stakeholder values into goals and policies that guide future growth. Kaiser and Godschalk (1995) observe, "Not only do [land use] plans help decision makers to manage urban growth and change, they also provide a platform for the formation of community consensus about land use issues, now among the most controversial items on local government agendas" (p. 365).

Given widespread usage and significance of comprehensive plans, it is surprising that plans are not routinely evaluated against accepted plan quality standards. Despite the centrality of land use plan making to the profession of city and regional planning and the growing number of evaluation studies by academic researchers, there is a gap in our knowledge about the quality of most plans. This is partly due to the complexity and future-oriented nature of the plans, which bring together factors, issues, and aspirations from a number of sources to focus on long-range outcomes. It may also be due to differing academic views about the purpose of plans, for example, whether they are concrete documents intended to underpin development regulations and public investments or more transitory provisional schemes used as tools for professional deliberation

(e.g., Hopkins 2001). The evaluation gap also is due to the uniqueness of individual plans, which are designed to suit the needs and objectives of particular localities, and to the range of legal foundations for land use planning, from constitutional issues to mandatory and permissive state statutes. Finally, the lack of evaluation may be due to the perception that plans are in large part works of art—designs that defy rational analysis.

If plans are to achieve their full potential, they should reflect the highest quality of thought and practice. Only systematic evaluation enables us to identify their specific strengths and weaknesses, to judge whether their overall quality is good, and to provide a basis for ensuring that they reach a desirable standard.

This article discusses an emerging effort to design and apply a systematic approach for evaluating plan quality. We initially review the evolution and rationale of plan quality evaluation, and the basic concepts and criteria for plan quality. We then provide a meta-analytic comparison of findings from plan quality studies based on sixteen selected studies, followed by a critique of evaluation instrument design, content analysis procedures, and research designs used by these studies. We conclude with recommendations about future directions in applications of plan quality evaluation, including applications to related land policy domains (e.g., equity, environmental quality, transportation, disaster resiliency, and sustainable development). The future should see important methodological advances, including simulation models, GIS scenarios, and statistical evaluation techniques. One possible future development is that some jurisdictions may begin to *require* that planners conduct both internal and external plan quality evaluations as part of the overall plan preparation and review processes.

### Why Assess Plan Quality?

In a performance-oriented society, people expect to be able to judge how well plans achieve their objectives and how well planning processes have been conducted. We can look back at the outcomes of historic plans, such as the famous Daniel Burnham 1909 plan for Chicago that defined the City Beautiful Movement in the United States and proposed the city's distinctive lakefront parks and roadways, the Magnificent Mile, and the Navy Pier (Smith 2006) or the Robert Moses plans that shaped the New York system of freeways, bridges, and parks (Caro 1974; Ballon and Jackson 2007). With the advantage of

decades of hindsight we can form opinions about how effective these plans were in achieving their objectives. We can also evaluate the planning processes that were carried out in these cases, both of which took place well before the participatory era of urban planning.<sup>1</sup>

It is more difficult to evaluate the outcomes of contemporary plans, whose effects will be realized in a future time when conditions have changed and different standards of evaluation may have been formulated. However, we can evaluate the plans themselves according to contemporary standards of good practice. Such evaluations also enable us to judge the quality of plan making, both to review the effectiveness of past processes and to guide future processes. Plan quality evaluation thus functions as a *learning process* that yields important planning lessons and guidelines.

If we do not evaluate our plans and planning processes, we miss a valuable opportunity to learn how to improve them. For example, evaluation of plans prepared under a statewide affordable housing mandate in Illinois showed that while most of the plans met the minimum legal requirements (procedural compliance), their content included widespread skepticism about the need for, and resistance to, the local actions necessary to implement this unfunded mandate (Hoch 2007).<sup>2</sup> The author concludes that it may be difficult to meet the goals of the unpopular mandate without additional incentives to turn the paper goals into practice commitments.

The essence of a profession, such as city and regional planning, is its capacity to set and enforce high standards of practice. Good practitioners learn from reflecting on their experience and on the quality of their work (Schon 1983); their reflection is assisted by professional norms of good practice. Over time, this professional learning shapes criteria for best practices in land use planning as well as other areas of planning. Findings and lessons from best practices in plan quality evaluation, while not extensive, are available in the published literature.<sup>3</sup>

### Conceptual Foundation for Plan Quality Evaluation

Baer (1997) has comprehensively evaluated the plan evaluation literature, trying to answer the question of how you would know a good plan if you saw one. He tackles both modernist and postmodernist issues, reviews published criteria, and proposes a vocabulary for plan evaluation.<sup>4</sup> He notes that the

type of evaluation will depend on the type of plan under consideration, which can include plans as visions, blueprints, land use guides, remedies to existing problems, responses to administrative requirements or mandates, process-oriented activities (e.g., public participation and interjurisdictional coordination), and pragmatic actions aimed at improving legal or statutory procedural frameworks.

Baer (1997) suggests a set of general criteria for plan assessment, to be used during plan making. His framework, drawn from the literature, is organized around the following categories:

- Adequacy of content (political context, administrative authority, role of preparer, background information, client, purpose, funding, etc.)
- “Rational Model” considerations (assessment criteria, problem identification, goals and objectives, tone, coordination with other agencies, regional context, alternatives considered, etc.)
- Procedural validity (groups involved in plan formulation, data and models used, transformation of technical matters into policy, public comment, etc.)
- Adequacy of scope (consideration of relevant issues, efficiency and equity issues, cost–benefit distribution, financial or fiscal implications, legal implications, political feasibility, etc.)
- Guidance for implementation (appropriate provisions, priorities, costs, time span, scheduling and coordination, impact analysis, responsible agency, etc.)
- Approach, data, and methodology (technical bases, wide data spectrum, flexibility in adding data, data and methodology sources cited, etc.)
- Quality of communication (client and public identified, convincing presentation, rationales for decisions given, proposals consistent with objectives, etc.)
- Plan format (size and format conducive to use, date of publication, authors listed, table of contents, graphic, etc.)

Other studies have employed additional criteria for plan evaluation. Kaiser, Godschalk, and Chapin (1995) and Kaiser and Davies (1999) focus on the conceptual dimensions of plans themselves that define their quality, including their goals, policies, and fact bases. Hopkins (2001) suggests including the external validity of plans, addressing their relevance in fitting the needs of local situations. Berke et al. (2006) and Talen (1996) propose methods for evaluating the implementation success of plans.

