LEED-NC 2009 SILVER TO GOLD CERTIFIED PROJECTS IN THE US IN 2012–2017: AN APPROPRIATE STATISTICAL ANALYSIS

Svetlana Pushkar[1] and Oleg Verbitsky [1]

ABSTRACT

This study aims to evaluate the Silver-to-Gold LEED-NC 2009 (Leadership in Energy and Environmental Design for New Construction and Major Renovations) cross-certification performance and categorize the cross-certification performance in eight US states in 2012–2017. The following three statistical analyses were used: (a) pooling LEED projects within a single state and single year in a single-state-year group with the subsequent use of a replication method, (b) pooling the medians of the LEED projects in each state from all years in a state-and-total-years group, and (c) pooling the LEED projects from all states and years in a total states-and-years group. The Silver-to-Gold cross-certification performance has a low propelling effect. Considering the Silver-to-Gold category cross-certification performances, the Energy and Atmosphere (EA) category has a high propelling effect, the Sustainable Sites (SS) and Environmental Quality (EQ) categories have moderate propelling effects, the Water Efficiency (WE), Materials and Resources (MR), and Innovation in Design (ID) categories have low propelling effects. Six of the eight states used an EA-high emphasized strategy, and two of the eight states used a SS/EA/WE/EQ/ID-moderate emphasized strategy. The single-state-year group and state-and-total-years group analyses are more robust than the pooling LEED projects using the total state-and-year group analysis.

KEYWORDS

LEED-NCv3, Silver level, Gold level, Three-valued logic, Replication method.

INTRODUCTION

The triple bottom line of sustainability includes environmental sustainability (saving and regenerating natural resources), social sustainability (ensuring the well-being of occupants and workers), and economic sustainability (decreasing the life cycle cost of building performance) (Illankoon et al. 2017a). However, the current green rating systems mostly evaluate the environmental sustainability and dismiss social sustainability and economic sustainability (Illankoon et al. 2017b). In this respect, the Leadership in Energy and Environmental Design (LEED), which was developed by the US Green Building Council (USGBC) for the US green building

1. Department of Civil Engineering, Ariel University, Israel, Email: svetlanap@ariel.ac.il, olegv@ariel.ac.il

practice, is not an exception. As was reported by Illankoon et al. (2017b), 74.59%, 18.03%, and 0.82% of all New Construction and Major Renovations (LEED-NC) credits are devoted to environmental sustainability, social sustainability, and economic sustainability, respectively.

LEED contains four levels of certification: Certified, Silver, Gold, and Platinum. The minimum numbers of points required for the Certified, Silver, Gold, and Platinum classification levels are 40, 50, 60 and 80, respectively. Each level includes six basic categories: Sustainable Sites (SS), Water Efficiency (WE), Energy and Atmosphere (EA), Materials and Resources (MR), Indoor Environmental Quality (EQ), and Innovation in Design (ID). There is one additional category, Regional Priority [RP], which contains four bonus points that can be awarded in any of the five main categories (i.e., SS, WE, EA, MR, and EQ). Points are accumulated in all categories, and each category has a maximum number of points: SSmax = 26 pts, WEmax = 10 pts, EAmax = 35 pts, MRmax = 14 pts, EQmax = 15 pts, IDmax = 6 pts, and RPmax = 4 pts.

LEED certification has been adapted for different building types, including Commercial Interior (LEED-CI), Core and Shell (LEED-C&S), and Existing Buildings (LEED-EB), among which LEED-NC is one of the most popular rating schemes (Cheng and Ma 2015). With the goal of continuous improvement, LEED-NC is constantly undergoing modifications (i.e., LEED-NCv2.0 [2000], LEED-NCv2.1 [2002], LEED-NCv2.2 [2005], LEED-NCv3 [2009], and LEED-NCv4 [2013]). One important practice that has enabled these improvements is a survey analysis of the LEED-certified projects (Wu et al. 2017).

To make further improvements to the LEED system, two types of research have been undertaken: (i) LEED rating system analyses (Pushkar and Shaviv 2016; Illankoon et al. 2017b) and (ii) LEED-certified project analyses [empirical studies] (Fuerst 2009; Ma and Cheng 2016; Wu et al. 2016; Wu et al. 2017; Wu et al. 2018; Pushkar and Verbitsky 2018). The goal of empirical LEED studies is to reveal trends (practitioner preferences) in achieving LEED categories, which discloses high- and low-performing credits and categories. This can help to improve unpopular, low-performing credits in further LEED versions (Wu et al. 2017).

Among the empirical studies that have analyzed the LEED system, some of the more enlightening are those presented by Fuerst (2009), Ma and Cheng (2016), Wu et al. (2016), Wu et al. (2017), and Pushkar and Verbitsky (2018). These studies considered projects certified with different LEED subsystems and versions. The analyses were performed on different levels of evaluation: cross-certification performance (an analysis of the differences in the total number of achieved points in all categories for any two adjacent certification levels; for example, the Silver to Gold [Ma and Cheng 2016]) and category cross-certification performance (an analysis of the achieved points for each individual category; for example, SS for any two adjacent certification levels, such as Silver to Gold [Wu et al. 2017]).

