VIEW FROM THE TOP

Data in SEM: A 50-Year Perspective

By Tom Green

One of the great throughlines of SEM is the use of data. Foundational to creating SEM as a field and as a practice, data usage has evolved and become more known, practiced and refined as our field has become more common and as expectations have increased. Perhaps its greatest utility is the way in which it allows us to start conversations with our institutional peers and understand the student experience. That throughline is likely to extend well into SEM's next 50 years, and beyond.

As we consider the 50 years since John "Jack" Maguire (1976) wrote for the first time about enrollment management in Boston College's *Bridge* alumni magazine, much has changed, but much remains the same. One of the throughlines of the field has been the use of data to inform decision-making and improve enrollment outcomes. While the adage that data puts the "strategic" in strategic enrollment management (SEM) has been around for many years, Hutton (2021) summarizes the current state of data and research use in the field and makes the case for this adage as a cornerstone of the field.

SEM has evolved and continues to evolve. This article will track that evolution from the perspective of data and its role in SEM with the goal of drawing the overarching view of data as a fundamental element of SEM. It is in broad strokes, with highlights from the last five decades.

The Foundational Years of SEM

Maguire's original article (1976) cited declining birth rates and a projection that high school graduates would drop by 30 percent by the late 1980s. This attention to U.S. Census data and the use of data projections would become a constant element of enrollment managers' work for the next 50 years.

Three years later, the Western Interstate Commission on Higher Education (WICHE) released its first publication on high school graduation rates. *High School Graduates: Projection for the Fifty States* (WICHE 1979) was an extension of the original publication that year that covered WICHE's thirteen-state region. The data projections were grouped by region and followed by state projections within that region, looking out fifteen years.

Savvy admission deans like Maguire and Tom Huddleston at Bradley University had a huge advantage in these early years. High school graduate numbers were still growing, or at least very healthy in most states. Early SEM practitioners gained market share from regions outside their historic recruitment areas. This early positioning allowed these institutions to establish relationships in markets that would benefit them during the years to come, and to reposition their insti-

tutions as more selective by growing application pools before the declines in high school graduates arrived (McDonough 1997).

New tools emerged in the early 1970s that helped schools target their new territories. Both ACT (multiple references point to a 1971 annual report of the American College Testing Service, but that original report is unavailable) and the College Board (Lucier 2006) began offering colleges the names of test-takers as a service that allowed them to market directly to these students. This served to increase access and awareness of college opportunities and also signaled the start of scientific marketing approaches to enrollment.

Measure Everything

The early years of SEM established important practices, including benchmarking and the measurement of return on investment (ROI) of recruitment activities (Hossler 1984). As admissions deans pushed into new geographic areas, they had to justify the expenses associated with sending recruiters to those areas, increased direct mail, and marketing. For example, if University X purchases 20,000 names of test-takers in a new state (California), what percent of those names returned an inquiry card, called the University for more information, set up a campus visit, came to a high school visit, or stopped at a table at a college fair? These tangible signs of interest were measured as conversion rates toward application and eventual enrollment.

Quantifying the expenses to measure ROI required additional data to be collected and analyzed. How much was spent on travel? What was the cost of purchasing names, printing and mailing letters with business reply cards, printing anything else that may be needed for recruitment fairs, costs for college fair tables, etc.? Enrollment managers began to segment their expenses and calculate the ROI of entering new markets. Typed reports, and later spreadsheets, were created and kept to compare ROI across years. Were we making progress on cracking into this new market, or did the costs outweigh tuition earned from these initiatives?

Questions also arose about expectations for ROI and response rates. What is the incremental increase in response if mailing a second letter? What should a school expect as the response rate from one or two mailings? What are other schools spending on recruiting trips, advertising, and printing? These early measurements became benchmarks shared informally among colleagues and then at sessions during AACRAO conferences or other meetings. In 2003, the National Association for College Admission Counseling (NACAC) produced its first *State of College Admissions* report and tracked the cost of recruitment, staff sizes, and other benchmarks that helped admissions and enrollment professionals compare expenses.

