GRESB

Advanced Ratings for Funds

JANUARY 2026

Parag Cameron-Rastogi
Director · Real Asset Analytics
GRESB

Charles van Thiel
Director · Real Estate Standard
GRESB

Chris Pyke
Chief Innovation Officer
GRESB

Executive Summary

This report presents proposed alternatives to the single, global, all-sector GRESB rating for funds. It explores how the usefulness of the rating system could be improved to provide better insights for participants and investors, and to address the challenges that have arisen from the evolution of participants' practices and increasing diversity of investor needs. A summary of this report is available on the **GRESB website**.

The current GRESB rating system assigns star ratings to entities based on their overall GRESB scores, using quantiles calculated from the current year's scores. This method ensures that a fixed percentage of entities receive each rating but can lead to significant fluctuations in star ratings from year to year. This report presents three alternatives:

- absolute-score-based thresholds (e.g., 0-20 for 1 star, 20-40 for 2 stars, etc.),
- thematic rating (e.g., Energy and GHG only),
- ratings within slices or groups defined by fund characteristics (e.g., region or sector).

The scores and ratings currently used in GRESB are a result of the evolution of consensus among stakeholders (managers and investors) about the relevance and importance of different themes (materiality). They are meant to be a **transparent**, **interpretable**, and **stable** system to assess the impact of investments, compare performance globally, and differentiate entities based on verifiable criteria and outcomes. *However, the system is also meant to be flexible to changes in the consensus, and such changes are the impetus behind this report.*

The proposals in this report represent a shift from a single, global rating system to a more nuanced and flexible framework that can better capture the complexities of real-world investment decision-making. This evolution from a single rating to multiple ratings based on themes, regions, sectors, and/or absolute score thresholds will provide more relevant and actionable insights for investors and managers, allowing them to make more informed decisions based on their specific priorities and contexts.

The impacts of each option on ratings over the whole GRESB Real Estate Assessment dataset are presented here to collect feedback from the wider industry. In Q2 2026, GRESB will incorporate this feedback to further refine the proposals presented here. Each participant will receive output ratings per fund in an accessible format. As a reminder, the survey is available **here**.

1 Introduction

GRESB Assessments evaluate entities (funds) across several dimensions of sustainability, including management practices, performance metrics, and stakeholder engagement. The results of an Assessment are presented in the form of scores, which provide a clear and concise summary of an entity's sustainability performance and risk profile compared to every other entity that undertook the assessment in the given year. The scores are then interpreted using a global rating system that divides entities (funds) by overall score (0-100) into five ordered groups of equal size (20% each) receiving 1-5 stars (1 is lowest, 5 is highest).

A shifting paradigm for ratings

The proposals in this report represent a shift from a single, global rating system to a more nuanced and flexible framework that can better capture the complexities of real-world investment decision-making. This shift is illustrated in Figure 1, which shows the evolution from a single rating to multiple ratings based on themes, regions, sectors, and absolute score thresholds. This new approach aims to provide more relevant and actionable insights for investors and managers, allowing them to make more informed decisions based on their specific priorities and contexts.

We begin this report with an overview of the challenges to the current system and the vision for improving it. We will then present an overview of the impact of the advanced alternative ratings that could result from this vision. We will conclude with a summary of findings and recommendations. While this report is focused on the Real Estate Assessment, the same approaches will be trialed with the Infrastructure Assessment in due course.

If you would like to respond to the findings contained in this report or know of colleagues that did not get a copy, please have them email **info@gresb.com**.

2 Challenges with the Current System

The current system was originally developed in 2015 for a smaller, more homogeneous set of participants. Since then, both the size and diversity of the entities participating in GRESB have increased significantly, leading to a wider range of priorities, styles, and portfolio compositions. A more flexible framework could better serve the needs of real-world investment decision-making. Providing a first proposal of such a framework is the purpose of this report. Additional context and background about these challenges is provided in Section 5.

The interaction of today's data with the background and original intention of the system creates the three challenges identified in this report:

- **Generalist Rating**: A single, all-encompassing rating system for all themes and subjects is no longer suitable for many current uses of GRESB outputs.
- **Global-Only Rating**: A global rating system that does not account for regional or sectoral differences of context and priorities creates unintended disparities in ratings outcomes that are not due to differences in sustainability performance.

- **Unclear score-ratings relationship**: A rating system that conveys only relative performance based on the given year's total score complicates the interpretation of how changes in scores line up with changes in ratings.

Before we consider how to address these challenges, it is useful to consider how they arise, their consequences for participants, and where and when they do work.