Fuerst (2009) evaluated the LEED certification level, studying over 2,000 LEED-NC, LEED-CI, LEED-CS, and LEED-EB certified projects registered as of March 2009. The author reported the relatively low point values observed for each certification level; the projects were clustered slightly above the lower threshold of each certification level: 40, 50, 60 and 80 in the Certified, Silver, Gold, and Platinum certifications, respectively. However, in this study, all types of buildings were evaluated together in one pooled sample, and no rigorous statistical evaluations were presented.

Ma and Cheng (2016) also evaluated the overall LEED certification level, studying only LEED-NCv3 certified projects (1,000 pooled samples). The authors confirmed the low point values previously reported by Fuerst (2009) and performed a category cross-certification analysis.

In this respect, they reported that the EA category was the most propelling category that moved a project from a lower to higher certification level. However, no significance test was performed.

Wu et al. (2016) applied significance tests to a study of LEED-NC certified projects (5,340 pooled samples). The results were similar to those reported by Ma and Cheng (2016), with low scores for Certified, Silver, Gold, and Platinum certifications and EA as a propelling category. However, Wu et al. (2016) analyzed projects certified under LEED-NCv2.2. There are three important differences between LEED-NCv2.2 and LEED-NC 2009 (v3): (i) the total number of points in LEED-NCv2.2 was 69 pts and 110 pts in LEED-NC 2009; (ii) in LEED-NCv2.2, all credits were given the same weighting of 1 pt for each credit, whereas different weights were given to different credits in LEED-NC 2009; and (iii) the RP category was launched in LEED-NC 2009 but was not included in LEED-NCv2.2 (LEED-NC 2005; LEED-NC 2009).

Wu et al. (2017) investigated the LEED-NC 2009 category cross-certification performance between the Silver and Gold certifications levels. The authors collected LEED-NC 2009 certified projects from all over the world over five years and pooled them into two unpaired groups (i.e., the LEED Silver level group and LEED Gold level group) in a single cluster. The sample sizes for the Silver and Gold LEED-certified project groups were n_1 = 1310 and n_2 = 1201, respectively (p. 375, Table 9). The authors concluded that no fewer than three categories (EA, SS, and WE) were propelling categories that moved projects from the Silver to Gold certification (p. 375, Table 7). Consequently, if the LEED Silver and Gold level projects are pooled into a single cluster, then the influences of other green building policies, such as Energy Standard for Buildings Except Low-Rise Residential Buildings (ASHRAE 90.1), on the LEED performance of a project are ignored (Pushkar and Verbitsky 2018), it can lead to misinterpretation of the results (Pushkar 2018).

Pushkar and Verbitsky (2018) analyzed the LEED-NC 2009 category cross-certification performance between the Silver and Gold certification levels when data were collected separately for ten different US states (CA, FL, GA, IL, MA, NY, OH, TX, VA, and WA) in 2016. The ten US states were selected to develop the following two criterions: (i) LEED Silver projects and LEED Gold projects should belong to the same state, and (ii) the minimum number of LEED projects for each of the Silver and Gold levels was n = 8. Thus, each state was defined as a single cluster. In this context, it was suggested that "any two LEED projects in the same state share more similar green building policy conditions compared to any two LEED projects from different states" (Pushkar 2018). As a result, three different strategies for certified projects in transition from Silver to Gold were revealed: (a) energy-emphasizing (e.g., California), (b) nonenergy-emphasizing (e.g., New York), and (c) integrated (e.g., Georgia) strategies (Pushkar and Verbitsky 2018). It was shown that the energy-emphasized strategy occurs when the EA category is responsible for the transition between adjacent (from low to high) certification levels of LEED. The nonenergy-emphasized strategy occurs when the non-EA categories are responsible for the transition, such as SS or EQ. In the integrated strategy, three categories, one of which is the EA category, are responsible for the transition between adjacent certification levels (Pushkar and Verbitsky 2018).

However, the lack of information on the standardized effect size is a significant drawback to the statistical conclusions obtained in the studies presented by Wu et al. (2017) and Pushkar and Verbitsky (2018). Information obtained from P-values can be used in conjunction with the effect size (i.e., a combination of the statistical significance and practical significance) to produce a correct statistical conclusion (especially when the sample sizes of the compared groups are substantially different) (Nakagawa and Cuthill 2007). It should be noted that a small *P*-value, such as that of Wu et al. (2017) due to their large sample size, can be associated with a negligible or small effect size (Sullivan and Feinn 2012). Based on a study by Pushkar and Verbitsky (2019, accepted), a large effect size can correspond to the practical significance when the propelling categories for the transition from Silver to Gold certification are evaluated. Additionally, a replication method must be used to improve the robustness of the statistical conclusion (Crandall and Sherman 2016).