Because test-taker names were sold in geographic groupings, enrollment managers became experts in zip codes. So did the testing companies that sold them. Population-rich metropolitan areas were broken into zip code groups that required purchasing multiple groups to cover the city and its suburbs. Was it better to buy the whole state? ROI calculations became critical to decision-making and budget stretching.

As the internet changed the world, recruiting and marketing approaches and techniques were forced to adapt. Response rates to email campaigns were different from postal rates, and information could be spread across multiple messages and a website. Landing page visits and time on specific pages of the website were now visible and measurable, allowing enrollment managers to focus resources on the pages most important to prospective students. Marketing response rates went from general activity to highly specific click rates and responses to specific keywords, as these new and improved tools emerged (Hasan, Morris, and Probets 2009; Marek 2011).

The growth in data from communications required that new information systems be developed to capture, manage, measure, and analyze them. Spreadsheets were no longer manageable. Higher education borrowed tools from business and implemented customer relationship management (CRM) systems. This helped enrollment managers gain insights into specific campaigns, topics, and headlines that earned the highest response and click rates. The concept of A/B testing emerged from this data that allowed experimentation with messages

and subject lines, providing precise measurement of preferred or resonant topics and content.

Institutional Aid Optimization and Enrollment Models

One of the most influential practices in SEM is the use of data to inform institutional financial aid strategy (Lassila 2011; Nguyen, Pike, and Hou 2018). Using econometric theories of price elasticity and supply and demand, a small group of enrollment managers began to experiment with offering differential amounts of institutional scholarships or merit aid to admitted students. These studies used a regression discontinuity approach to compare those who received aid to those who did not (Leeds and DesJardins 2015; van der Klaauw 2002). Further, the approach was based on the premise that increasing the amount of gift aid (which did not have to be repaid, nor required work to receive) had a positive impact on the enrollment decision. It was observed that even a shift to grants over loans had a positive impact on enrollment and disproportionately so for students from underserved backgrounds (Linsenmeier, Rosen, and Rouse 2006).

The downturn in first-time students in the mid-1980s and early 1990s accelerated interest in data analysis and strategic use of institutional aid. Price sensitivity drove private, independent institutions that lacked large endowments and donor bases to try a new approach. They inflated the sticker price of tuition and used a small portion of gross revenue to create institutional aid pools. The new pools of institutional aid were used initially as a blunt instrument to drive up headcount, then in more targeted and refined ways to align aid strategy with institutional enrollment goals. By analyzing the yield rates among different groups of students and the amount of gift aid they received, the enrollment manager could strategically add aid to increase yield or redistribute aid from high-yield areas to those that needed increases to reach enrollment goals (Lassila 2011).

The outcomes of aid optimization were used to enhance enrollment models. Colleges and universities wanted to know the broader impact of enrollment changes on academic programs, residential facilities, and net tuition revenue from different cohorts, such as undergraduate, international, and graduate students. There are examples of enrollment models that predate SEM. Enrollment projection models that are not tied to aid appeared in 1972 at Kansas State University (Orwig, Jones, and Lenning 1972); econometric studies that analyzed enrollment by socioeconomic groups appear that same year (Corazzini, Dugan, and Grabowski 1972). More refined models appeared less than a decade later, when researchers examined the variances among incoming first-year students by academic program at the University of Minnesota in 1978 (Weiler 1980).

Modeling that incorporates price and aid began to appear in the late 1980s. Leslie and Brinkman (1987) provided a foundation for assessing the impact of lowering price by offering gift aid that offsets that price, using increments of $100 to assess changes in enrollment behavior. Broad studies of college-going groups reinforced theories that aid worked to influence enrollment behavior; these were initially focused on the efficacy of federal grants (McPherson and Schapiro 1991).