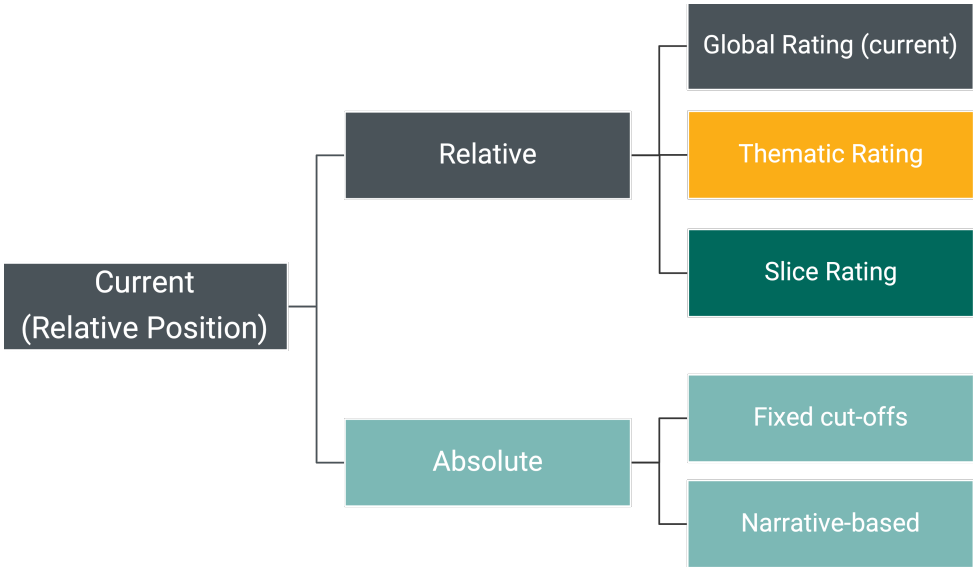

Figure 1: *The evolution of ratings from a single, global rating to multiple ratings based on the proposals in this report.*

2.1 Generalist Rating

While the headline outcomes of the Assessment are the overall score and rating derived from it, included within these are several sustainability themes, such as energy, carbon, water, human health, and physical climate risk. There are no specific restrictions or narratives attached to the ratings vis-à-vis themes and minimum achievement within them. For example:

- A 5-star entity could have entirely skipped Water or GHG reporting in 2025.
- A 3-star entity could have skipped the entire management component and risk management.
- A 1-star entity could have excellent performance in energy, carbon, water, and waste, but nothing else.

The scores for the individual indicators and themes are open and transparent in the Benchmark Reports. However, since the headline score is a sum of the indicator scores, there is an unavoidable loss of thematic specificity, when *only the overall score and rating are considered*. This could be a problem for investors and managers who have specific thematic priorities or mandates, but could, in principle, be easily rectified by interrogating indicator scores. Indeed, many investors already do this.

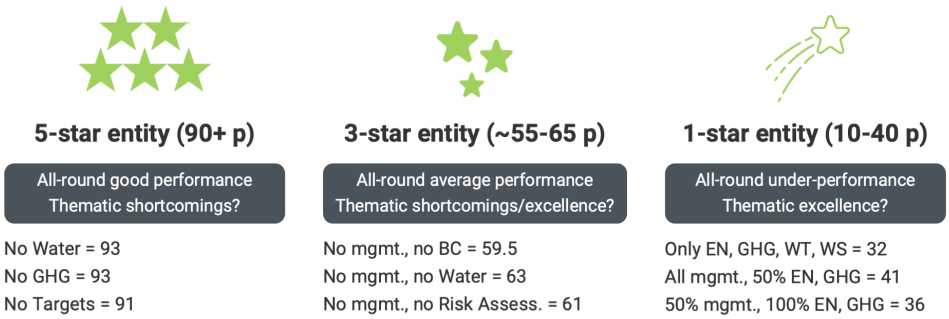

Figure 2: *There are many ways to reach the same overall score and rating, but with very different thematic profiles.*

2.2 Global-Only Rating

The current rating system is global and does not account for regional or sectoral differences in context, priorities, and challenges. For example, water scarcity may be a critical issue in some regions but less so in others. Similarly, the importance of certain sustainability themes may vary across sectors, such as the need for occupant engagement in the residential sector. Differences in metric relevance, data availability, and data quality across sectors can create artificial "ceilings" on the scores and ratings of entities operating in those contexts.

This one-size-fits-all approach can create accidental winners and losers. One demonstration of that is the percentage of entities from each sector that end up in each star rating, shown in Table 1. Each row of the table represents all the entities in a sector (adds up to 100%). In principle, each cell in each row should contain a value close to 20%, indicating that the chance of entities from each sector ending up in each star category is roughly the same. That is, there is no deliberate sector bias in the Assessment and so the percentage of entities from each sector falling in each star rating should be about 20%. In practice, we see that Office and Retail have a disproportionate number of top star ratings, while Industrial and Residential have disproportionately few. This suggests that the current global rating system may not be adequately accounting for sector-specific challenges and contexts, leading to disparities in ratings that are not solely based on sustainability performance.

Table 1: *Percent of entities in each sector (row) per star rating (column).*

	1	2	3	4	5
Industrial	29.1	22.9	19.6	14.5	13.9
Office	8.8	11.6	15.7	24.8	39.2
Others	18.0	21.5	22.6	22.0	15.8
Residential	27.1	23.0	20.3	16.7	12.9
Retail	22.7	16.3	13.8	15.9	31.2

2.3 Score-Rating Relationship

Scores are the primary translation of real-world outcomes into an abstract numerical system for comparison and evaluation. Ratings are a secondary interpretation of these scores, designed to simplify the communication of performance.