In other words, to verify the reliability of the results that were obtained in 10 US states in 2016, a replication analysis of the same 10 US states should be performed by using a retrospective longitudinal design with an interval in a given year (e.g., the first year that LEED-NC 2009 was used) to the current year. Using the wrong statistical analysis may lead to misunderstanding the main tendencies of the cross-certification performance and category cross-certification.

The aim of the present study was to identify trends (practitioner preferences) in (i) cross-certification and (ii) category cross-certification performance for LEED-NCv3 practices in the US during 2012–2017. Each of the two performance assessments was descriptively and statistically evaluated by applying three types of statistical analysis: (i) pooling LEED projects from a single state (i.e., spatial) and a single year (i.e., temporal) in a single state-year group using a replication method; (ii) pooling the medians of the LEED projects in each state over the years in a state-and-total-years group, and (iii) pooling the LEED projects from all states and years in a total states-and-years group.

METHODS

Data collection and sample size limitations

First, we collected LEED-NC 2009 Silver and Gold projects in the same states that were analyzed in the study of Pushkar and Verbitsky (2018) (CA, FL, GA, IL, MA, NY, OH, TX, VA, and WA) from 2009–2017. Then, we eliminated from the analysis (i) the projects that were certified in 2009–2011 and (ii) the projects from two states (GA and NY). This is because of the requirements of the Wilcoxon–Mann–Whitney (WMW) test (i.e., the test for two unpaired groups) that needs to be used in this analysis due to the ordinal scale of the LEED data being represented. In particular, the minimum sample size for two groups for the WMW test should be $n_1 = n_2 = 4$ (Fay and Proschan 2010). This sample size requirement was not met in the groups of projects described above (for example, see the data in 2011 and the GA and NY data, Table 1). As a result, in this study, LEED-NCv3 Silver and Gold projects certified in eight states, CA, FL, GA, IL, MA, NY, OH, TX, VA, and WA, in 2012–2017 were considered (Table 1).

The USGBC scorecards of the considered projects were downloaded from the USGBC LEED for New Construction Projects Directory. From the scorecards, information about the points awarded in the SS, WE, EA, MR, EQ, ID, and RP categories was obtained, and the bonus RP points were assigned to one of five basic categories: SS, WE, EA, MR, and EQ.

Data analysis

Delta measurements were used to assess the cross-certification performance. The delta value was calculated as the difference between the project certification performance (the summary performance of all six categories, SS, WE, EA, MR, EQ, and ID) and the bottom boundary of certification (for Silver—50 points; for Gold—60 points) (briefly: —Silver delta and —Gold delta). The Silver delta and Gold delta were compared.

TABLE 1. Number of LEED-NCv3 Silver and Gold projects in ten states in 2012–2017 (i.e., new projects certified each year).

n	State	Level	2011*	2012	2013	2014	2015	2016	2017	Total
1.	California (CA)	Silver	4	21	28	37	50	40	41	217
		Gold	3	14	27	36	51	49	42	219
2.	Florida (FL)	Silver	5	19	13	28	20	31	14	125
		Gold	3	20	24	13	11	11	13	92
3.	Georgia** (GA)	Silver	2	4	8	8	11	11	4	—
		Gold	3	3	9	6	5	7	4	—
4.	Illinois (IL)	Silver	1	6	19	11	20	11	15	82
		Gold	1	9	13	12	20	14	6	74
5.	Massachusetts (MA)	Silver	0	5	7	5	10	15	15	57
		Gold	0	7	7	15	14	12	18	73
6.	New York** (NY)	Silver	0	3	5	10	12	17	13	—
		Gold	3	4	8	8	15	9	6	—
7.	Ohio (OH)	Silver	3	14	12	17	13	20	23	99
		Gold	2	6	7	10	7	7	7	44
8.	Texas (TX)	Silver	2	13	24	35	28	27	9	136
		Gold	2	4	11	13	14	8	27	77
9.	Virginia (VA)	Silver	1	4	10	16	23	34	23	110
		Gold	0	7	12	6	5	11	11	52
10.	Washington (WA)	Silver	0	8	13	23	19	10	14	87
		Gold	1	6	8	9	14	11	6	54
Total		Silver								913
		Gold								685

* These US states were not included in the statistical analysis in 2011 because the number of LEED projects (i.e., sample size) was $n < 4$ in one of the levels. ** These US states were not included in the statistical analysis because the number of LEED projects (i.e., sample size) was $n < 4$ in one of the years.

The category performance was considered when assessing the category cross-certification performance. For each of the six categories (SS, WE, EA, MR, EQ, and ID), the Silver and Gold performances were compared.

Design of the study

A cross-sectional design was previously used to evaluate the performances of the LEED-NCv3 Silver and Gold certifications in 10 US states in 2016 (Pushkar and Verbitsky 2018). In the present study, a retrospective longitudinal design is used to evaluate the performances of the LEED-NCv3 Silver and Gold certifications in the same eight of the ten US states in 2012–2017. To clearly describe the design of the study, the relevant statistical terminology is defined below.

