Synthetic populations of building office occupants and behaviors

Citation to Publisher Version: Senick, Jennifer, Andrews, Clinton J., Chandra Putra, Handi, Tsoulou, Ioanna & Sorensen Allacci, MaryAnn. (2018). Synthetic populations of building office occupants and behaviors. In W. Preiser, A. Hardy, U. Schramm (Eds.), *Building Performance Evaluation* (63-72). Springer. https://doi.org/10.1007/978-3-319-56862-1_5.

Citation to *this* Version: Senick, Jennifer, Andrews, Clinton J., Chandra Putra, Handi, Tsoulou, Ioanna & Sorensen Allacci, MaryAnn. (2018). Synthetic populations of building office occupants and behaviors. In W. Preiser, A. Hardy, U. Schramm (Eds.), *Building Performance Evaluation* (63-72). Springer. Retrieved from http://dx.doi.org/doi:10.7282/t3-a1zs-x315.

Chapter 5
Synthetic Populations of Building Office Occupants and Behaviors

Jennifer A. Senick, Clinton J. Andrews, Handi Chandra Putra, Ioanna Tsoulou and MaryAnn Sorensen Allacci

5.1 Introduction

The goal of this chapter is to convey a novel approach to overcoming the limitations of case study research of building occupant behavior in workplace settings by pooling samples and creating a synthetic population of building occupants and behaviors. Synthetic populations can be used by researchers and designers of buildings to develop more accurate models of performance and behavior (Andrews et al. 2016). In the example presented here, three disparate field studies of workplace settings are combined into a larger database that is enhanced through the generation of a statistically similar synthetic data set.

5.2 Building Occupant Behavior and Synthetic Databases

Behavioral researchers know that occupants influence energy use and other aspects of building performance by their heterogeneous choices over building environmental conditions and adaptive behaviors (Bordass et al. 2001; Hewitt et al. 2015). Computer simulation models are increasingly used to assist in the building design process, but few incorporate the influence of occupant behavior (Andrews et al. 2016; Hong et al. 2015). Similarly, designers struggle to optimize comfort, satisfaction, health, and productivity of commercial office building occupants (Leaman and Bordass 1999; Heerwagen 2000), without knowing attendant parameters a priori, e.g., before the building or its occupants exist.

J.A. Senick (✉) · C.J. Andrews · H.C. Putra · I. Tsoulou · M.S. Allacci
Rutgers University, New Brunswick, NJ 08901, USA
e-mail: jsenick@rci.rutgers.edu

W.F.E. Preiser et al. (eds.), *Building Performance Evaluation*,
DOI 10.1007/978-3-319-56862-1_5

The synthetic population approach offers advantages over current methods for incorporating quantitative insights about occupant behavior into the design process: it preserves co-varying and interdependent relationships among behavioral and environmental variables and does not risk compromising confidential data. Additionally, it makes insights more transferable across behavior settings because data on contextual variables accompany the behavioral data. This allows hypothesis testing regarding the generalizability, or context-dependence, of behavioral patterns.

While procedures for developing synthetic populations have not previously been applied to the study of building occupant behavior, they have been well represented in other fields such as transportation planning and public health. An important resource in this area is FRED (Framework for Reconstructing Epidemic Dynamics), an open-access census-based synthetic population used in agent-based epidemic modeling (Grefenstette et al. 2013). Another large synthesized database created by RTI International assigns US population in 50 states and the District of Columbia to schools, workplaces, military bases, dormitories, prisons, and nursing homes and other contextual settings (Wheaton et al. 2009).

In generating synthesized populations, there are four commonly utilized approaches: deterministic reweighting, conditional probability, Monte Carlo simulation, and combinatorial optimization, e.g., simulated annealing. Each of these has advantages and disadvantages, depending on such factors as amount and quality of data, and computing resources available (Harland et al. 2012; Wheaton et al. 2009). This implementation using R, an open source platform for statistical computing, relies on classification and regression trees (CART) to generate synthetic populations (Nowok et al. 2015).

5.3 Creating a Synthetic Population of Building Occupants

This section presents a 10-step process of assembling three diverse occupant behavior datasets to serve as the basis for a synthetic population of building occupants.

- Establishment of a theoretical/methodological rationale based on variables of interest.
- Identification of studies that include these variables and measures.
- Location of data and coding for each study.
- Identification of common variables of interest.
- Establishment of an equivalence basis for converting the coding schemes to a common one.
- Pooling of the discrete datasets into one database.
- Analysis of key multivariate relationships in the underlying datasets.