The thresholds, or edges, of each star rating are the scores corresponding to four equally spaced quantiles of the distribution of each year's scores–20th, 40th, 60th, 80th, with minimum and maximum scores to close the edges. This means that each year, only 20% of entities can achieve the top 5-star rating, regardless of the distribution of scores, and 20% must fall in each of the other star ratings down to the bottom 1-star rating. This system has been effective in providing a relative measure of performance, but it has also led to unintended consequences such as:

- Overall scores are stable year on year (Figure 3), but cut-offs for star ratings are unpredictable. This global stability is useful for high-level comparisons, but it is not helpful for entities near the cut-offs.
- The bands for star ratings 2 to 5 are compressing every year (Figure 3), making it harder to see differences between funds across these bands.

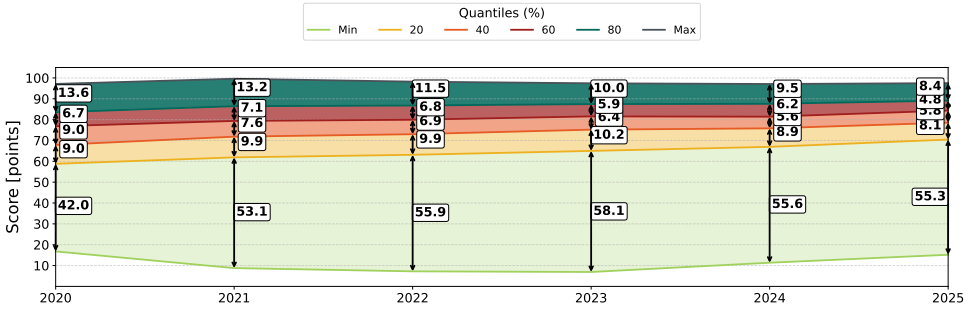

Figure 3: *The evolution of total score thresholds over time. The numbers in the white boxes are the width of each star rating band in points.*

3 Solution Exploration

The solution exploration is divided into three types of proposals. These proposals are not mutually exclusive and can be combined in various ways to create a more tailored rating system that meets the diverse needs of GRESB participants.

- **Thematic Ratings**: Ratings focused on specific themes or groups of themes to allow participants to demonstrate focused improvement and excellence, for example:
  - Energy and Carbon Rating
  - Health and Wellbeing Rating
  - Measured Performance Excluding Building Certifications
- **Sliced Ratings**: Consistent, comparable "slices" (subsets) of the population based on region, sector, or other relevant criteria to create more contextually relevant ratings, for example:
  - Industrial, USA
  - Residential, Europe
  - Office, Japan

- **Absolute Score Thresholds**: Narrative-based fixed cutoffs for scores to create more predictable and stable ratings that can evolve over time, for example:
  - Uniform width, difficult five star: [0, 25, 50, 75, 95, 100]
  - Width halves with each step up: [0, 51.5, 77.5, 90.5, 97, 100]
  - Custom thresholds based on stakeholder input

We now present the impacts of the various proposed solutions on the entire population of funds submitted to, and scored by, the Real Estate Assessment in 2024 and 2025. The change in star ratings for each proposed advanced rating type is calculated by comparing the star ratings assigned within each new method to the original GRESB star ratings assigned based on the global score distribution. The graphs presented here show the change in all entities across all slices–we will soon release fund-specific results in the GRESB portal for existing participants that show the impact of specific design choices. The histogram of ratings changes in each section includes data for 2024 and 2025 to show that the results are consistent across years. If you would like a quick primer on how to interpret Sankey/alluvial diagrams, please refer to **this article**.

3.1 Scope and Thematic Ratings

A thematic rating should provide a focused assessment of performance in a key area of sustainability, allowing participants to demonstrate excellence and improvement in a specific domain or group of related domains. Examples of thematic ratings include:

- **Energy and Carbon Rating**: A rating based solely on measured energy, water, and carbon data coverage and performance.
- **Everything but Building Certifications**: Excluding only the Building Certifications since feedback suggests that many participants consider this indicator to be redundant.
- **Performance Without Building Certifications**: Including all indicators in the “Performance” module only except Building Certifications.
- **Tenants, Communities, and Risk Rating**: Indicators related to tenant engagement, community impact, and risk management only.

The impact of these thematic ratings on the distribution of star ratings is shown in Figures 4 to 6. The x-axis in the bar chart in Figure 4 is the difference between the current rating and a thematic rating for any given entity. For example, a value of +2 means that the entity received a star rating two levels higher in the thematic rating compared to the original GRESB rating. The y-axis is the percentage of entities that experienced that change. The Sankey/alluvial diagrams in Figures 5 and 6 illustrate how entities transition between star ratings when moving from the original GRESB rating to the thematic ratings.

The extent of change from any thematic rating compared to the current, generalist rating depends on the level of differentiation among participants in the selected themes, i.e., themes that have not been dropped. When dropping an indicator appears to have little to no impact on the rating outcome, it is because most of the population scores about the same in an indicator and, thus, there is no meaningful differentiation between entities. For example, the Energy and Carbon rating shows a larger shift in star ratings compared to the others, indicating that

performance in these areas varies more significantly across entities. However, note that even in this case, the majority of entities (about 80%) see no change in their star rating, indicating that while there is some differentiation, many entities perform similarly in these themes.