According to Picquelle and Mier (2011, p. 2), A “primary sampling unit” is defined as an element within a “sampling frame” that is sampled and statistically independent of other sampling units within the frame. The sampling frame is defined as the collection of all the elements (primary sampling units) accessible for sampling in the population of interest.

Using this terminology and taking into account the spatial [i.e., state(s)] and temporal [i.e., year(s)] properties of the LEED projects, the cross-certification and category cross-certification performances were evaluated using three types of statistical analysis in conjunction with the sampling units and sampling frame:

1. If the LEED projects pooled from a single state and single year are the primary sampling units, then this collection of LEED projects is a sampling frame.
2. If the medians of the LEED projects pooled from a single state and all years are the primary sampling units, then this collection of the medians of LEED projects is a sampling frame.
3. If the LEED projects pooled from all states and years are the primary sampling units, then this collection of LEED projects is a sampling frame.

In the first type of statistical analysis, each of the eight states was studied separately each year for six years. In the second type of statistical analysis, the eight medians (one median for each of the eight states) in the six years were studied together. It should be noted that this type of statistical analysis decreases the spread of the data with outliers (Bickel 2003). In the third type of statistical analysis, all eight states and all six years were studied together.

Statistical analysis

The LEED data are presented on ordinal scales. Accordingly, two nonparametric tests, Cliff’s δ (Cliff 1993) and the WMW test (Mann and Whitney 1947), are used to compare the two unpaired groups. The data are presented as the median ± interquartile range (IQR, 25th–75th percentiles).

Cliff’s δ. Cliff’s δ is used to measure the substantive significance (effect size) between two unpaired groups. Cliff’s δ (Cliff 1993, p. 495) is expressed as:

$$\delta = \frac{\#\left(x_1 > x_2\right) - \#\left(x_1 < x_2\right)}{\left(n_1 n_2\right)}$$

where x_1 and x_2 represent scores within group 1 and group 2, respectively; n_1 and n_2 represent the sizes of the sample groups (group 1 and group 2, respectively); and # indicates the frequency.

Cliff’s δ ranges between –1 and +1; positive (+) values indicate that group 1 is larger than group 2, 0 indicates equality or overlap, and negative (–) values indicate that group 2 is larger than group 1 (Cliff 1993).

The effect size is considered (i) negligible when $|\delta| < 0.147$, (ii) small when $0.147 \le |\delta| < 0.33$, (iii) medium when $0.33 \le |\delta| < 0.474$, and (iv) large when $|\delta| \ge 0.474$ (Zhang et al. 2016).

According to Cohen (1992, p. 156), "a medium effect is visible to the naked eye of a careful observer. A small effect is noticeably smaller than the medium effect but is not trivial. A large effect is the same distance above the medium as small is below it." It should be noted that the effect size is not an "iron-clad criterion" (Durlak 2009) but rather is only a general rule of thumb that might be followed in the absence of knowledge of the area (Volker 2006).

WMW test. The WMW test is used to determine the significant difference (*P*-value) between two unpaired groups. WMW tests could be applied in two forms: the approximate form and the extract form (Fagerland and Sandvik 2009). When the sample size is $n_1 = n_2 \geq 9$, then an approximate WMW test is used (Marx et al. 2016, p. 56). Mann and Whitney (1947, p. 50) noted that "the distribution is almost normal" when the sample size is $n_1 = n_2 = 8$. When the sample size for one of the two groups is $4 \leq n_1 \geq 8$, an exact WMW test is used (Fay and Proschan 2010, p. 19; Marx et al. 2016). In both tests, a two-tailed *P*-value is applied.

Neo-Fisherian significance assessments. Historically, in research but not in industrial quality control, studies using statistical analysis were based on the following paradigm: fixing α, i.e., the level of significance, and dichotomizing the scale of the *P*-values, i.e., $P \leq \alpha$ or $P > \alpha$, determining "as a sort of mechanical Occam's razor" (Beninger et al. 2012) whether there was a 'significant or 'nonsignificant' difference, e.g., between two groups. This paradigm is a hybrid of the Paleo–Fisherian and Neyman–Pearsonian paradigms or null hypothesis significance testing (NHST) (Hurlbert and Lombardi 2009). In the current study, neo-Fisherian significance assessments (NFSAs) were performed (Hurlbert and Lombardi 2009). NFSAs include calculations of both the extract *P*-value and effect size, but there is no need to specify α. A high *P*-value leads only to suspended judgments. The interpolation of the extract *P*-value based on a three-valued logic is as follows: "it seems to be positive" (i.e., there seems to be a difference between group 1 and group 2), "it seems to be negative" (i.e., there does not seem to be a difference between group 1 and group 2), and "judgment is suspended" regarding the difference between group 1 and group 2 (Hurlbert and Lombardi 2009; 2012).

The use of the NFSA paradigm is based on three approaches that are considered in parallel: (i) the results of the inferential statistics, (2) an analysis of the literature, and (3) the author's experience (Hurlbert and Lombardi 2009). According to Pushkar and Verbitsky (2019, accepted), the difference between the Silver level and Gold level "seems to be positive" when there are a large size effect, and a low *P*-value, "seems to be negative" when there is a negligible or small effect size with any *P*-value. Judgment is suspended when there is a medium effect size and an intermediate or low *P*-value. Further observations are required to increase the accuracy of the NFSA conclusions (Pushkar 2018).