For land use plans, we propose a refined approach to plan quality evaluation that recognizes their essential nature (Berke, Godschalk, and Kaiser 2006, 69–82). In our view, a “high-quality plan provides a clear

and convincing picture of the future, which strengthens the plan’s influence in the land planning arena” (p. 69). We believe that two conceptual dimensions should be included in plan quality evaluation: (1) *internal plan quality* that includes the content and format of key components of the plan (e.g., issues and vision statement, fact base, goal and policy framework, implementation, monitoring) needed to guide land use in the future and (2) *external plan quality* that accounts for the relevance of the scope and coverage to reflect stakeholder values and the local situation to maximize use and influence of the plan.<sup>5</sup>

### **Identification of Plan Quality Studies and Evaluation Criteria**

A systematic procedure was derived to identify published studies on plan quality. We used EBSCOhost (2006) databases on Academic Search Premier and Government Collections and the peer-reviewed journals indexed in the bibliographic listings of the *Journal of Planning Literature* between 1995 and 2007 to identify prospective studies for our meta-analysis. We then applied four inclusion criteria in our listing to produce the studies used in our analysis. Specifically, we included studies that (1) presented the mean or proportional score of any findings in plan quality, (2) reported the number of plans that were evaluated, and (3) only included studies published in peer-reviewed journals and books and (4) we selected one publication when multiple publications were based on the same plan quality data set. Application of these criteria to our list of prospective studies resulted in a final set of sixteen plan evaluation studies that we use to examine plan quality.

Table 1 indicates that the selected studies cover a range of topics, research designs, intercoder reliability procedures (and scores), settings, and samples. Natural hazard mitigation within plans is the most frequent topic (seven of the sixteen studies), with the remaining studies covering a diversity of topics, including smart growth, sustainable development, watershed protection, housing affordability, landscape ecosystems, coastal resources, and human rights of indigenous people. The plan evaluation method has been applied to groups of plans in domestic and international settings, with twelve studies based solely in the United States, three studies in New Zealand, one in Holland, and one that included New Zealand and U.S. plans (Berke, Dixon, and Ericksen 1997). Nine studies adopted intercoder reliability procedures and

**Table 1**  
**Summary of Plan Quality Studies**

Investigators	Topic	Intercoder Agreement	Design	Setting	Sample
1. Burby and May 1997	Hazards	Y (.90)	CM	USA	90 mandated local plans 90 nonmandated local plans
2. Berke et al. 1997	Hazards	N	CM	NZ USA	16 regional plans 7 FL regional plans
3. Deyle and Smith 1998	Hazards	N	CM	USA	18 FL local plans
4. Berke et al. 1999	Hazards	Y (.70)	CM	NZ	34 local plans 16 regional plans
5. Godschalk et al. 1999	Hazards	Y	CS	USA	44 state plans
6. Berke and Manta-Conroy 2000	Sustainable development	Y (.84)	CM	USA	10 sus dev local plans 20 non-sus dev local plans
7. Berke et al. 2002	Human rights	Y (.80)	CS	NZ	34 local plans
8. Nelson and French 2002	Hazards	N	CS	USA	19 CA local plans
9. Brody 2003a	Hazards	N	CM/ TS <sup>a</sup>	USA	30 FL local plans 29 WA local plans
10. Brody 2003b	Ecosystems	Y (.97)	CS	USA	30 FL local plans
11. Brody et al. 2004	Watersheds	Y (.92)	CS	USA	35 FL local plans
12. Norton 2005	Coastal	N	CS	USA	40 NC local plans
13. Brody et al. 2006	Smart growth	Y (.97)	CS	USA	46 FL local plans
14. Termorshuizen et al. 2007	Ecosystems	N	CS	Holland	38 regional plans
15. Hoch 2007	Housing affordability	N	CS	USA	36 IL local plans
16. Edwards and Haines 2007	Smart growth	Y (.86)	CS	USA	30 WI local plans

Note: Y = yes; N = no; CS = cross sectional; CM = comparative; TS = time series; NZ = New Zealand; sus dev = sustainable development. Values in parentheses are intercoder agreement scores.

a. Time series: 1991 and 1999 plan quality data were collected for the Florida and Washington samples of plans.

all studies used one or more types of research designs (to be discussed).

We then identified the major plan quality characteristics used in the sixteen selected studies. Table 2 shows the plan quality characteristics and examples of specific criteria grouped under each characteristic. As noted, we categorized the characteristics based on the internal and external conceptual dimensions of plan quality (Berke, Godschalk, and Kaiser 2006, 69-82). We identified seven internal characteristics, with characteristics 1 through 6 reflecting the sequence of tasks in making plan elements that comprise a comprehensive plan. The sequence starts with issue identification and visioning (1), followed by direction-setting elements that include goals (2), fact base for policy selection (3), and policies for guiding future settlement patterns (4). Characteristics 1 through 4 provide the foundation for plan implementation actions (5), and monitoring and evaluation (6) that tracks and assesses the effectiveness of the plan in resolving issues and achieving goals. Finally, internal consistency (7) addresses how well the first six plan elements are integrated. Three external characteristics include organization and presentation

(8) to foster comprehension and understandability of the plan, interorganizational coordination (9) to facilitate coordination among other plans (e.g., transportation, open space, housing, hazard mitigation), and compliance to ensure consistency with federal and state mandates (10).

The characteristics and associated criteria can guide an evaluation of existing plans, and the preparation of a new plan or update of an existing one. Practitioners can use them as a checklist of possible considerations during plan preparation and to stimulate plan authors (planners and the public) to devise variations that are pertinent to local contexts and have continuity among diverse viewpoints (Hoch 2002). Researchers can use them to guide content analysis of plans and create quantitative indices of plan quality. The indices can be used for empirical studies that determine the causes of plan quality (e.g., state planning mandates, local commitment and capacity), and effects of plan quality on various outcomes (e.g., traffic congestion, vulnerability to hazards, affordable housing, ecosystem health) while controlling for local contextual conditions (income, population size and growth rates, education).