- Reproduction of these investigations utilizing the aggregated database.
- Creation of synthesized database.
- Validation of synthetic database.

Rationale for Combining Datasets (Step 1):
The rationale for combining datasets is to gain greater predictive power in understanding occupant behaviors, while establishing the foundation for a synthetically enhanced database as a robust and reliable representation of expected building occupant behavior. For instance, in Andrews et al. (2016), it is demonstrated that better information about adaptive behaviors greatly improves energy model accuracy.

Identification of Existing Datasets (Steps 2–3):
The individual datasets drawn upon here are comprised of workplace-based, cross-sectional, longitudinal, time-series POE (post-occupancy evaluation) studies, and a large-scale ASHRAE RP-884 dataset. They are variable in terms of occupant age, sex, daily average outdoor temperature, relative humidity, and indoor air temperature. However, all sets show some significantly consistent patterns, especially regarding variables of interest, e.g., availability of portable space heaters, availability of portable fan, and their frequency of use. Links to all three datasets are provided in the bibliography.

- Cross-Sectional Dataset

The cross-sectional dataset contains variables on thermal and lighting comfort and satisfaction as drawn from 6 separate POEs conducted by Rutgers Center for Green Building researchers (2016) in 16 low- to mid-rise commercial office buildings between 2009 and 2014. Two of the studies include time-series data on occupants' responses to building conditions during actual and simulated demand response events, i.e., load shedding experiments; the majority of the data is cross-sectional. A total of 954 occupant records were collected.

- Longitudinal Dataset

This dataset consists of 24 occupant records of thermal comfort of responses to questions about thermal comfort and related behavioral adaptations from a longitudinal case study of a single office building. Twice-daily online surveys were administered for two-week periods in four seasons of one year for 2012 and 2013 and were accompanied by more frequent observations through datalogger measurements of indoor and outdoor temperatures and other environmental factors (Langevin 2015; Langevin et al. 2015).

- ASHRAE-RP-884 Dataset

Data on thermal comfort were collected from multiple projects and various researchers of 160 buildings worldwide to assemble this mix of cross-sectional and longitudinal data recorded during 1982–1997. All projects utilized a standard template that organized records in the following groups of variables: basic identifiers, thermal comfort questionnaire, indoor climate physical observations, calculated indices, personal environmental control, and outdoor meteorological observations. The resulting 20,215 occupant records additionally incorporated quality control measures throughout an adaptive modeling method (de Dear et al. 1997; de Dear and Brager 1998).

Establishing the Basis for the Pooled Database (Steps 4–5):
The individual datasets, although diverse, share a large number of variables in common. In some cases, the common variables of interest were coded differently. Such differences need to be addressed to create the pooled database. There were also a number of fields that existed in only some of the sets, but could be easily reproduced. Lastly, there were a number of fields that were relevant to the research scope but did not have values in all sets.

The next step is to create a standard template for the pooled database that takes into account similarities and differences among the datasets. The resulting template includes all common fields from the 3 sets, the number of fields that could be easily reproduced, as well as those that were incomplete, but still relevant to the scope of the research project and deemed valuable for future use.

In order to comply with the final database template fields, units and coding, the underlying datasets require modification. Illustratively, to the cross-sectional dataset 7 fields were added and sources, units and coding were standardized relating to such items as meteorological data, indoor environmental observations, lighting, and thermal adjustment access and use. Similar modifications were made also to the longitudinal and ASHRAE RP-884 datasets.

Creation of the Final Database (Steps 6–8):
To locate common explanatory variables to predict occupant behaviors, logistic regressions were run against the three data sets and the combined data set. Table 5.1 shows that with significantly more observations than the other two data sets, ASHRAE RP-884 performs the best in predicting a variable of interest—use of portable fan, while cross-sectional data performs the worst. Regression results for the pooled data set show that mean indoor air temperature, occupant's age, and occupant's gender significantly predict the use of portable fan at the 0.001 level. Outdoor air temperature explains the use of portable fan at the 0.01 level of significance.

The final database acquires characteristics of the three underlying datasets in proportion to the number of observations in each. As expected, the larger ASHRAE RP-884 dataset is most influential in the combined database, while the cross-sectional dataset, which is the smallest, is least influential. Compared to the others, ASHRAE RP-884 has a higher mean age, more naturally ventilated buildings, and more measurements from warmer climates.