Building on these foundations, enrollment models could estimate changes in policy (*i.e.*, distribution of institutional aid) and improvements in performance (*i.e.*, increased retention). Once tied to revenue, the models could produce estimates of the net tuition generated by investments in aid and in student support programs and resources. The models could also estimate the variable outcomes and impacts on different segments of the student body, such as campus residents, academic programs, and resident/nonresident populations for public institutions (Goenner and Pauls 2006).

National Benchmarking and Data Initiatives

Tuition discounting, the practice of taking some gross tuition revenue to create pools of institutional aid, became so widespread that in 1994 the National Association of College and University Business Officers (NACUBO) developed a standard formula for calculating the percentage of institutional aid against gross tuition and fee revenues. This discount rate formula allowed private institutions to benchmark their practices against

peers. While the practice spread to public institutions in the early 2000s, it was not until 2020 that NACUBO began to also evaluate the rate of discount at public universities.

Starting in 1979, researchers from WICHE were able to compile data from their member Western states, then broadened their work to include the entire United States. Starting in 1984, *Knocking at the College Door* was released every four years. The report attempted to account for changes in birth rates, mortality, and migration patterns, among other variables. This quickly became a valuable resource to explore what states and regions of the country were growing in high school graduates, allowing enrollment managers to target those areas and position resources to capture that growth. As data elements in the report expanded, more information about the makeup of these students emerged and targeting became more precise.

An equally powerful and valuable set of tools emerged when the National Student Clearinghouse established its Research Center (NSCRC) in 2010. The early reports provided timely access to information on enrollment, sliced in various ways, across the country. While the data collected through the Integrated Postsecondary Education Data System (IPEDS) surveys is more complete and contains much richer data, the lag time between collection and use meant any analysis of those data was dated and perhaps no longer relevant. The NSCRC data came out just a few months after the term began, allowing for more immediate shifts in enrollment strategy.

Perhaps the most influential reports from the NSCRC have been on the number of people in the United States with some college but no degree (NSCRC n.d.). The illumination of the data on this population shifted the course of enrollment strategy across the country and informed public policy across many states. As the reports expanded, information by state became available and institutions were able to create delivery modalities and patterns, including online, hybrid, accelerated degrees, and short-term certificates to attract and educate the tens of millions of people who started but never completed a degree.

Data Reporting: From Bar Paper to Visualizations

Early computers were innovative tools that were capable of storing student application and academic record data. Output often took the form of lists, and some calculating power existed that allowed advanced users to create reports (number and type of applications, admissions, enrolled, course counts, etc.). These were printed on green-bar paper from dot matrix printers. Computer time was at a premium and the reporting took time, as the computer had to compile the information, often over several hours or overnight (Panzatis 2017; University of Washington n.d.).

Spreadsheet tools such as Lotus 1–2-3 and Excel emerged, allowing end users to take lists of data and manipulate them into reports. These could be disaggregated into segments, such as recruiting territories, academic programs, or preparation characteristics (GPA, test scores, etc.). These provided greater insights into application and enrollment patterns; they demanded that enrollment managers gain fluency with spreadsheets (Ewell 1985).

Static data downloaded from an information system gave way to real-time views of the data. Early versions resembled early report forms, largely rows and columns of information. Data scientists began to apply data visualization techniques that had evolved over hundreds of years (Korengold 2023) to these data. These graphs, which used various shapes and colors—often sized in proportion to the number of cases represented in a larger dataset—allowed report recipients to understand the data more quickly.

These early forms became the modern dashboards that are common and expected in today's CRM reports. Users can drill down into the data to discover greater levels of detail about groups of students within the data, or even individual cases.

Predictive Analytics and Artificial Intelligence

It is difficult to pinpoint when predictive analytics entered the enrollment field, as the practice likely began early but went undocumented. In 1992, Whitman Col-

lege Director of Institutional Research Ronald Urban published a case study on the use of data to assign weights to admitted students in the 1988 entering class (Urban 1992). This is significant because it went beyond regression analysis to convert historical data into predictive weights for the likelihood that an individual admitted student might enroll.