**Table 2**  
**Characteristics of Plan Quality That Serve as Evaluation Criteria**

---

**Internal characteristics**

*Issue identification and vision:* Description of community needs, assets, trends, and future vision

Assessment of major issues, trends, and impacts of forecasted change

Description of major opportunities for and threats to desirable land use and development

A vision that identifies what the community wants to be

*Goals:* Reflections of public values that express desired future land use and development pattern

Statements of future desired conditions that reflect breadth of community values

*Fact base:* Analysis of current and future conditions and explanation of reasoning

Present and future population and economy

Existing land use and land supply, and future land demands for various uses (e.g., housing, commercial, industrial, public facilities)

Existing capacity and future demand for public infrastructure

State of natural environment resources and constraints

Clear maps and tables that support reasoning, and enhance relevance and comprehensibility

*Policies:* Specification of principles to guide public and private land use decisions to achieve goals

Sufficiently specific (not vague) to be tied to definite actions

Spatial designs that specify future land use, infrastructure, transportation, and open space networks that are sized to accommodate future growth

*Implementation:* Commitments to carry out policy-driven actions

Timelines for actions

Organizations identified that are responsible for actions

Sources of funding are identified to supporting actions

*Monitoring and evaluation:* Provisions for tracking change in community conditions

Goals are based on measurable objectives, e.g., 40 percent of residents within a quarter mile of transit stop

Indicators of objectives to assess progress, e.g., annual percentage of residents within a quarter mile of transit stop

Organizations identified responsible for monitoring

Timetable for updating plan based on monitoring of changing conditions

*Internal consistency:* Issues, vision, goals, policies, and implementation are mutually reinforcing

Goals must be comprehensive to accommodate issues and vision

Policies must be clearly linked back to goals and forward to implementation actions

Monitoring should include indicators to gauge goal achievement and effectiveness of policies

**External characteristics**

*Organization and presentation:* Provisions to enhance understandability for a wide range of readers

Table of contents, glossary of terms, executive summary

Cross referencing of issues, vision, goals, and policies

Clear visuals, e.g., maps, charts, and pictures, and diagrams

Supporting documents, e.g., video, CD, Web page

*Interorganizational coordination:* Integration with other plans or policies of public and private parties

Vertical coordination with plans or policies of federal, state, and regional parties

Horizontal coordination with plans or policies of other local parties within or outside local jurisdiction

*Compliance:* Consistent with the purpose of plan mandates

Required elements are included in plan

Required elements fit together

---

## Meta-analysis of Plan Quality Studies

While there are a few summaries of findings from plan quality evaluations (Baer 1997; Berke, Godschalk, and Kaiser 2006, ch. 3), none have employed the formal tools of meta-analysis. Meta-analysis is a procedure that allows researchers to systematically analyze previous individual analyses to determine the consistency of empirical findings within a given research program (Glass 1977; Wolf 1986). Reviews of research are undertaken in a format that permits standardization of measurement of

findings and statistical analysis (Glass 1977). The accumulated findings from multiple studies are regarded as data points similar to data points collected under a single study.

Meta-analysis offers an alternative to the traditional narrative discussions of research studies, which are subject to several shortcomings: (1) selective inclusion of studies often based on the reviewer's own impressionistic view of the quality of a study, (2) subjective weighting of studies in the interpretation of findings, and (3) misleading interpretation of study findings (Wolf 1986).

The planning field has given limited attention to the methodological problems of research integration. Little headway can be made by pooling the words in the conclusions of a set of studies, but progress can be enhanced through statistical pooling of findings from multiple studies. While meta-analysis is more frequently used in the fields of epidemiology, education, sociology, psychology, and human communications, planning research has expanded to the point where we are witnessing a critical need for this approach. Bartholomew's (2007) recent meta-analysis of eighty regional planning scenario projects is an exemplar. His meta-analysis reveals best practices in formulating scenarios of alternative future land development patterns, and their impacts on reduction of vehicle miles traveled (VMT) and  $\text{NO}_x$  (nitrogen oxide) emissions. While Bartholomew could have gone further by examining the sources of variation (e.g., duration of time horizons, reallocation of transportation investments, alternative densities and mixes of uses) on VMT and emissions, this study demonstrates the important practical and scholarly implications of meta-analysis in the planning field.

### Computation of Values for Plan Quality

It is not possible in our meta-analysis to directly use values from each study because of individual study differences in how the plan quality characteristics are measured (e.g., scales and number of items for each criterion vary) and how plan quality scores are computed. As discussed in an early and influential article on meta-analysis by education psychologist Gene V. Glass (1977), a critical element of the meta-analytic procedure involves transformation of the statistics of interest (e.g., means and standard deviations) into standardized scores that permit analysis of findings across studies. In our study, such a transformation makes scores comparable across plan quality characteristics.<sup>6</sup> In studies that reported standardized proportionate scores we could directly use the findings (e.g., Berke et al. 1996; Brody 2003a, 2003b).<sup>7</sup> In other studies, we transformed scores by plan quality characteristic by first identifying the maximum possible score of characteristics for each study and then dividing the reported score of each characteristic by the total maximum score to determine a proportionate score (e.g., Burby and May 1997; Nelson and French 2002).<sup>8</sup>

Mean scores for plan quality characteristics from our meta-analytic transformations were computed for each internal and external characteristic, with scores

ranging from a low of 0 to a high of 1. The overall means of characteristics across studies were weighted based on the sample size used in each study (see table 2 for sample sizes) to give equal weight to each case (i.e., mean of each study finding), thereby accounting for differences in sample sizes among the study findings. The main methodological limitation is the small number of scores for several of the characteristics. Hence, statistical significance tests for mean values to compare overall scores across plan quality characteristics are not particularly useful. Thus, our interpretation of findings is based on overall patterns rather than on statistical results. These data limitations reveal the exploratory orientation of the comparisons that follow. A full meta-analysis would require a greater number of studies to use inferential statistics to examine differences across plan quality characteristics (e.g., *t*-test, ANOVA, chi square) and regression models to examine the many sources of variation in plan quality (e.g., OLS, binary, ordinal, multinomial, Poisson).

### Meta-analysis Findings

Mean scores on a scale of 0 to 1 for plan quality characteristics and overall weighted means (with standard deviations) are presented in table 3 for internal characteristics and table 4 for external characteristics. Several findings emerge from the analysis.

*Internal characteristics.* The clarity with which plans identify issues received a moderate overall mean score (.48) compared to overall scores for all internal characteristics ranging from .25 to .63. This finding indicates that some plans clearly define issues, but others only vaguely explain the issues or they are not explained at all. If issues are clearly articulated early in the plan document, then subsequent plan elements are more apt to squarely address issues deemed important by the community. In contrast, poorly defined issues are likely to create policy solutions that do not fit community needs and capabilities. Findings from studies in New Zealand shed light on how plan quality for issue identification could be improved. In the context of resource management, regional plans scored higher in issue identification (.61) than local plans (.47) because regional government planners initiated public engagement earlier in the planning process (Berke et al. 1999). Similarly, local governments that engaged in early consultation with tribal authorities of indigenous people (Maori) scored higher in identification of issues (Berke et al.