Generating and Validating Synthetic Population (Steps 9 and 10)
Generating a synthetic version of data fits the original dataset to the assumed distribution and obtains its parameters estimates. For detailed instruction on how to generate the synthetic population using R, interested readers may consult the package documentation (Nowok et al. 2015).

Table 5.1 Regression results comparing the three data sets and synthetic database *Source* Authors 2016

Selection	Longitudinal		Cross-sectional		RP-884		Combined	
Data set	Observed	Synthetic	Observed	Synthetic	Observed	Synthetic	Observed	Synthetic
Dep. Var.	Hportfan							
Expl. Var.	Coeff (SD)							
Dayav_ta	−0.014* (0.006)	0.001 (0.006)	−0.03 (0.05)	0.004 (0.041)	0.032*** (0.004)	0.059*** (0.004)	0.01** (0.003)	0.03*** (0.003)
TA_M	0.182*** (0.04)	0.131*** (0.039)	0.43* (0.24)	0.05 (0.175)	0.421*** (0.011)	0.336*** (0.009)	0.419*** (0.010)	0.342*** (0.009)
Age	0.450*** (0.05)	0.427*** (0.046)	−0.24* (0.137)	−0.129 (0.169)	−0.283*** (0.023)	−0.246*** (0.023)	−0.067*** (0.018)	−0.04* (0.018)
Sex	0.405** (0.41)	0.266* (0.124)	−0.82* (0.367)	−0.535 (0.432)	0.185** (0.054)	0.296*** (0.063)	0.281*** (0.047)	0.366*** (0.05)
_cons	−7.55*** (0.96)	−6.12*** (0.915)	−8.12 (5.84)	−1.18 (4.54)	−11.2*** (0.263)	−9.88*** (0.246)	−11.29*** (0.249)	−9.984*** (0.232)
Number of obs	2023	2071	147	128	10,496	10,289	12,666	12,488
LR $chi^2(5)$	193.92	167.67	13.18	2.67	4975.87	4488.54	4719.94	4204.32
Prob > chi^2	0.000***	0.000***	0.01*	0.6152	0.000***	0.000***	0.000***	0.000***
Pseudo R^2	0.0785	0.0642	0.0674	0.0191	0.354	0.3241	0.2813	0.2524

Asterisks show which variables are significant at the 0.1 level (*), 0.01 level (**), and 0.001 level (***)

Table 5.2 Descriptive statistics of pooled observed data and synthesized data. *Source* Authors 2016

Variable	Source	Mean	SD	Minimum	Maximum	N
Age	Observed	2.72	1.42	1	7	21,990
	Synthetic	2.73	1.42	1	7	22,004
Sex	Observed	1.51	0.5	1	2	27,500
	Synthetic	1.51	0.5	1	2	27,465
Dayav_ta	Observed	18.5	9.92	−24.9	35	28,299
	Synthetic	18.48	9.89	−24.9	35	28,262
TA_M	Observed	22.22	7.91	0	42.67	27,814
	Synthetic	22.21	7.92	0	42.5	27,804
Hportfan	Observed	0.37	0.48	0	1	16,780
	Synthetic	0.38	0.48	0	1	16,551

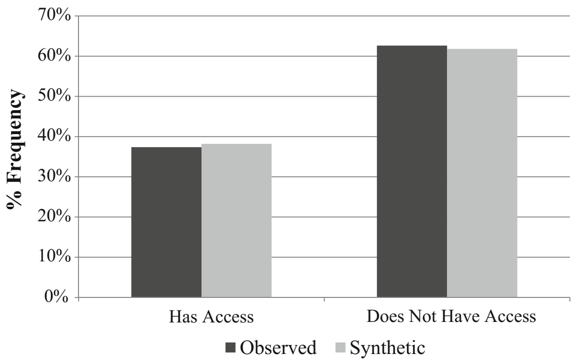

Fig. 5.1 Relative frequency distribution of use of portable fan (hportfan) for observed and synthetic data. *Source* Authors, 2016

After running the synthesis process, the resulting data needs to be validated with the original observed data by comparing statistical inferences. In this example, an ordered logistic regression is estimated where the dependent variable is the use of portable fan. For explanatory variables, the model uses outdoor air temperature, indoor air temperature, sex, and age. Descriptive statistics compare actual data and synthesized data. Results from the descriptive statistics support the conclusion that both data have similar measures of central tendency and similar measures of spread (see Table 5.2).