This practice expanded over the next 20 years, and the precision of the predictions increased as well. With several data points, the likelihood of enrollment moved from the broad low-medium-high categories used by Urban (1992) to percentages for individual cases. The statistical expertise to complete this work often resided outside the enrollment management unit or admissions office. Companies and consultants provided the modeling and resulting likelihood-to-enroll percentages, or "scores." They were helpful in determining where to place limited resources, and strategies emerged to target students from particular score ranges.

Early predictive modeling had its drawbacks. Because it was based on historical data, it produced high scores for those applicants who looked like earlier successes in recruitment. In the tails of the models, perhaps two or three standard deviations from the mean, were the aspirations of strategy—honors students, nonresidents, and students from historically excluded backgrounds. They also froze the results in time. Because they were often externally scored, a file of prospective students was sent out and returned. Any new data that was received or generated, such as changes in preference of program, campus visits, or relationships that were discovered, could not be used to update the score (CustomViewbook 2026).

Models to predict persistence and completion were also developed. Many early studies of retention were univariate, posing a single data point (*e.g.*, high school GPA) against retention rates. While the independent variable may change (*e.g.*, test scores, residency, program of study), the interaction of variables found in more complex regression models was less common.

As statistical techniques advanced, multivariate analyses became more commonly employed to understand the profiles of students who were more or less likely to persist or complete their programs of study. By 2007, EDUCAUSE produced a case study review (Campbell, DeBlois, and Oblinger 2007) of four institutional types and their deployment of advanced statistical models. Each focused on different use cases and the variety of institutional types was intentional. No one model was the solution for all institutional types.

The Campbell article accurately predicted that analytics would play a greater role in student success in the future. More and more institutions sought tools to understand the factors that led to success or departure. Institutional researchers brought greater levels of sophistication to the task.

Classification trees, an early form of machine learning, used algorithms to create nodes and branches that showed how data related to the dependent variable (*i.e.*, persistence or completion). These revealed an important fact of student success: one variable could be positively associated with one group of students and negatively associated with another. For example, student loans could enable one group of students to persist and complete, while they were a barrier for another group in the same study. These results gave greater and more individualized insights than logistic regression, which provided the power of a variable to explain the outcome of the independent variable but only in one direction and at an aggregate level (Loh 2014).

Artificial intelligence (AI) promises to be one of the most transformational data tools in the last 50 years. The science of neural networks is not new and can be traced back to the 1940s (Hardesty 2017). The network learns from training sets of data and then creates nodes and layers of nodes or clusters of data that appear to be related. In somewhat similar fashion to classification trees, these nodes can have multiple relationships (or multivectors). The network attempts to answer the question of the original dependent variable, which can be expressed in natural language such as the likelihood an applicant will enroll.

Advances in computing power and data storage have created an explosion in both predictive and generative forms of AI. Rather than relying on third-party analyses of static data, institutions can now predict enroll-

ment, persistence, or completion within information and engagement systems, using real-time data to update predictions continually. One of the major advances is the availability of predictive analytics to front-line recruitment and student success personnel. Generative AI allows users to query the data with natural language prompts rather than to master statistical techniques to investigate it. Recruiters, advisors, tutors, academic planners, managers, and executives can have predictive AI tools embedded into their daily practice and systems. As these new tools proliferate, they place demands on institutions for greater data governance and for training end users on the ethical and practical uses of new AI tools (Oliphant and Smith 2025).

Summary

SEM is and has been grounded in data. From its origins, it sought to understand internal and external environments through data (U.S. Census, student records, institutional capacity) and, from these observations, to create measurable strategies. Enrollment managers have required increasingly sophisticated tools to analyze data and to gain insights from them. The need to prove return on investment for innovative approaches created pressure to measure, record, and assess each strategy. As data increased, they created demand for better methods and tools to manage and understand data (Fingerson and Troutman 2019).