**Table 3**  
**Summary of Findings for Internal Characteristics of Plan Quality**

Investigators	Issue	Fact Base	Goal	Policy	Internal Consistency	Implement	Monitor and Evaluate
1. Burby and May 1997							
Mandated plans	.26	.34	.13				
Nonmandated plans	.05	.09	.03				
2. Berke et al. 1997							
New Zealand		.13	.68	.11			
Florida		.45	.53	.23			
3. Deyle and Smith 1998							
4. Berke et al. 1999							
Local plans	.47	.06			.66		.39
Regional plans	.61	.12			.62		.21
5. Godschalk et al. 1999		.24	.33	.24		.24	.24
6. Berke and Manta-Conroy 2000							
Sustainable development				.32			
Nonsustainable development				.36			
7. Berke et al. 2002	.44	.15			.60		
8. Nelson and French 2002		.21	.18				
9. Brody 2003							
<i>t</i> <sub>1</sub> = 1991		.09	.10	.05			
<i>t</i> <sub>2</sub> = 1999		.12	.13	.12			
10. Brody et al. 2003a, 2003b		.22	.37	.44		.58	
11. Brody et al. 2004		.25	.36	.42		.30	
12. Norton 2005		.65	.62	.69	.62	.67	.58
13. Brody et al. 2006				.12			
14. Termorshuizen et al. 2007		.30	.16	.87			
15. Hoch 2007		.47		.14			
16. Edwards and Haines 2007			.67	.22			
Overall means <sup>a</sup>	.48	.23	.31	.25	.63	.44	.38
Standard deviation	.06	.16	.19	.24	.02	.19	.15

a. Overall means are weighted based on sample size of each study.

2002). This helped build trust and a stronger sense of ownership of the planning process among Maori, leading to greater willingness to participate and define issues that were important to them.

Goals, facts, and policies scored the lowest among all characteristics under the internal dimension (facts = .23, goals = .31, and policy = .25). This finding is troublesome since these characteristics serve the critical direction-setting framework of plans. Goals identify what the community wants to become. Policies rely on a range of tools, including regulations (e.g., zoning and subdivision codes), incentives (e.g., development density bonuses, tax abatements), and public infrastructure investments (e.g., transportation, water, and sewer) to achieve the desired future spatial form. And facts provide the bases to ensure decision making in setting goals and policies are well-informed. A weak direction-setting framework means that a community is less likely to exert control over its planning agenda and ensure that long-range public interests supersede short-range interests and private

concerns. This also means that plans will not provide a clear, relevant basis for the implementation and monitoring and evaluation elements.

In contrast to this overall pattern of findings, Norton's (2005) study of forty local coastal plans in North Carolina found a strong direction-setting framework (facts = .65, goals = .62, and policies = .69). Norton contends that these coastal communities have been motivated to produce high-quality plans to contend with the dual pressures of guiding intense economic growth while protecting fragile natural ecological ecosystems. Also, North Carolina coastal plans are mandated by the state. Compared with non-mandated plans, mandated plans are consistently higher quality because mandates encourage local governments to develop better information bases, be more inclusive and incorporate a broader array of goals, and have stronger policies that are more likely to influence development outcomes (Burby and May 1997). Similarly, Termorshuizen, Opdam, and van den Brink's (2007) study of thirty-eight landscape

**Table 4**  
**Summary of Findings for External Characteristics of Plan Quality**

Investigators	Organization and Presentation	Coordination	Compliance
1. Burby and May 1997			
2. Berke et al. 1997			
3. Deyle and Smith 1998			.44
4. Berke et al. 1999			
Local plans	.48	.43	.35
Regional plans	.37	.46	.65
5. Godschalk et al. 1999		.23	
6. Berke and Manta-Conroy 2000			
7. Berke et al. 2002			.70
8. Nelson and French 2002			
9. Brody 2003a			
10. Brody 2003b		.51	
11. Brody et al. 2004		.43	
12. Norton 2005			
13. Brody et al. 2006			
14. Termorshuizen et al. 2007		.74	
15. Hoch 2007			.91
16. Edwards and Haines 2007			
Overall means <sup>a</sup>	.44	.46	.63
Standard deviation	.05	.17	.21

a. Overall means are weighted based on the sample size of each study.

ecological plans in Holland scored highest (.87) for policies among all studies. The Dutch mandate regional landscape-scale plans and pay careful attention to landscape ecological sustainability in response to the country's limited land resource base and pressure to accommodate growth and outward urban expansion.

The overall mean scores are moderate for implementation (.44) and somewhat lower for monitoring and evaluation (.38). These findings indicate that some plans clearly specify organization responsibility and timelines for actions for implementation and monitoring, as well as monitoring indicators to gauge progress in goal achievement. However, other plans only vaguely specify these characteristics of plans or they are not specified at all. If plans score high on all elements except implementation and monitoring and evaluation, then plans may become paper documents that are not carried out and regularly evaluated and improved. In contrast to the overall moderate scores, Brody (2003b) found high implementation scores (.58) for thirty Florida plans in twenty-three watersheds south of Lake Okeechobee, and Norton (2005) found high scores for implementation (.67) and monitoring (.58) in his study of North Carolina coastal plans. Both studies concluded that the pressures of high population growth in sensitive ecological settings created public pressure to ensure that plans are

not rendered meaningless and that they include strong provisions for implementation (and monitoring in North Carolina).

Finally, internal consistency received the highest overall mean score (.63) among all internal characteristics. Strong internal consistency will enhance prospects that issues deemed important enough by the community to be included in plans will be the focus of comprehensive attention throughout the plans and thus to influence actions resulting from them. Findings suggest that planners and their communities are reliably linking issues to goals, goals to policies, policies to implementation, and implementation to monitoring and evaluation of outcomes. Furthermore, changes in outcomes are linked to timetables in plans that require regular plan updates to account for results of monitoring and evaluation. However, no matter how strong the internal consistency the other elements must be strong for a plan to be effective.

*External characteristics.* The compliance characteristic with plan mandates received the highest overall mean score (.63) among the three external characteristics (see table 4). While this finding reveals that in most cases planners and their communities are writing plans that comply with the "letter of the law," the remaining plan quality characteristics must be sufficiently strong to at least achieve or go beyond the

minimum thresholds of planning mandates. As noted, Hoch (2007) found that while plans may meet the minimum legal requirements, their content revealed skepticism and resistance to local actions necessary to implement a mandate (Hoch 2007).