Another way to compare both data sets is by evaluating the relative frequency distributions of key variables. In this example, the frequency distributions of the use of portable fan in the observed data set and the synthetic one are compared (see Fig. 5.1).

Alternatively, the synthesis process can be validated by comparing goodness-of-fit for both the observed and synthesized data (see Fig. 5.2). Estimates of the 95% confidence interval and Z statistics from a logistic regression of the use of portable fan for both data were relatively similar. These findings show that in both data sets, mean indoor air temperature and gender are significant at 0.001 level. Both data sets also show that outdoor temperature is significant at the 0.001 level for the cohort of buildings with HVAC.

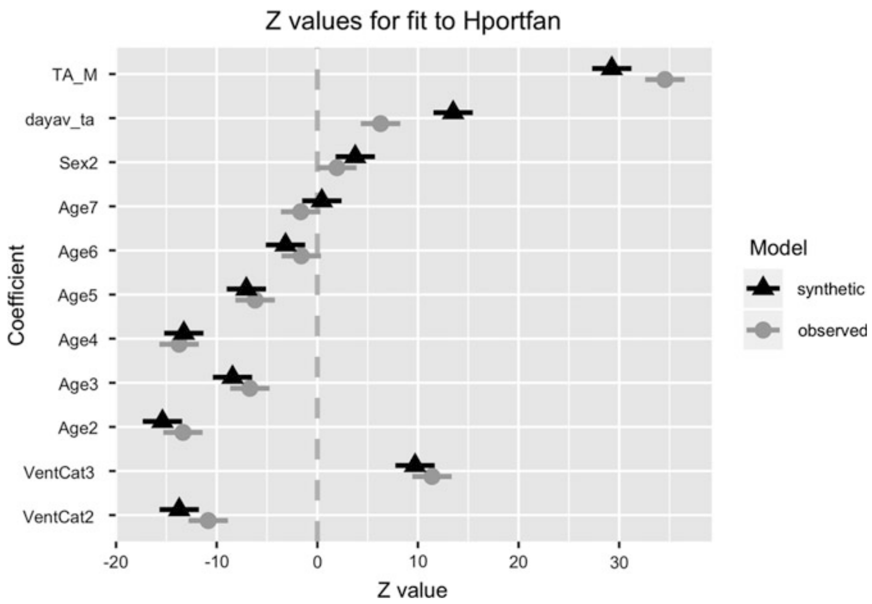

Fig. 5.2 Estimates for Z statistics from a logistic regression of the use of portable fan for both observed and synthetic data. *Source* Authors, 2016

5.4 Uses of Synthesized Data

The creation of a synthetic population of building occupants and their behaviors enables researchers and designers of buildings to use these synthetic data in Building Energy Modeling (BEM), Agent Based Modeling (ABM), and other co-simulation methods thereby extending the framework previously advanced by Andrews et al. (2011), and Andrews et al. (2012). Synthesized data is additionally a promising basis for supplementing the practice of generative POE (Wener et al. 2016), allowing for grounded-hypotheses derived from case study data to be evaluated using the larger synthetic database.

5.5 Methodological Implications and Limitations

This chapter demonstrates that creating a synthetic set of generic building occupants is feasible for leveraging smaller POE-based studies of building occupant behavior. Guidance for assembling disparate databases is provided, resulting in an adequate foundation of building occupant characteristics and behaviors from which to generate a representative population using the R statistical software package. The synthetic database preserves confidentiality while capturing building occupants' interactions and covariation in commercial buildings. The application of long-standing synthetic population techniques to the study of building occupant behavior presents an exciting frontier for overcoming the current constraint of many case studies. Nevertheless, a few cautionary notes are appropriate. The process of assembling the combined database requires attention to detail and a willingness to make subjective judgments about contexts from which data was collected, missing data fields, and needed data transformations. Pooled datasets may be relatively rich in data for some variables and lean for others. In the combined dataset, there is more data on thermal comfort, and less on lighting, adaptive responses and social or organizational aspects of behavior. To help overcome these challenges, it is important for behavioral researchers to work together to utilize a more standardized data collection template.

5.6 Conclusion: Theoretical Implications and Future Research

In future and on-going work, the authors extend the synthetic database approach beyond standard hypotheses about building occupant behavior to a series of analyses that delve deeper into the roles of organizational determinants of behavior and contexts, and seek to integrate energy modeling and agent based modeling simulations. These analyses intend to draw on data derived from POEs as well as from large scale representative databases such as the US Department's Commercial Building Energy Consumption (CBECS) database and the American Community Survey of the US Census Bureau. Other opportunities for enhancing studies of building occupant behavior with synthetic data reside in 3-D imaging techniques for more detailed characterizations of building contexts.