Reporting and sharing data is one way enrollment managers inform their peers in the academy and forge relationships with them. It requires enrollment managers to understand and communicate the insights gained from data analysis, and to listen to and understand the needs and questions of academic peers.

Like SEM, data use in this context continues to evolve. Enrollment managers continue to adapt their practices and strategies in response to new tools and expectations. The advent of AI requires that data be coordinated and unified across the institution. With dozens of disparate systems creating and storing data on student experiences and outcomes, coordinating and unifying these data becomes a tremendous challenge. The use of potentially transformative AI tools is limited by the availability of data on the student experience, both in and outside the classroom.

As we look forward to the next 50 years of data in SEM, it is clear that data will maintain its role as the bedrock of our profession. Tools and techniques will continue to evolve, and access to them for end users and front-line personnel will increase. Enrollment managers will be called upon to have accurate, up-to-the-minute data on prospective and current students, and to have insights about what those data may indicate about progress toward institutional goals and aspirations. Finally, data will continue to serve as the calling card for conversations and strategic planning with peers across the institution. Mastery of data techniques and tools will help enrollment managers drive practices and policies that attract and support students toward their educational goals. This mastery also demonstrates the value that enrollment managers bring to their institutions.

References

Campbell, J. P., P. B. DeBlois, and D. G. Oblinger. 2007. Academic analytics: A new tool for a new era. *EDUCAUSE Review.* 42(4): 40–57. Retrieved from: <er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era>.

Corazzini, A. J., D. J. Dugan, and H. G. Grabowski. 1972. Determinants and distributional aspects of enrollment in U.S. higher education. *The Journal of Human Resources.* 7(1): 39–59. Available at: <doi.org/10.2307/145056>.

CustomViewbook. 2026. *Predictive vs. Engagement Scoring: Identify Real Intent.* Chesapeake, VA: Author. Retrieved from: <customviewbook.com/predictive-vs-engagement-scoring/>.

Ewell, P. T. 1985. *Recruitment, Retention and Student Flow: A Comprehensive Approach to Enrollment Management Research* (NCHEMS Monograph No. 7). Boulder, CO: National Center for Higher Education Management Systems.

Fingerson, L., and D. R. Troutman. 2019. Measuring and enhancing student success. *New Directions for Institutional Research.* 2019(184): 33–46. Available at: <doi.org/10.1002/ir.20320>.

Lucier, M. 2006. How all those colleges got your name. *The New York Times.* April 23. Retrieved from: <nytimes.com/2006/04/23/education/edlife/how-all-those-colleges-got-your-name.html>.

Goenner, C. F., and K. Pauls. 2006. A predictive model of student enrollment. *Research in Higher Education.* 47(8): 935–956. Available at: <doi.org/10.1007/s11162-006-9021-8>.

Hardesty, L. 2017. Explained: Neural networks. *MIT News.* April 14. Retrieved from: <news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.

Hasan, L., A. Morris, and S. Probets. 2009. Using Google Analytics to evaluate the usability of e-commerce sites. *Lecture Notes in Computer Science.* 5619: 697–706. Available at: <doi.org/10.1007/978-3-642-02806-9_81>.

Hossler, D. 1984. *Enrollment Management: An Integrated Approach*. New York: College Entrance Examination Board.

Hutton, A. 2021. Putting "strategic" in strategic enrollment management. *Strategic Enrollment Management Quarterly.* 9(3).

Korengold, A. 2023. Revealing data: Visualizations in historical collections. *Circulating Now.* July 13. Retrieved from: <circulatingnow.nlm.nih.gov/2023/07/13/revealing-data-visualizations-in-historical-collections/>.

Lassila, N. E. 2011. The relationship of institutional tuition discounts with enrollment. *Journal of Student Financial Aid.* 40(3): Article 2. Available at: <doi.org/10.55504/0884-9153.1030>.

Leeds, D. M., and S. L. DesJardins. 2015. The effect of merit aid on enrollment: A regression discontinuity analysis of Iowa's National Scholars Award. *Research in Higher Education.* 56: 471–495. Available at: <doi.org/10.1007/s11162-014-9359-2>.