Replication method. The need to use a replication method was vividly described by Meehl, (Meehl, 1990, p.111): "Any working scientist is more impressed with 2 replications in each of 6 highly dissimilar experimental contexts than he is with 12 replications of the same experiment." In the present study, for each state and for each of the six years, observations were defined as six replications (i.e., six replications in a single state). Based on NFSAs, six replications for each state have individual sets that present the following statistical conclusions: it seems to be positive (briefly: "positive"), judgment is suspended (briefly: "suspended"), and it seems to be negative (briefly: "negative") differences. To determine the real contribution of delta to the cross-certification performance as well as each category in the category cross-certification performance, the sets from the positive, suspended, and negative NFSA decisions are accounted for and converted to three levels of performance: "high," "moderate," and "low."

RESULTS AND DISCUSSION

Cross-certification performance

(i) Pooling LEED projects in a single state-year group and replication results

Descriptive statistics (median ± interquartile range [IQR, 25th–75th percentiles]) for the Silver and Gold projects were obtained, and an inferential statistical analysis (Cliff's δ and P-value) for comparing the Silver-to-Gold cross-certification performances was performed using Silver "delta" and Gold "delta" measurements (difference between median Silver projects and the bottom level for the Silver level [50 Pts] and between median Gold projects and bottom level for Gold level [60 Pts]). These analyses were performed separately for each of the eight US states in each of six years (2012–2017) (Appendix A, Table A1).

The replication results of the Silver-to-Gold cross-certification performances for each of eight US states for a six-year period (2012–2017) are presented below (Table 2). The replication results present the total number of "positive" (where the difference between the Silver delta and Gold delta seems to be positive), "suspended" (where the judgment about the difference between the Silver delta and Gold delta is suspended), and "negative" (where the difference between the Silver delta and Gold delta seems to be negative) decisions in the Silver-to-Gold cross-certification derived from the results in Appendix A (Table A1).

As shown in Table 2, an analysis of the replication results for 2012–2017 reveals that mostly negative decisions were obtained. This means that there is no difference between the Silver and Gold deltas. Similar results were reported by other researchers (Fuerst 2009; Wu et al. 2016; Wu et al. 2017; Pushkar and Verbitsky 2018). It may be suggested that building practitioners prefer to pay for a certification level rather than for proper green building performance, as was originally suggested by Sandoval and Prakash (2016).

(ii) Pooling medians of LEED projects in median state-and-total-years group

The results of the descriptive statistics (median ± interquartile range [IQR, 25th–75th percentiles]) for the Silver and Gold projects and an inferential statistical analysis (Cliff's δ and P-value) for comparing the Silver-to-Gold cross-certification performances for the eight medians (one median for each of the eight US states) are presented in Table 3. It may be concluded that judgment about the difference between the Silver delta and Gold delta should be suspended (Table 3).

TABLE 2. LEED-NCv3 Silver-to-Gold cross-certification performance: replication results for the single state-year group statistical analysis.

NFSA	CA	FL	IL	MA	OH	TX	VA	WA
Positive	0	0	0	1[a]	1[a]	1[b]	1[a]	0
Suspended	0	3	1	0	0	1[b]	2[b]	1[b]
Negative	6	3	5	5	5	4	3	5
Performance	low	low	low	low	low	low	low	low

Note: [a]Silver delta was greater than Gold delta; [b]Gold delta was greater than Silver delta

TABLE 3. LEED-NCv3 Silver-to-Gold cross-certification performance: median state-and-total-years group analysis in eight US states in 2012–2017[1].

Sample size		Median±IQR		Inferential statistics			
Silver	Gold	Silver delta	Gold delta	Cliff's δ	Effect size	*P*-value	NFSA
8	8	2.0 ± 0.0	3.0 ± 0.8	–0.500	large	0.093	suspended

Note: [1] The analysis was performed using "delta" measurements (difference between the median of the raw data of the Silver projects and the bottom level for the Silver level [50 Pts] and between the median of the raw data of the Gold projects and the bottom level for the Gold level [60 Pts]).

(iii) Pooling LEED projects in total states-and-years group

The results of the descriptive statistics (median ± interquartile range [IQR, 25th–75th percentiles]) for the Silver and Gold projects and an inferential statistical analysis (Cliff's δ and *P*-value) for comparing the Silver-to-Gold cross-certification performances for the eight US states and six years pooled together are presented in Table 4. The difference between the Silver and Gold deltas seems to be negative because there was a negligible effect size despite a very small *P*-value (Table 4). Table 4 illustrates the mistakenness of the application of the NHST paradigm when there is a large sample size. Hurlbert and Lombardi (2009) allegorically noted the following: "it lurks quietly in the darkness, waiting for researchers to pass by who are too focused on obtaining adequate sample sizes. If sample sizes are too large, one may be 'in danger' of getting very low *p*-values and establishing the sign and magnitude of even small effects with too much confidence." Thus, there was no difference in the certification performance when the Silver deltas and Gold deltas were compared using the total states-and-years group analysis. These results are consistent with the results of a single-state-year group and replication results analysis (Tables 2).