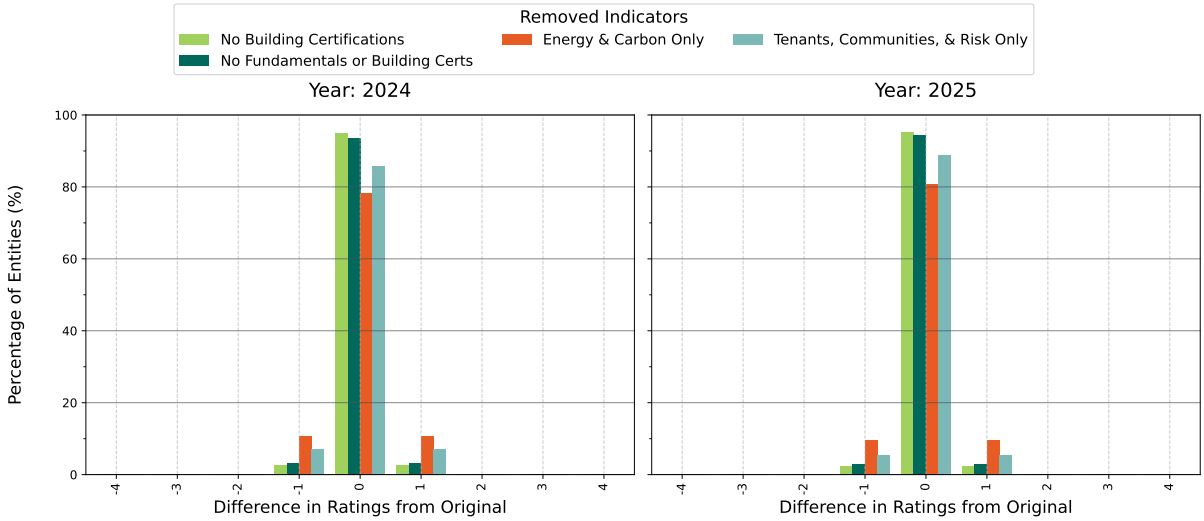

Figure 4: *Histogram of difference in ratings after removing selected indicators for thematic modules. None of the proposed themes create large differences in ratings for any individual entity, nor do they significantly alter the overall distribution of ratings.*

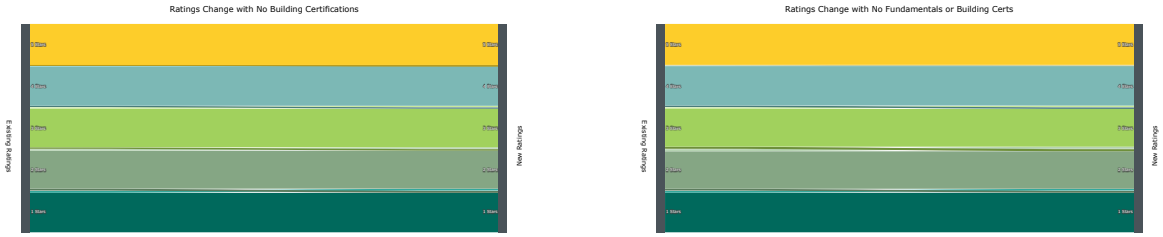

Figure 5: *Ratings change after removing selected indicators for thematic modules (1 of 2).*

3.2 Sliced Ratings

To calculate star ratings for each slice (e.g., USA–all sectors, USA–Office, Japan–all sectors, Listed vs. Non-listed), entities within each slice are ranked by their overall scores. Star ratings are then assigned based on the quantiles of the score distribution within that slice, consistent with the current global approach. This section explores the impact of creating ratings within specific slices of the population based on region, sector, and ownership type (listed vs non-listed), accounting for location-driven (regional, national, etc.), sectoral, and ownership-driven differences in sustainability priorities and challenges. Figures 7 to 9 illustrate how entities transition between star ratings when moving from the original GRESB rating to the sliced ratings based on country, sector, and ownership type (separately and in combinations).

Slicing entities by ownership type (listed vs non-listed) has the least impact on star ratings, with most entities retaining their original ratings. This suggests that ownership type may not be a significant differentiator in sustainability performance. Slicing by sector and location

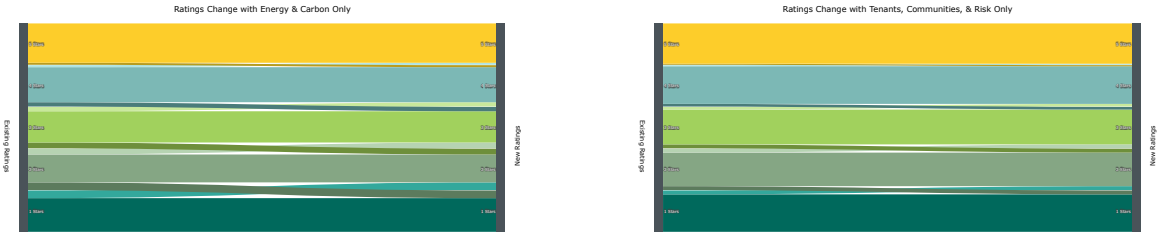

Figure 6: *Ratings change after removing selected indicators for thematic modules (2 of 2).*

shows a moderate impact, with a noticeable number of entities experiencing changes in their star ratings (both up and down). Finally, combining sector and location slicing results in the most significant changes in star ratings, indicating that these factors might together create a more contextually relevant assessment of sustainability performance. Note that each sankey diagram in Figures 8 and 9 contains a ”zero” star rating on the right side. This is because some slices contain fewer than six entities, making it impossible to assign all five star ratings within that slice. Entities that would have received a star rating in such slices, and have star ratings in the current, global system, are instead assigned a ”zero” star rating. This is the obvious downside of slicing the population too finely. Currently, this is mitigated in the peer group system (which uses location and building subtype to begin the enumeration of peer groups) by escalating to a higher-level on each dimension (e.g., country $\rightarrow$ subregion $\rightarrow$ region) until a sufficient number of entities are found to create a peer group. A similar approach could be taken here to ensure that each slice has enough entities to assign all five star ratings.