Plans received a moderate overall mean score for interorganizational coordination (.46). Successful implementation and ultimate influence of plans require that local comprehensive plans be coordinated with actions of other organizations. For example, large open spaces should not be purchased by a local government in areas served by local water and sewer utilities, or designated for new or expanded roadways by a state transportation department. Notably, the Termorshuizen, Opdam, and van den Brink's (2007) study of Dutch plans scored high (.74) in interorganizational coordination. This was due to Dutch law that requires strong vertical consistency for coordination among national, regional, and local government plans and horizontal consistency requirements among local governments at the regional landscape scale.

Finally, the overall organization and presentation score of plans was moderate (.44). This finding indicates that some plans organize and present information and offer policy solutions in ways that make them useful and relevant in promoting awareness and support of the public interest of the community at large. Others are poorly organized and presented in ways that cannot be clearly understood by decision makers. Because these scores were derived from only one study of regional and local plans in New Zealand (Berke et al. 1999), we are not able to derive insights about how this characteristic can be improved (or degraded) based on comparisons with other studies.

In sum, meta-analysis of internal and external characteristics of a small sample of studies on plan quality revealed several preliminary findings. First, with the exception of a strong internal consistency score, internal characteristics of plans had mixed results. Issue identification, implementation, and monitoring and evaluation were moderately strong, but direction-setting frameworks (facts, goals, and policies) were weak. Second, with the exception of strong compliance scores, external characteristics of coordination and organization and presentation were moderate in strength. Third, despite these mixed results, plan authors have been able to achieve strong internal consistency of plans. Finally, while plans are strong in compliance with external mandates, the mixed results for other plan quality characteristics may indicate limited commitment and capacity to prepare plans that can be effectively translated to action.

## Evaluation Instrument Design

Analyses of plan quality are obviously influenced by the design of the evaluation instruments that are used to conduct the analyses. Because no standard instrument format has been universally accepted, the instruments vary according to the purpose for which they are intended. Some, such as the Florida Department of Community Affairs checklist for preparing a comprehensive plan update, focus on assessing the compliance of plans with state laws or guidelines. Others, such as the Canada Mortgage and Housing Corporation report card, assess the degree to which smart growth policies are incorporated into the plans, while yet others, such as the Community Land use Evaluation for Natural hazards (CLUE) published by the Institute of Business and Home Safety, assess the robustness of hazard mitigation policies and actions contained in plans.<sup>9</sup>

The instruments also vary according to whether they are intended for use in research or practice. Practice assessments tend to be conducted by practitioners, either employed by a state agency overseeing local plans or by the local planners themselves, and tend to use checklist formats that can be used to determine the presence or absence of a particular item. Research assessments are conducted by academic researchers and tend to follow more demanding research protocols, based on theoretical frameworks and social science standards; they may assess both the presence of an item and its degree on a qualitative scale. Both types are useful in improving overall plan quality.

The studies included in our meta-analysis are all research studies. However, even within this category there are significant variations in instrument purpose and design. Evaluation scores are substantially influenced by the specific set of items used to construct a given measure, such as the fact base, both in terms of the content of the items and the number of items used. Researchers may tend to construct instruments that represent the ideal plan content for the purpose of the evaluation. In the case of a hazard mitigation plan, for example, the items may include every possible mitigation policy or technique that might reasonably be employed; in this case it would be either impossible or highly unlikely that any plan would have a perfect score. In fact, the best scores would be more likely to fall in the 60 percent range, as do the scores in our meta-analysis. The number of items included also can affect the score, when the scores are standardized by dividing items observed by the number of possible items, which is typically done. Here again, when

there are a high number of items, the scores will be likely to fall more toward the mid-range.

These instrument design issues point up the need for caution in making judgments about the validity of plan quality studies. This is especially true for studies such as the ones reported in this article that compare findings across various plan quality analyses that use different types and numbers of evaluation items. While there has been significant convergence in terms of plan quality criteria usage, differences in applications and conceptual framing will exist. As more studies adopt uniform criteria, this will become less of an issue. In the meantime, we should all proceed cautiously.

### **Intercoder Reliability and Content Analysis of Plans**

The application and reporting of intercoder reliability procedures is a critical component of the content analysis of plans. Content analysis is a research technique for the objective description of the content of information contained in a written document like a comprehensive plan, oral messages like radio and television broadcasts, and tape recordings of interviews. When intercoder reliability is not assessed and reported, then doubt is cast on the validity of the data generated by the content analysis (Krippendorff 2004).

Recent debates in human communications research revealed the misconceptions that circulate in the content analysis literature regarding how reliability measures are used and can deceive content analysts (and public policy decision-makers) in the computation and reporting of the reliability of coded data (Krippendorff 2004; Lombard, Snyder-Duch, and Bracken 2002). In response to the concerns about the quality and rigor of content analysis, Krippendorff (2004) proposes four conditions to maximize the intercoder reliability: (1) use multiple coders who work independently, (2) select an appropriate intercoder agreement technique, (3) choose an acceptable level of agreement, and (4) test and report individual reliability scores. We apply these conditions to our sixteen selected studies of plan quality.

#### **Use Multiple Coders Who Work Independently**

To ensure that reliable data are generated, content analyses should be conducted by two or more coders who explicitly follow common coding instructions

and who work independently.<sup>10</sup> Coders must be trained to ensure consistency in the application of instructions.<sup>11</sup> Table 1 indicates that only 56 percent ( $n = 9$ ) of the sixteen studies reported use of multiple coders. While all this subset of studies mentioned coding instructions in terms of types of measurement scales used for each variable, the discussion about training is almost always too brief and underdeveloped. For example, only Godschalk et al. (1999) mentioned use of pretesting procedures to improve reliability of the instructions. By comparison, a study by Lombard, Snyder-Duch, and Bracken (2002) of two hundred published articles between 1994 and 1998 that used content analysis in the human communications research field found that 69 percent ( $n = 137$ ) used multiple coders, and the majority of this subset discussed both coding instructions and training procedures.

#### **Select an Appropriate Intercoder Agreement Technique**

Intercoder reliability should measure the level of agreement for each criterion that falls under each characteristic of a plan. It should indicate the likelihood that conclusions drawn from the data generated by the content analysis are valid beyond chance. Of the subset of nine plan evaluation studies that used multiple coders, 89 percent ( $n = 8$ ) reported an intercoder reliability score (see table 1). All these studies rely on the percentage agreement technique among coders that measures the number of disagreements in coding divided by the number of agreements plus disagreements.