Acknowledgements This research was supported by the Consortium for Building Energy Innovation, sponsored by the US Department of Energy Award Number DE-EE0004261.

References

Andrews, C. J., Yi, D., Krogmann, U., Senick, J. A., & Wener, R. E. (2011). Designing buildings for real occupants: An agent-based approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 41*(6), 1077–1091.

Andrews, C. J., Senick, J. A., & Wener, R. E. (2012). Incorporating occupant perceptions and behavior into BIM. In S. Mallory-Hill, W. Preiser, & C. Watson (Eds.), *Enhancing building performance* (pp. 234–246). Oxford, UK: Blackwell Publishing.

Andrews, C. J., Allacci, M., Senick, J. A., Putra, H. C., & Tsoulou, I. (2016). *Using synthetic population data for prospective modeling of occupant behavior during design.* Accepted for publication in Energy and Buildings special issue on occupancy behavior.

Bordass, B., Leaman, A., & Ruyssevelt, P. (2001). Assessing building performance in use 5: Conclusions and implications. *Building Research & Information, 29*(2), 144–157.

de Dear, R. J., Brager, G. S. & Cooper, D. (1997). ASHRAE RP-884 dataset link: http://sydney.edu.au/architecture/staff/homepage/richard_de_dear/ashrae_rp-884.shtml. Accessed 04 May 2016.

de Dear, R. J., & Brager, G. S. (1998). Towards an adaptive model of thermal comfort and preference. *ASHRAE Transactions, 104*(1), 145–167.

Grefenstette, J. J., Brown, S. T., Rosenfeld, R., DePasse, J., Stone, N. T. B., Cooley, P. C., et al. (2013). FRED (a framework for reconstructing epidemic dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. *BioMed Central Ltd*. doi:10.1186/1471-2458-13-940.

Harland, K., Heppendstall, A., Smith, D., & Birkin, M. (2012). Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques. *Journal of Artificial Societies and Social Simulation 15*(1):1. http://jasss.soc.surrey.ac.uk/15/1/1.html. Accessed 04 May 2016.

Heerwagen, J. (2000). Green buildings, organizational success and occupant productivity. *Building Research & Information, 28*(5), 353–367.

Hewitt, E., Andrews, C., Senick, J., Wener, R., Krogmann, U., & Sorensen Allacci, M. (2015). Distinguishing between green building occupants' reasoned and unplanned behaviors. *Building Research & Information*. doi:10.1080/09613218.2015.1015854.

Hong, T., D'Oca, S., Turner, W. J. N., & Taylor-Lange, S. C. (2015). An ontology to represent energy-related occupant behavior in buildings, Part I: Introduction to the DNAS framework. *Building and Environment, 92*(2015), 764–777. doi:10.1016/j.buildenv.2015.02.019.

Langevin, J. (2015). Longitudinal dataset. http://en.openei.org/datasets/dataset/one-year-behavior-environment-data-for-medium-office. Accessed 04 May 2016.

Langevin, J., Wen, J., & Gurian, P. L. (2015). Simulating the human-building interaction: Development and validation of an agent-based model of office occupant behaviors. *Building and Environment, 88,* 27–45. doi:10.1016/j.buildenv.2014.11.037.

Leaman, A., & Bordass, B. (1999). Productivity in Buildings: the "Killer" variables. *Building Research & Information, 27*(1), 4–20.

Nowok, B., Raab, G. M., & Dibben, C. (2015). *Synthpop: Bespoke creation of synthetic data in R. Package*. vignette http://cran.r-project.org/web/packages/synthpop/vignettes/synthpop.pdf. Accessed 22 Nov 2015.

Rutgers Center for Green Building. (2016). Cross-sectional dataset (2009–2014). http://en.openei.org/datasets/dataset/ob-commercial-building. Accessed 04 May 2016.

Wener, R., McCunn, L., & Senick, J. (2016). Did that plan work? Post occupancy evaluation. In R. Gifford (Ed.), *Research methods for environmental psychology* (pp. 249–270). Hoboken, NJ: Wiley-Blackwell.

Wheaton, W. D., Cajka, J. C., Chasteen, B. M., Wagener, D. K., Cooley, P. C., Ganapathi, L., et al. (2009). Synthesized population databases: A US geospatial database for agent-based models. *Methods Report*. doi:10.3768/rtipress.2009.mr.0010.0905.