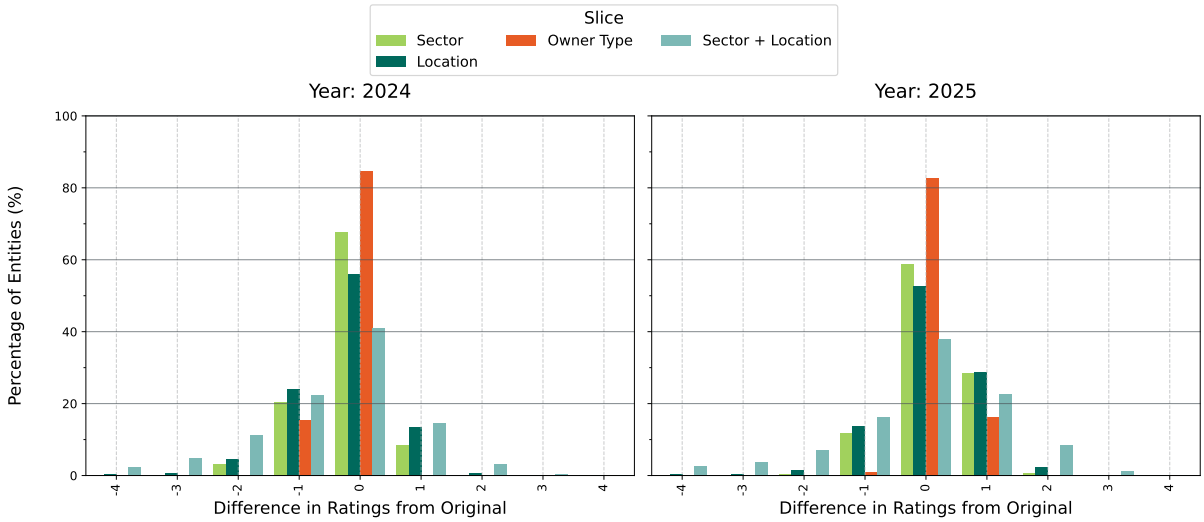

Figure 7: *Histogram of difference in ratings after slicing by sector, location, ownership type, and sector-location combinations.*

3.3 Absolute Score Thresholds

Fixed thresholds are, by definition, predictable and stable. This should mean that entities can plan for and interpret changes in ratings more easily. However, fixed thresholds do not adapt to

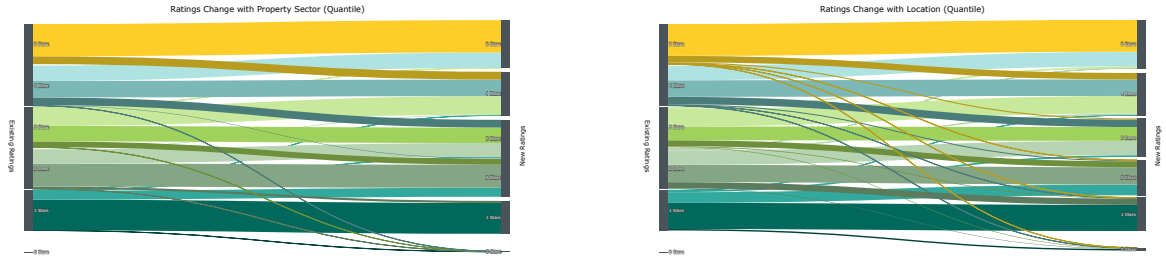

Figure 8: *Ratings change for slices of [left to right]: property sector and location (1 of 2).*

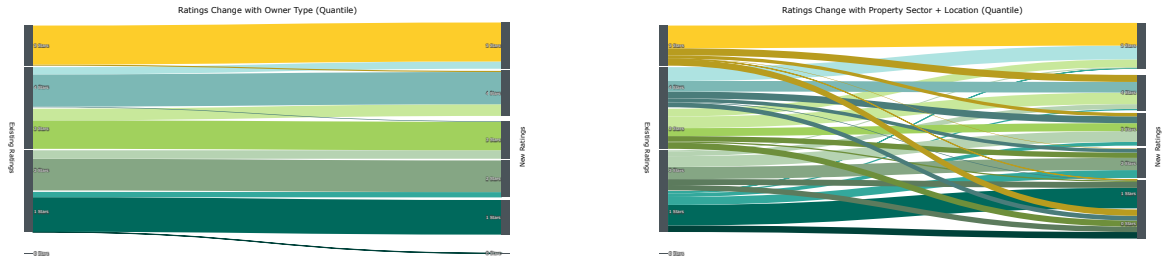

Figure 9: *Ratings change for slices of [left to right]: owner type and sector + location combined (2 of 2).*

changes in the overall score distribution, which could lead to unintended consequences if the population's performance improves significantly over time and the thresholds are not updated.