There are several misconceptions about the accuracy of the percentage agreement technique in reporting reliability scores. Most importantly, it does not account for the concern that percentage agreement becomes more difficult to achieve as the number of coding categories increases (e.g., binary versus ordinal scales). To adjust for this effect, Krippendorff (2004) recommends that scores that are computed based on the percentage agreement technique should be divided by the expected chance of disagreement (e.g., 50 percent for binary scale and 66 percent for a 3-point ordinal scale). Krippendorff indicates that in addition to correcting for number of categories per scale, corrections are needed for the effects of the number of pairs of coders and the size of the sample used by multiple coders, which is drawn from the sample used in the study (not the population of ultimate research interest). While we do not have sufficient space here to discuss the limitations and

advantages of the range of techniques for computing agreement scores, both Lombard, Snyder-Duch, and Bracken (2002) and Krippendorff (2004) provide straightforward explanations and penetrating reviews of these techniques.

### **Choose an Acceptable Level of Agreement**

An acceptable level of agreement below which coded data should be rejected due to an intolerable unreliability must be selected. Setting the cutoff depends on the costs of drawing invalid conclusions. For example, the acceptable level should be very high if a decision to attack another country depends on results of content analysis of intercepted messages from a terrorist cell. The level should not be as high when content analysis is intended to gauge and improve the quality of a land use plan. Table 1 indicates that agreement coefficients among the eight studies that report agreement coefficients range from .70 to .97. Suggested ranges for most studies require coefficients of at least .80, and where tentative conclusions need only basic similarity in agreement the lowest level is suggested to be .67 (Krippendorff 2004).

### **Test and Report Individual Intercoder Reliability Scores**

Krippendorff (2004) recommends that intercoder reliability scores should be computed for each individual plan quality criterion. When reporting on an index score composed of multiple criteria for each plan quality characteristic, the reliability of each criterion should be measured separately and the smallest reliability among them should be taken as the reliability of the entire index. Averaging the agreement measures of all the criteria in each characteristic of a plan should be discouraged since the aggregate average can be misleading about the reliability of the coded data.

Counter to the guidance suggested by this condition, reliability scores shown on table 1 represent the overall average of all criteria for all plan quality characteristics that were examined in each of the eight studies that report scores. These scores are inflated and may give plan evaluators and users of plan evaluation research an unfounded sense of confidence about the reliability of the data.

## **A Critique of Research Designs**

Three types of research designs have been employed.<sup>12</sup> Cross-sectional samples of plans were

used by 62.5 percent of the studies (ten of sixteen) listed on table 1. These studies yield results of scores for various plan quality indexes for a national sample of state mitigation plans in the United States, national samples of regional plans in Holland and regional and district plans in New Zealand, and state-level samples of local plans in Florida, California, Illinois, North Carolina, Washington, and Wisconsin. Within this subset of studies, several entail a descriptive assessment of strengths and weaknesses of different characteristics of plans (e.g., Brody, Highfield, and Carrasco 2004; Edwards and Haines 2007; Hoch 2007; Termorshuizen, Opdam, and van den Brink 2007). Other studies are aimed at an evaluation of plan quality and various local conditions that affect plan making and plans, including, for example, state planning mandates for local plans, public participation schemes, threats to biodiversity, and natural disaster history (e.g., Berke et al. 2002; Burby and May 1997; Brody, Highfield, and Carrasco 2004; Deyle and Smith 1998; Nelson and French 2002). A criticism of the cross-sectional approach is that researchers have no control of selecting key factors (i.e., independent variables) that may affect plan quality scores.

Some studies have introduced a comparative research design where investigators selected a group of plans where a key factor is present and then selected another group of plans where the factor is not present or is altered in some way. Burby and May (1997) employed such an approach in a comparative study of ninety local plans in three states with local planning mandates and ninety local plans in three states without mandates. Berke and Manta-Conroy (2000) compared a sample of plans that integrated the concept of sustainable development with a sample that did not integrate this concept. Finally, Deyle and Smith (1998) used regulatory implementation theory to guide a comparison of compliance rates of local plans with Florida's state mandate under two different state planning office administrations.

The main shortcoming of the comparative approach is that only differences between the groups subject to an intervention (e.g., state-mandated local plans) compared to the control group (e.g., nonmandated local plans) are taken as evidence of the impact. However, there may be considerable differences between each group at the start of the intervention, and the change resulting from the intervention may not be accurately detected given differences at the outset. A superior research design was employed by Brody (2003a) that combines the comparative and time-series approaches. Here a more accurate distinction can be estimated by

considering both pre- and postintervention information in both targeted and control groups of plans. Brody compared differences in the absolute change between one group of plans before and after a state mandate intervention with the absolute change in a control group during the same time period.

## Conclusions and Recommendations

Plan quality evaluation is emerging as a valuable tool for systematic analysis of the goodness of plans. The methodology offers an objective and straightforward tool for studying plan quality and guiding plan preparation. It has been applied by university researchers in a variety of settings in the United States and abroad to a number of plan components, including those dealing with natural hazards, sustainable development, human rights, ecosystems, watersheds, coastal area management, and housing affordability.

A meta-analysis of sixteen published plan quality evaluations reveals patterns of plan strength and weakness. While these qualities tended to vary with the plan element topic and setting, we found some clear patterns. In terms of *internal* plan dimensions, consistency scored the highest, while fact base scored lowest. It is heartening to learn that plan authors are preparing consistent documents, given the complexity of linking issues, goals, policies, implementation actions, and monitoring indicators within the plans. However, it is disappointing to learn that the planners are not providing stronger fact bases for some since these data undergird the proposals of the plans. Evaluation of plan goals and policies also showed weaknesses. *External* dimension of plan quality tended to fare better. Compliance characteristic received the highest overall mean score among the three external characteristics. Coordination and organization and presentation received moderate overall scores.

Where is plan quality evaluation likely to head in the future? We expect to see a broadening scope of applications related to land policy, including equity, environmental quality, transportation, sustainability, smart growth, and others. The capacity of plan quality evaluation to highlight plan strengths and weaknesses should prove especially valuable in the analysis of the impacts of controversial or innovative land use policies in a variety of fields.

We also expect to see important methodological advances in the future. Improvements in intercoder reliability will support the generation of more valid and reliable plan quality data. More sophisticated

research designs will allow researchers to better detect the influence of various plan-making strategies on plan quality as well as gauge the influence of plan quality on the sustainability of outcomes. Advances will be made as the plan quality approach is refined and linked to simulation models, GIS scenarios, and various goal-oriented techniques. The quantitative evaluation of plan quality relative to proposed visions, goals, and policies could open a new field of land use research.