Category cross-certification performance

(i) Pooling LEED projects in single-state-year group and replication results

Descriptive statistics (median ± interquartile range [IQR, 25th–75th percentiles]) for the Silver and Gold projects were obtained and an inferential statistical analysis (Cliff's δ and *P*-value) was performed separately for each of eight US states in each of six years (2012–2017) to compare the Silver-to-Gold category cross-certification performances (Appendix A, Table A2-A9).

TABLE 4. LEED-NCv3 Silver-to-Gold cross-certification performance: total states-and-years group analysis (i.e., eight US states and six years pooled)[1].

Sample size		Median±IQR		Inferential statistics			
Silver	Gold	Silver delta	Gold delta	Cliff's δ	Effect size	*P*-value	NFSA
913	685	2.0 ± 3.0	3.0 ± 4.0	–0.107	negligible	<0.000001	negative

Note: [1]The analysis was performed using "delta" measurements (difference between the raw data of the Silver projects and the bottom level for the Silver level [50 Pts] and between the raw data of the Gold projects and the bottom level for the Gold level [60 Pts]).

The replication results of the Silver-to-Gold category cross-certification performances for each of the eight US states for the six-year period (2012–2017) are presented below (Table 5). The replication results present the total number of "positive" (where the difference between the Silver and Gold category performances seems to be positive), "suspended" (where judgment about the difference between the Silver and Gold category performances is suspended), and "negative" (where the difference between the Silver and Gold category performances seems to be negative) decisions in the Silver-to-Gold category cross-certification derived from the results in Appendix A, Tables A2-A9.

According to the replication results (Table 5), EA had high performance in six states, CA, FL, IL, MA, VA, and WA, and moderate performance in two states, OH and TX. SS had moderate performance in six states, FL, IL, MA, TX, VA, and WA, and low performance in two states, CA and OH. EQ had moderate performance in five states, FL, IL, MA, TX, and WA, and low performance in three states, CA, OH, and VA. ID had moderate performance in two states, MA and TX, and low performance in six states, CA, FL, IL, OH, VA, and WA. WE had moderate performance in a single state, OH, and low performance in the other seven states. MR had low performance in all eight states. Thus, six of the eight states, CA, FL, IL, MA, VA, and WA, showed a trend for the use of the high EA-high emphasizing strategy, whereas the other two states, OH and TX, showed the use of the SS/EA/WE/EQ/ID-moderate emphasizing strategy in projects certified in 2012–2017 (Table 5).

Thus, the propelling EA category vs. the other non-propelling categories (revealed in this six-year analysis) are consistent with the results from Pushkar and Verbitsky (2018), who studied the LEED-NCv3 data in ten states under the same conditions but in a one-year certification period (2016). Pushkar and Verbitsky (2018) supposed that such high EA propelling may be because the EA performance (specifically, "EAc1 Optimize Energy Performance") is directly connected with the ASHRAE 90. regulation performance applied in the US. The authors revealed that states that had adopted ASHRAE 90.1 2010 (CA, FL, IL, MA, VA, and WA) preferred the energy-emphasizing strategy, whereas the states (NY, OH, TX, and GA) that used ASHRAE 90.1 2007 preferred the moderate several-category-emphasizing strategy.

(ii) Pooling medians of LEED projects in the median state-and-total-years group

The results of the descriptive statistics (median ± interquartile range [IQR, 25th–75th percentiles]) for the Silver and Gold projects and of an inferential statistical analysis (Cliff's δ and *P*-value) for comparing the Silver-to-Gold category cross-certification performances for the eight medians (one median for each of eight US states) are presented in Table 6.

In the SS, EA, EQ, and ID categories, the difference between the median of the Silver projects and the median of the Gold projects seems to be positive due to a large effect size and a small *P*-value. In the WE and MR categories, the difference between the median of the Silver projects and the median of the Gold projects seems to be negative due to small and medium effect sizes and high *P*-values. Consequently, four propelling categories that moved the projects from Silver to Gold certification, SS, EA, EQ, and ID, were recognized. These results are aligned with the results of the state-year-group statistical analysis, as shown in Table 5.