A thought experiment showing the impact of using fixed thresholds on the current data is shown in Figures 10 to 12 using the thresholds described at the beginning of this section (Section 3). Freed from the constraint of five equal-sized groups most entities today achieve four or five stars. A second set of fixed thresholds based on making five star disproportionately difficult to achieve (95-100 only) means few entities achieve it. Finally, successively halving the width of each band as you go up means that it is progressively more difficult to achieve each higher star rating. This still results in no one-star entities and a preponderance of two and three star entities.

These bands are illustrative only, and an implementation of fixed thresholds should ideally flow from narratives. For example, to reach a three-star rating, an entity must demonstrate high data coverage and moderate performance across three out of five material themes. Or to reach a five-star rating, an entity must demonstrate excellent performance in energy and carbon, water, and waste, along with high data coverage in all themes.

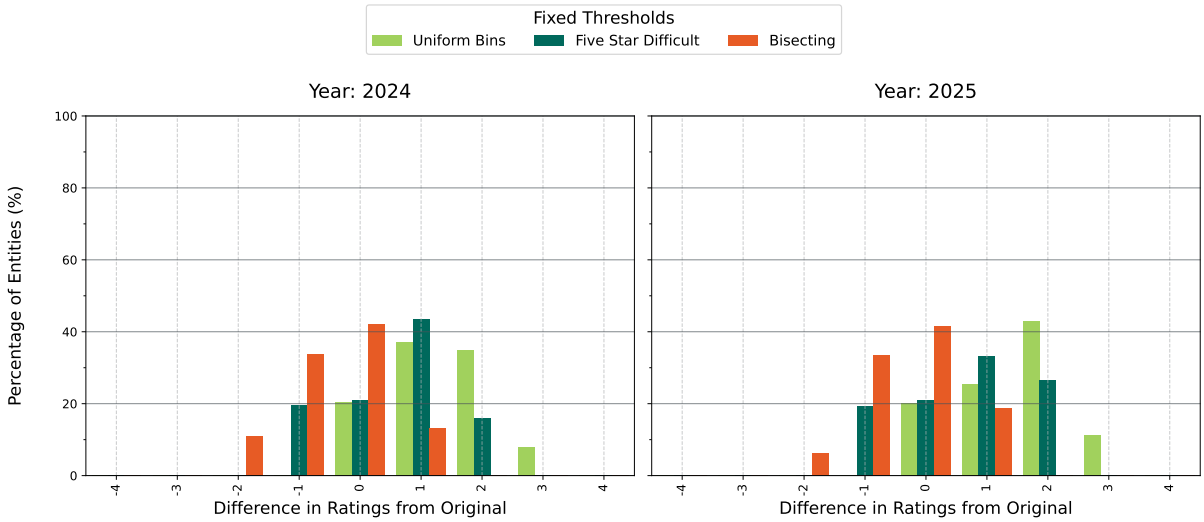

Figure 10: *Histogram of difference in ratings after applying fixed thresholds.*

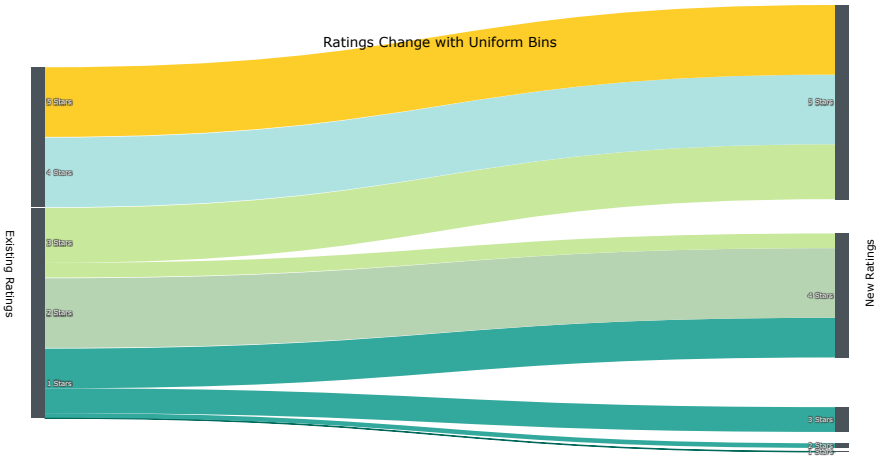

Figure 11: *Sankey diagram of ratings change for fixed thresholds (uniform bins).*

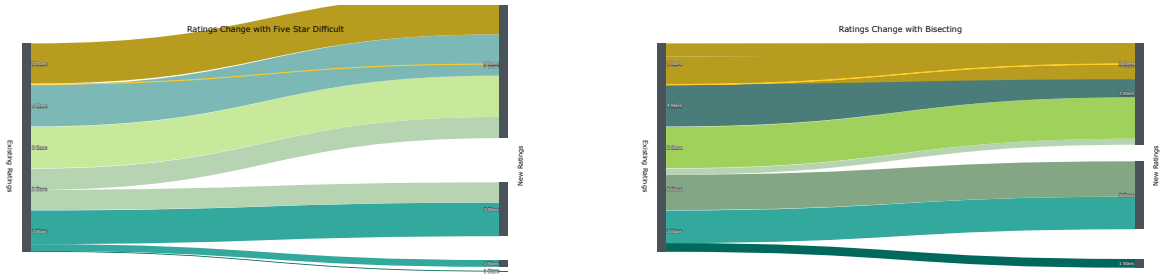

Figure 12: *Sankey diagrams of ratings change for fixed thresholds: five star difficult and bisecting.*

4 Conclusion and Next Steps

The proposals presented in this report demonstrate the potential for a more flexible and nuanced rating framework composed of multiple simultaneous ratings available to GRESB participants and investors. This exercise has demonstrated the use of novel approaches to rating that can better inform investment decision-making by providing more relevant, interpretable, and stable assessments of sustainability performance.