Author Biographies

Jennifer A. Senick Rutgers Center for Green Building and Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Senick holds a Ph.D. in Planning and Public Policy from Rutgers, an M.A. in Political Science from UCLA, and an A.B. in Government from Bowdoin College. She is Executive Director of the Rutgers Center for Green Building at the Edward J. Bloustein School of Urban Planning and Policy Development, Rutgers University and is well published on topics in green building and sustainable development, environmental behavior, evaluation methods. Current appointments include as Board Member and Treasurer of the Environmental Design Research Association (EDRA) and co-chair of the EDRA Sustainable Planning Design and Behavior Network. Senick has worked in numerous private sector positions in areas of emerging technology and has founded two successful start-up companies in energy planning and environmental assessments.

Clinton J. Andrews Rutgers Center for Green Building and Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Andrews, Ph.D. PE AICP, was educated at Brown and MIT as an engineer and planner. He is Director of the Center for Green Building, Professor of Urban Planning and Policy Development, and Associate Dean for Planning and New Initiatives at the Edward J. Bloustein School, Rutgers University, New Brunswick, New Jersey. His research addresses behavioral, policy and planning questions related to energy use in the built environment. He has worked in the private sector as a design engineer and technology assessor, helped launch an energy planning project at MIT, and helped to found a science policy program at Princeton. At Rutgers, he has launched initiatives in energy planning and green building. He has published over 100 scholarly and popular articles and his books include *Humble Analysis: The Practice of Joint fact-Finding*, *Regulating Regional Power Systems*, and *Industrial Ecology and Global Change*. He is the new editor (with Frank Popper) of the *Journal of Planning Education and Research*.

Handi Chandra Putra Rutgers Center for Green Building and Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Chandra Putra holds a Master of City and Regional Planning (MCRP) from Rutgers University. He is a Research Assistant at the Rutgers Center for Green Building and a doctoral student at the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. His dissertation investigates the effects of coastal flooding on housing market dynamics by using spatial econometrics and agent-based modeling (ABM). He has served as a reviewer to the Journal of Artificial Societies and Artificial Social Simulation.

Ioanna Tsoulou Rutgers Center for Green Building and Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Tsoulou received a Master of City and Regional Planning (MCRP) from the Edward J. Bloustein School of Planning and Public Policy, Rutgers University, NJ with concentration in Environmental Planning and Design and a Master of Engineering from the Department of Civil Engineering, University of Patras, Greece with Concentration in Transportation and Environmental Engineering. She is a Research Assistant at the Rutgers Center for Green Building and a doctoral student at the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Her research interests lay in the areas of Sustainability and Environmental Land Use Planning, Intelligent Transportation Systems, Green Buildings, Neighborhood Planning, Urban Design and Smart Growth.

MaryAnn Sorensen Allacci Rutgers Center for Green Building and Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Sorensen Allacci, has a Ph.D. from CUNY's Environmental Psychology Program. She has been the Research Project Coordinator for the Rutgers Center for Green Building and specializes in community based collaborative research and engagement with organizations and diverse groups including low income, ethnic minority communities, and individuals with disabilities. Her particular focus is environmental health and environmental justice concerns in housing and community environments, and she teaches these and other aspects of to interior designers. Dr. Sorensen Allacci is also president of the board for Projects for Environmental Health, Knowledge, and Action, Inc., a non-profit organization.

Wolfgang F.E. Preiser · Andrea E. Hardy
Ulrich Schramm *Editors*

Building Performance Evaluation

From Delivery Process to Life Cycle Phases

Second Edition

Springer

Building Performance Evaluation

Wolfgang F.E. Preiser · Andrea E. Hardy
Ulrich Schramm
Editors

Building Performance Evaluation

From Delivery Process to Life Cycle Phases

Second Edition

Springer

Editors
Wolfgang F.E. Preiser (deceased)
University of Cincinnati
Cincinnati, OH
USA

and

Arizona State University
Scottsdale, AZ
USA

Andrea E. Hardy
Creo Architects
Phoenix, AZ
USA

Ulrich Schramm
Bielefeld University of Applied Sciences
Minden, Bielefeld
Germany

ISBN 978-3-319-56861-4 ISBN 978-3-319-56862-1 (eBook)
DOI 10.1007/978-3-319-56862-1

Library of Congress Control Number: 2017939094

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In Memory of Wolfgang F.E. Preiser
colleague, teacher, mentor, and friend