It is also possible that some jurisdictions may begin to *require* that planners conduct both internal and external plan quality evaluations as part of the overall plan preparation and review processes. For example, states with smart growth requirements may come to see that the standard checklist approach to plan compliance with state guidelines is inadequate to address the full intent of smart growth standards. These states may ask not just about the presence of a policy in a plan but also whether the policy is well integrated with the goals, facts, and issues expressed in the plan.

The bottom line is that we believe that plan quality evaluation is likely to be increasingly prominent in both research and practice. The approach can provide invaluable insights into plan performance and can open new vistas for understanding how to prepare better plans. Such a powerful tool deserves widespread usage.

## Notes

1. Both plans would be classified as elitist by today's standards. The client for the Chicago Plan was the Commercial Club of Chicago, although generations of schoolchildren in the city studied the plan in a manual prepared in 1911 by Walter D. Moody. Moses used the power he accumulated through setting up independent commissions to push through his freeways and large construction projects, although the beneficiaries were often lower-income families, who gained access to new housing and recreation facilities.

2. Hoch (2007) read the thirty-six written plans, evaluating them for compliance, consistency, relevance, and commitment. He also surveyed local officials concerning their attitudes toward the mandate.

3. See, e.g., the criteria listed in Berke, Godschalk, and Kaiser (2006, 69-74) and incorporated into Esnard et al. (2006); the Florida Department of Community Affairs checklist for preparing a comprehensive plan update (<http://www.dca.state.fl.us/fdcp/DCP/EAR/1Introduction.pdf>); Tomalty and Alexander (2005); and the Community Land Use Evaluation for Natural Hazards (CLUE) criteria established by the Institute for Business and Home Safety (<http://www.ibhs.org/publications/downloads/543.pdf>).

4. Baer (1997) notes the difficulty of formulating postmodern criteria, given their objections to rationality, clarity, consistency,

and instrumentality. He comes down on the side of an informed and cautious modernist approach that does not become too enamored with techniques and expertise. To clarify the unsystematic vocabulary of the field, he recommends that plan “critique” be used to refer to an outside review after the plan is published, that plan “assessment” be used to label the testing of alternatives during the plan-making process, that “comparative research” be used to describe the postadoption (but preevaluation) analysis of plans, and that “post-hoc plan outcomes” be the title of work that empirically evaluates plan implementation over time.

5. An illustrative application of the *internal* and *external* dimensions of the plan quality evaluation framework was conducted for the 2000 *Denver Comprehensive Plan* and the 2002 *Blueprint Denver: Land Use and Transportation Plan* (see Godschalk 2004 and Berke, Godschalk, and Kaiser 2006, 74-76).

6. Of course, there would be improved reliability and validity of data in using standardized measurements (see Kalkhoff and Thyne 2006 as an exemplar in use of standardized measurement to conduct meta-analysis in the area of experimental social psychology). However, as discussed in the section Evaluation of Instrument Design, there is no acceptable standard of measurement of plan quality. This endeavor is evolving and needs improvement in conceptual framing and measurement. However, Glass’s (1977) observation of the state of education psychological research over three decades ago is consistent with the issues confronting contemporary plan quality scholarship—that is, meta-analysis should not be precluded based on the argument that an accepted standard of measurement of plan quality has yet to be determined and standardized.

7. Proportionate scores were calculated but the scales varied across studies; for example, Brody (2003a, 2003b) used a 0-to-10 scale, and Berke et al. (1996) used 0-to-100 scale. The scores from these studies were standardized to a 0-to-1 scale. In studies that reported the percentage of plans that included a particular item—e.g., Termorshuizen, Opdam, and van den Brink (2007) reported the percentage of plans that included clear delineation of natural areas—the percentage is simply converted as a proportionate score on a 0-to-1 scale.

8. Several studies included multiple items for a particular plan quality characteristic. For example, Hoch (2007) included two items for the compliance characteristic: “compliant with the mandate requirements” and “consistent internally with the mandate.” We combined the individual mean proportionate scores for each of these items and then computed an overall mean proportionate score for compliance for this study (see table 4 under Hoch 2007).

9. See note 3 for citations of plan quality evaluation instruments from practice.

10. By reliability data, we mean the sample of data drawn from the study sample that should be representative of data (i.e., plan documents) whose reliability is in question (not of the population of ultimate research interests).

11. A single coder is more likely to drift from the rules established for coders since there is no procedure to cross-check his or her accuracy.

12. While contextual quasi-controls (e.g., population size and growth rates, and income levels of local communities) are potentially important factors that affect plan quality, we only focus on the strengths and weaknesses of the three research designs shown on table 2 in determining plan quality scores. A discussion of the implications of combining a quasi-experimental design with each of the three designs would be too long and not consistent with our main goal aimed at findings on plan quality.

## References

- Baer, W.C. 1997. General plan evaluation criteria: An approach to making better plans. *Journal of the American Planning Association* 63 (3): 329-45.
- Ballon, H., and K. T. Jackson, eds. 2007. *Robert Moses and the modern city: The transformation of New York*. New York: Norton.
- Bartholomew, K. 2007. Land use transportation scenario planning in an era of global climate change. *Transportation* 34 (4): 397-412.
- Beatley, T. 1999. *Green urbanism: Learning from European cities*. Washington, DC: Island Press.
- . 2005. *Native to nowhere: Sustaining home and community in a global age*. Washington, DC: Island Press.
- Berke, P. R., M. Backhurst, L. Laurian, J. Crawford, and J. Dixon. 2006. What makes plan implementation successful? An evaluation of local plans and implementation practices in New Zealand. *Environment and Planning B: Planning and Design* 33 (4): 581-600.
- Berke, P. R., J. Crawford, J. Dixon, and N. Ericksen. 1999. Do cooperative environmental planning mandates produce good plans? Empirical results from the New Zealand experience. *Environment and Planning B: Planning and Design* 26: 643-64.
- Berke, P. R., J. Dixon, and N. Ericksen. 1997. Coercive and cooperative intergovernmental mandates: A comparative analysis of Florida and New Zealand environmental plans. *Environment and Planning B: Planning and Design* 24: 451-68.
- Berke, P. R., N. Ericksen, J. Crawford, and J. Dixon. 2002. Planning for indigenous people: Human rights and environmental protection in New Zealand. *Journal of Planning Education and Research* 22: 115-34.
- Berke, P. R., D. R. Godschalk, and E. J. Kaiser, with D. A. Rodriguez. 2006. *Urban land use planning*. 5th ed. Urbana: University of Illinois Press.
- Berke, P. R., and M. Manta-Conroy. 2000. Are we planning for sustainable development? An evaluation of 30 comprehensive plans. *Journal of the American Planning Association* 66 (1): 21-33.
- Berke, P. R., D. Roenigk, E. Kaiser, and R. Burby. 1996. Enhancing plan quality: Evaluating the role of state planning mandates for natural hazard mitigation. *Journal of Environmental Planning and Management* 39 (1): 79-96.
- Brody, S. 2003a. Are we learning to make better plans? A longitudinal analysis of plan quality associated with natural hazards. *Journal of Planning Education and Research* 23 (2): 191-201.
- . 2003b. Implementing the principles of ecosystem management through local land use planning. *Population and Environment* 29 (6): 511-40.
- Brody, S., V. Carrasco, and W. Highfield. 2006. Measuring the adoption of local sprawl reduction planning policies in Florida. *Journal of Planning Education and Research* 25: 294-310.
- Brody, S., W. Highfield, and V. Carrasco. 2004. Measuring the collective planning capabilities of local jurisdictions to manage ecological systems in southern Florida. *Landscape and Urban Planning* 69: 32-50.
- Bullard, R., ed. 2007. *Growing smarter: Achieving livable communities, environmental justice, and regional equity*. Cambridge, MA: MIT Press.