4.1 Roadmap for Implementation

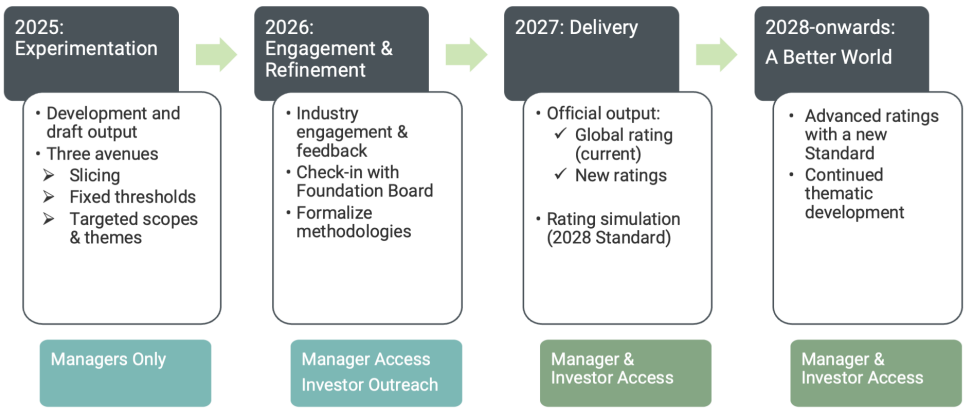

Figure 13: *Proposed roadmap for implementing multiple simultaneous ratings in GRESB.*

Figure 13 presents a proposed roadmap for implementing multiple simultaneous ratings in GRESB. This report is the result of the 2025 experimentation. Stakeholder engagement in 2026 will shape delivery and methods, with the new ratings available in beta alongside the 2026 results. The ratings will become co-official with the current, global rating in 2027 and continue to evolve alongside the evolution of the standard.

4.2 Call to Action

Your Feedback Is Valuable

We invite all GRESB participants and investors to provide feedback on the proposals presented in this report. Your insights and perspectives are crucial in shaping the future of GRESB ratings. Please fill in **this survey**.

5 Appendix: Background and Context

5.1 Score-Rating Relationship

Any system of scores and ratings is inherently a simplification of complex realities. The GRESB Assessment takes a metric calculable for an asset (e.g., energy use intensity in kWh/m^2), and converts it into a score for that asset for that indicator (e.g., EN1) from 0 to `max_score` using a "scoring function". The formulation of this function depends on the benchmark group to which an asset belongs, which is determined by the type of asset and its location. These translation functions are designed to improve fairness by uplifting scores for metrics in markets that have very low median scores, which is taken to be a reflection of market conditions or difficulties in a given year. If a market has high scores in general, the scoring function never penalizes participants, defaulting instead to the simplest one-to-one translation from metric to score. In other words, scores and metrics are strictly positively correlated. The scores for each asset for each indicator are then aggregated into scores for the entity using weighted averages, where the weights are a combination of an asset's area and Gross Asset Value. See the **GRESB Real Estate Assessment Reference Guide** and **Aggregation Handbook** for worked examples. The upshot is that while metrics are a reflection of real-world outcomes, especially for measured or calculated performance, their translation into scores is mediated by the relative performance of the benchmark group to which an asset belongs, meaning *the same scores in different benchmark groups may reflect different real world outcomes*.

Table 2: *Comparison of different thresholds against the current quantile-based system. Last year refers to using the breaks from the previous survey year (2023 or 2024) for the year mentioned (2024 or 2025).*

	1	2	3	4	5
Quantiles (Current)	67.4	76.5	82.5	88.1	100.0
Last Year (2024)	65.0	75.2	81.6	87.4	100.0
Last Year (2025)	66.9	75.8	81.4	87.6	100.0
Fixed Breaks	20.0	40.0	60.0	80.0	100.0

Table 2 compares three different methods of assigning star ratings to entities based on their overall GRESB scores. The current method uses quantiles calculated from the current year's scores to define the thresholds for each star rating, ensuring that a fixed percentage of entities receive each rating. Using last year's quantiles increases transparency but not stability, since the quantiles are still calculated from each year's scores. Fixed bins are stable and predictable, but do not intrinsically adjust to changes in the overall score distribution, i.e., achievement by participants.