(iii) Pooling LEED projects in total state-and-years group

The results of the descriptive statistics (median ± interquartile range [IQR, 25th–75th percentiles]) for the Silver and Gold projects and of the inferential statistical analysis (Cliff's δ and

TABLE 5. LEED-NCv3 Silver-to-Gold category cross-certification performance: replication results for single-state-year group statistical analysis.

n	State	NFSAs	SS	WE	EA	MR	EQ	ID
1	CA[1]	Positive	0	0	6	0	0	0
		Suspended	2	2	0	0	2	1
		Negative	4	4	0	6	4	5
		Performance	low	low	high	low	low	low
2	FL[1]	Positive	2	1	4	0	1	1
		Suspended	2	0	1	1	3	0
		Negative	2	5	1	5	2	5
		Performance	moderate	low	high	low	moderate	low
3	IL[1]	Positive	2	0	4	0	0	0
		Suspended	1	0	1	2	3	1
		Negative	3	6	1	4	3	5
		Performance	moderate	low	high	low	moderate	low
4	MA[1]	Positive	1	0	4	0	1	2
		Suspended	2	1	0	1	3	2
		Negative	3	5	2	5	2	2
		Performance	moderate	low	high	low	moderate	moderate
5	OH[2]	Positive	2	2	3	1	0	1
		Suspended	0	1	1	1	0	0
		Negative	4	3	2	4	6	5
		Performance	low	moderate	moderate	low	low	low
6	TX[2]	Positive	2	1	2	0	3	3
		Suspended	2	1	2	1	1	1
		Negative	2	4	2	5	2	2
		Performance	moderate	low	moderate	low	moderate	moderate
7	VA[1]	Positive	1	1	3	2	0	0
		Suspended	3	1	2	0	2	1
		Negative	2	4	1	4	4	5
		Performance	moderate	low	high	low	low	low
8	WA[1]	Positive	2	1	5	0	2	0
		Suspended	1	0	1	0	1	1
		Negative	3	5	0	6	3	5
		Performance	moderate	low	high	low	moderate	low

Note: [1]Energy-emphasizing strategy; [2]non-energy-emphasizing strategy.

TABLE 6. LEED-NCv3 category cross-certification performance: median state-and-total-years group analysis in eight US states in 2012–2017 (sample sizes for Silver and Gold projects, $n_1 = n_2 = 8$).

Category	Median ± IQR		Inferential statistics			
	Silver level	Gold level	Cliff's δ	Effect size	*P*-value	NFSA
SS	17.8 ± 2.0	20.5 ± 2.8	–0.828	large	0.004	positive
WE	6.0 ± 0.8	6.0 ± 0.5	–0.422	medium	0.137	negative
EA	11.0 ± 2.0	16.3 ± 2.0	–1.000	large	0.0002	positive
MR	6.0 ± 1.3	6.0 ± 1.5	–0.219	small	0.493	negative
EQ	9.0 ± 0.0	10.0 ± 0.8	–0.781	large	0.006	positive
ID	4.0 ± 0.0	5.0 ± 0.3	–0.875	large	0.001	positive

P-value) for comparing the Silver-to-Gold category cross-certification performances for the eight US states and six years pooled together are presented in Table 7.

Only in the EA category did the difference between the Silver and Gold performances seem to be positive due to a large effect size and a small *P*-value. In all other categories, (SS, WE, MR, EQ, and ID), the differences seemed to be negative because there were negligible or small effect sizes despite very small *P*-values.

This means that only EA is a propelling category in the Silver-to-Gold category cross-certification performance. These results are not aligned with the results in Table 5, which shows that in addition to the EA propelling category, the SS, EQ, and ID categories are also recognized as propelling in the Silver-to-Gold category cross-certification performance.

TABLE 7. LEED-NCv3 category cross-certification performance: total states-and-years group analysis in eight US states in 2012–2017 (sample sizes: Silver, $n_1 = 913$ and Gold, $n_2 = 685$ projects).

Category	Median ± IQR		Inferential statistics			
	Silver	Gold	Cliff's δ	Effect size	*P*-value	NFSA
SS	17.0 ± 7.3	20.0 ± 6.0	–0.301	small	<0.000001	negative
WE	5.0 ± 3.0	6.0 ± 3.0	–0.172	small	<0.000001	negative
EA	11.0 ± 6.0	17.0 ± 9.0	–0.539	large	<0.000001	positive
MR	6.0 ± 2.0	6.0 ± 2.0	–0.028	negligible	0.331	negative
EQ	9.0 ± 2.3	10.0 ± 3.0	–0.230	small	<0.000001	negative
ID	4.0 ± 2.0	5.0 ± 2.0	–0.286	small	<0.000001	negative

CONCLUSIONS

The present study focused on the LEED-NCv3 Silver-to-Gold (i) cross-certification and (ii) category cross-certification performances. Eight US states (CA, FL, IL, MA, OH, TX, VA, and WA) were studied during the period 2012–2017 to evaluate these performances. The performances were evaluated by applying three types of statistical analysis: (a) the single state-year-group and replication method, (b) the median of the state-and-total-years group, and (c) the total of the states-and-years group. The following conclusions were drawn:

1. *LEED-NCv3 Silver-to-Gold cross-certification performance.* The Silver delta vs. Gold delta cross-certification has a low propelling effect. Thus, in both certifications, the eight US states showed results slightly above the relevant certification boundaries (50 pts and 60 pts for the Silver and Gold certifications, respectively). This result was confirmed by applying three types of statistical analysis: (a) the single-state-year group and replication method, (b) the median of the state-and-total-years group, and (c) the total of the states-and-years group.
2. *LEED-NCv3 Silver-to-Gold category cross-certification performance.* Several categories with different propelling effects for moving the projects from Silver to Gold certification were recognized. The EA was revealed as the most propelling category vs. the other categories that had moderate (SS and EQ) or low propelling effects (WE and MR). These findings resulted from (a) the single-state-year group and replication method and (b) the median of the state-and-total-years group analyses. However, according to (c) the total of the states-and-years group analysis, only the EA was revealed to be a propelling category moving the projects from Silver to Gold certification. Thus, in the LEED-NCv3 Silver-to-Gold category cross-certification evaluation, pooling the LEED projects in the total state-and-year group method may be misleading.