Linsenmeier, D. M., H. S. Rosen, and C. E. Rouse. 2006. Financial aid packages and college enrollment decisions: An econometric case study. *The Review of Economics and Statistics.* 88(1): 126–145. Available at: <jstor.org/stable/40042964>.

Leslie, L. L., and P. T. Brinkman. 1987. Student price response in higher education: The student demand studies. *Journal of Higher Education.* 58(2): 181–204.

Loh, W. Y. 2014. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 4(1): 14–23.

Maguire, J. 1976. To the organized go the students. *Bridge Magazine.* 39(1). Fall.

Marek, K. 2011. Getting to know web analytics. *Library Technology Reports.* 47(5): 5–10.

McDonough, P. M. 1997. *Choosing Colleges: How Social Class and Schools Structure Opportunity*. Albany, NY: State University of New York Press.

McPherson, M. S., and M. O. Schapiro. 1991. Does student aid affect college enrollment? New evidence on a persistent controversy. *The American Economic Review.* 81(1): 309–318. Available at: <jstor.org/stable/2006804>.

National Student Clearinghouse Research Center. n.d. *Some College, No Credential*. Herndon, VA: Author. Retrieved from: <nscresearchcenter.org/some-college-no-credential/>.

Nguyen, T.-D., G. R. Pike, and Y. Hou. 2018. Effects of tuition discounting on university performance. Paper presented at the European Financial Management Association Annual Meeting, Milan, Italy.

Oliphant, T., and C. A. Smith. 2025. Transforming enrollment management in the field of online learning: A systems-level approach from business process mapping to AI integration. *Online Learning.* 29(2).

Orwig, M. D., P. Jones, and O. Lenning. 1972. *Enrollment Projection Models for Institutional Planning* (ACT Research Report No. 48). Iowa City, IA: American College Testing Program.

Pantazis, M. A. 2017. *Technological Change in Higher Education Information Systems (1980 to the Present): A Case Study of The Pennsylvania State University* (doctoral thesis). Harvard University, Graduate School of Education. Available at: <nrs.harvard.edu/urn-3:HUL.InstRepos:37705560>.

University of Washington. n.d. *A History of IT at the UW: Administrative Data Processing.* Seattle: UW-IT Historical Archive.

Urban, R. F. 1992. *Increasing Admitted Student Yield Using a Political Targeting Model and Discriminant Analysis: An Institutional Research-Admissions Partnership* (AIR Professional File No. 45). ERIC Document Reproduction Service No. ED350901/HE025859 Tallahassee, FL: Association for Institutional Research.

van der Klaauw, W. 2002. Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review.* 43(4): 1249–1287. Available at: <jstor.org/stable/826967>.

Weiler, W. C. 1980. A model for short-term institutional enrollment forecasting. *The Journal of Higher Education.* 51(3): 314–327. Available at: <doi.org/10.2307/1981012>.

WICHE. *See* Western Interstate Commission for Higher Education.

Western Interstate Commission for Higher Education. *About Knocking*. Boulder, CO: Author. Retrieved July 29, 2026 from: <wiche.edu/knocking/about/>.

About the Author



Tom Green

Tom Green, Ph.D., brings 40 years of higher education leadership success to his work. After serving as the chief enrollment officer at multiple universities where he increased enrollment, persistence, and net revenue at each, he served as a senior consultant for and later led AACRAO Consulting. He also served as the Associate Executive Director of the association, leading professional development for more than 11,000 members worldwide. Dr. Green has led hundreds of workshops on enrollment, financial aid optimization, and digital transformation, and has personally helped more than 100 universities address their enrollment, financial, and technology issues. He has taught at the undergraduate and graduate levels and is the author of numerous peer-reviewed articles and book chapters. Dr. Green holds a Ph.D. in Higher Education Leadership, Management and Policy from Seton Hall University.