5.2 Ratings and Perception

The choice of splitting entities into groups of equal sizes is a simple and effective way to create a rating system that is easy to understand and communicate. However, it also means that the thresholds for each star rating are determined solely by the distribution of scores in a given year, which can lead to significant fluctuations in star ratings from year to year based on one's arbitrary position relative to the benchmark or peer groups. This makes it harder to say that a "5-star" entity consistently conveys something specific each year, even if a 5-star entity always has a higher overall score than a 4-star entity, by definition. Fixed breaks

corresponding to narratives are a potential solution to this issue. Though we have not developed detailed narratives for the fixed breaks proposed here, the inclusion of this option is meant to show the potential impacts from such a choice.

The perception of a "5-star" entity may vary significantly among different stakeholders, but it will always be regarded as being better than a "4-star" entity. If the threshold between 4 and 5 stars is, for example, 92 points, then an entity scoring 91 points would be considered a "4-star" entity, while one scoring 93 points would be a "5-star" entity. While both scores would have been achieved through a combination of actions, such as improving energy efficiency, reducing emissions, enhancing tenant engagement, or implementing innovative sustainability practices, the performance medians of the benchmark groups to which the assets of an entity belong govern the translation of metrics to scores. This means that, for entities with similar scores, the difference in star rating may not reflect a significant difference in absolute performance or impact, leading to potential confusion and misinterpretation of the results. Though the 5-star entity has a higher score, it may not necessarily have a proportionately greater impact or performance improvement compared to the 4-star entity. This perception "cliff" creates a challenge for stakeholders trying to understand the true performance of entities.

5.3 Are 5 Stars Optimal?

There is no rule for why a group of entities must be divided into exactly five levels. **The choice of five levels is somewhat arbitrary and may not reflect the optimal distribution of entities** to maximize differentiation while minimizing complexity. This choice is similar to the question of how unlabelled data ought to be clustered or categorized. Unless there is a strong real-world reason to justify a specific number of levels, it is worth exploring alternative approaches to grouping entities.

We present the results of a clustering exercise based on the structure of the existing Real Estate Assessment in Figure 14, which shows the elbow method for determining the optimal number of clusters. All metrics in that figure are normalized such that the value at clusters=2 is 1, i.e., we assess the value of adding each new cluster (category) against just having two categories or stars. The three panels in Figure 14 correspond to three different metrics for assessing the quality of clustering. [Left] No clear elbow is visible, though there is a slight change of slope between 4-6 clusters, suggesting that range of clusters (divisions) is a reasonable compromise between the complexity of many clusters and the correct assignment of points to clusters. [Center] The average cosine similarity for each cluster/rating always converges to 0.9 (entities are very similar) at the higher end of GRESB Scores (i.e., total scores) for any number of total clusters. However, the average cosine similarity for lower GRESB scores tends to be much lower on average, i.e., entities are different, when the number of clusters is increased. [Right] The average $\pm$ standard deviation of maximum distance to cluster median or centroid across all points in a cluster. The centroids are chosen through k-means and the medians correspond to the entity with the median score in each cluster.

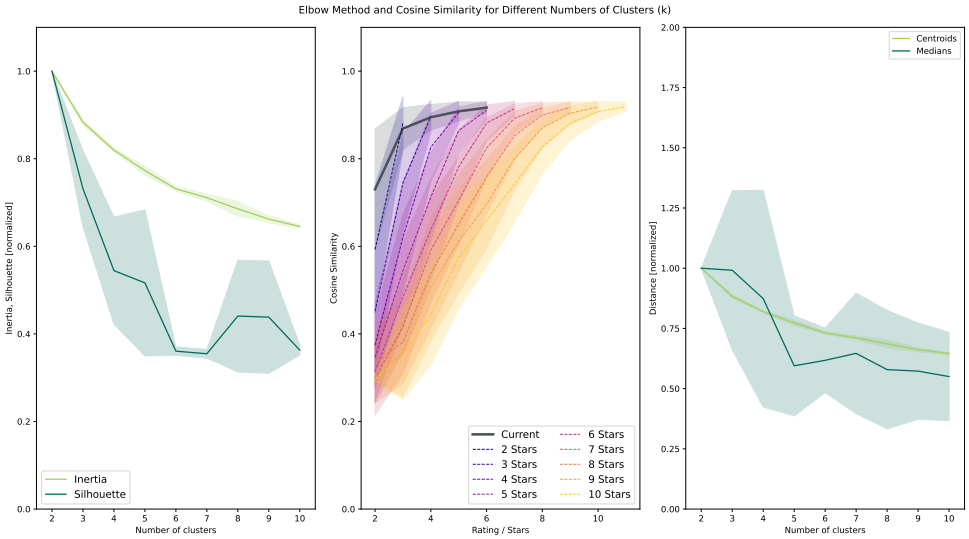

Figure 14: *Elbow method for determining the optimal number of clusters. All metrics are normalized such that the value at clusters=2 is 1.*

Table 3: *Fixed score breaks for different numbers of fixed star ratings.*

	1	2	3	4	5	6	7	8	9	10
2 Stars	50.0	100.0								
3 Stars	33.3	66.7	100.0							
4 Stars	25.0	50.0	75.0	100.0						
5 Stars	20.0	40.0	60.0	80.0	100.0					
6 Stars	16.7	33.3	50.0	66.7	83.3	100.0				
7 Stars	14.3	28.6	42.9	57.1	71.4	85.7	100.0			
8 Stars	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0		
9 Stars	11.1	22.2	33.3	44.4	55.6	66.7	77.8	88.9	100.0	
10 Stars	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	100.0


GRESB