- Burby, R. J., P. R. Berke, R. E. Doyle, S. F. French, D. R. Godschalk, E. J. Kaiser, J. D. Kartez, P. J. May, R. Olshansky, R. G. Paterson, and R. H. Platt. 1999. Unleashing the power of planning to create disaster-resistant communities. *Journal of the American Planning Association* 65 (3): 247-58.
- Burby, R. J., and L. C. Dalton. 1994. Plans can matter! The role of land use plans and state planning mandates in limiting the development of hazardous areas. *Public Administration Review* 54 (3): 230-38.
- Burby, R. J., and P. J. May, with Philip R. Berke, Linda C. Dalton, Steven P. French, and Edward J. Kaiser. 1997. *Making Governments Plan: State Experiments in Managing Land Use*. Baltimore, MD: Johns Hopkins University Press.
- Caro, R. A. 1974. *The power broker: Robert Moses and the fall of New York*. New York: Knopf.
- Deyle, R. E., and R. A. Smith. 1998. Local government compliance with state plan mandates: The effects of state implementation in Florida. *Journal of the American Planning Association* 64 (4): 457-69.
- Edwards, M. M., and A. Haines. 2007. Evaluating smart growth: Implications for small communities. *Journal of Planning Education and Research* 27 (1): 49-64.
- Erickson, N., P. Berke, and J. Crawford. 2004. *Plan-making for sustainability: The New Zealand experience*. London: Ashgate.
- Esnard, A., P. Berke, D. Godschalk, and E. Kaiser. 2006. *Hypothetical city workbook III: Exercises in GIS data to accompany urban land use planning*. 5th ed. Champaign: University of Illinois Press.
- Glass, G. 1977. Integrating findings: The meta-analysis of research. *Review of Research in Education* 5: 351-79.
- Godschalk, D. R. 2004. Land use planning challenges: Coping with conflicts in visions of sustainable development and livable communities. *Journal of the American Planning Association* 70 (1): 1-9.
- Godschalk, D. R., T. Beatley, P. Berke, D. Brower, and E. Kaiser. 1999. *Natural hazard mitigation: Recasting disaster policy and planning*, 327-91. Washington, DC: Island Press.
- Hoch, C. 2002. Evaluating plans pragmatically. *Planning Theory* 1 (1): 53-75.
- . 2007. How plan mandates work: Affordable housing in Illinois. *Journal of the American planning Association* 73 (1): 86-99.
- Hopkins, L. D. 2001. *Urban development: The logic of making plans*. Washington, DC: Island Press.
- Kaiser, E. J., and J. Davies. 1999. What a good plan should contain: A proposed model. *Carolina Planning* 24 (2): 29-41.
- Kaiser, E. J., and D. R. Godschalk. 1995. Twentieth century land use planning: A stalwart family tree. *Journal of the American Planning Association* 61 (3): 365-85.
- Kaiser, E. J., D. R. Godschalk, and S. Chapin Jr. 1995. *Urban land use planning*. 4th ed. Urbana: University of Illinois Press.
- Kalkhoff, W., and S. Thye. 2006. Expectation states theory and research: New observations from meta-analysis. *Sociological Methods and Research* 35 (2): 219-49.
- Knaap, G. J., C. Ding, and L. D. Hopkins. 2001. Do plans matter? The effects of light rail plans on land values in station areas. *Journal of Planning Education and Research* 21: 32-39.
- Krippendorff, K. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human Communications Research* 30 (3): 411-33.
- Lombard, M., J. Snyder-Duch, and C. Bracken. 2002. Content analysis in mass communication research: An assessment and reporting of intercoder reliability. *Human Communication Research* 28: 587-604.
- Meck, S., ed. 2002. *Growing smart legislative guidebook: Model statutes for planning and management of change*. 2 vols. Chicago: American Planning Association.
- Nelson, A., and S. French. 2002. Plan quality and mitigating damage from natural disasters: A case study of the Northridge Earthquake with planning considerations. *Journal of the American Planning Association* 68 (2): 194-207.
- Norton, R. 2005. More and better local planning: State-mandated local planning in coastal North Carolina. *Journal of the American Planning Association* 71 (1): 55-72.
- Schon, D. A. 1983. *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Smith, C. 2006. *The Plan of Chicago: Daniel Burnham and the remaking of the American city*. Chicago: University of Chicago Press.
- Talen, E. 1996. After the plans: Methods to evaluate the implementation success of plans. *Journal of Planning Education and Research* 16 (2): 79-91.
- Termorshuizen, J., P. Opdam, and A. van den Brink. 2007. Incorporating ecological sustainability into landscape planning. *Landscape and Urban Planning* 79: 374-84.
- Tomalty, R., and D. Alexander. 2005. *Smart growth in Canada: A report card*. Ottawa: Canada Mortgage and Housing Corporation.
- Wolf, F. 1986. *Meta-analysis: Quantitative methods for research synthesis*. London: Sage.

**Philip Berke and David Godschalk** are colleagues in the Department of City and Regional Planning at the University of North Carolina at Chapel Hill. They are coauthors of work on plan quality evaluation included in two books, *Urban Land Use Planning* (University of Illinois Press, 5th ed., 2006) and *Natural Hazard Mitigation* (Island Press, 1999), and have individually published journal articles and reports on the topic.