   Using the single-state-year group with the replication method shows that six of the eight states (CA, FL, IL, MA, VA, and WA) used the EA-high emphasizing strategy, while OH and TX used the SS/EA/WE/EQ/ID-moderate emphasizing strategy. This EA preferability may be related to ASHRAE 90.1, which has performance standards that are required in this category.

LIMITATIONS

Two main limitations of the study can be recognized. The first one is the LEED version that was used. We analyzed projects certified under version 3 (LEED-NC 2009) and not under the current version 4 (LEED-NC v4). This is because in empirical studies such as the present study, project certification became possible only after a relativity long time period (on average, two to three years are required to design and build a building). Thus, there was an insufficient sample size of certified projects that were designed under the current LEED-NCv4 that were revealed in the USGBC project directory.

The second limitation is the number of US states that were studied. Only eight of the 50 US states were analyzed. Again, a sufficient sample size of certified projects was only available for this six-year empirical study in the analyzed eight states.

Despite these limitations, this study may help LEED experts involved in further system development to understand the policy implications of balancing the LEED category performances to foster more category-embracing green building. In addition, this study may help LEED researchers who are performing similar empirical studies to apply more correct statistical methods, which will lead to more correct conclusions about green building strategies.

APPENDIX A

TABLE A1. LEED-NCv3 Silver-to-Gold cross-certification performance: single state-year-group analysis in eight US states in 2012–2017[1].

State	Year	Median ± IQR		Inferential statistics			
		Silver delta	Gold delta	Cliff's, δ	Effect size	*P*-value	NFSAs
CA	2012	3.0 ± 4.3	2.5 ± 4.0	0.037	negligible	0.866	negative
	2013	2.5 ± 4.0	3.0 ± 4.0	–0.061	negligible	0.705	negative
	2014	2.0 ± 3.0	4.0 ± 3.5	–0.242	small	0.076	negative
	2015	3.0 ± 2.0	2.0 ± 3.8	0.134	negligible	0.246	negative
	2016	3.0 ± 3.0	3.0 ± 6.0	–0.185	small	0.135	negative
	2017	2.0 ± 2.0	2.0 ± 4.0	–0.096	negligible	0.452	negative
FL	2012	3.0 ± 4.0	6.0 ± 12.5	–0.345	medium	0.068	suspended
	2013	1.0 ± 1.3	2.5 ± 6.5	–0.416	medium	0.035	suspended
	2014	3.0 ± 5.5	1.0 ± 3.5	0.220	small	0.268	negative
	2015	3.0 ± 2.5	3.0 ± 1.0	0.073	negligible	0.757	negative
	2016	1.0 ± 3.0	1.0 ± 6.3	–0.062	negligible	0.774	negative
	2017	1.5 ± 4.0	4.0 ± 7.3	–0.412	medium	0.073	suspended
IL	2012	2.0 ± 4.0	2.0 ± 12.0	–0.370	medium	0.249	negative
	2013	2.0 ± 3.0	3.0 ± 3.3	–0.413	medium	0.053	suspended
	2014	2.0 ± 4.0	2.0 ± 5.5	0.197	small	0.441	negative
	2015	1.5 ± 2.5	1.5 ± 4.0	–0.060	negligible	0.756	negative
	2016	1.0 ± 2.8	1.0 ± 5.0	–0.065	negligible	0.805	negative
	2017	3.0 ± 3.5	2.0 ± 6.0	0.156	small	0.600	negative
MA	2012	3.0 ± 4.0	6.0 ± 7.8	–0.257	small	0.502	negative
	2013	4.0 ± 3.8	5.0 ± 5.3	–0.450	medium	0.171	negative
	2014	5.0 ± 2.8	2.0 ± 3.8	0.613	large	0.043	positive
	2015	4.5 ± 3.0	2.0 ± 4.0	0.329	small	0.187	negative
	2016	2.0 ± 2.8	4.0 ± 5.0	–0.206	small	0.374	negative
	2017	2.0 ± 4.5	3.5 ± 6.0	–0.274	small	0.183	negative

TABLE A1. *(Cont.)*

State	Year	Median ± IQR		Inferential statistics			
		Silver delta	Gold delta	Cliff's, δ	Effect size	*P*-value	NFSAs
OH	2012	3.5 ± 4.0	1.5 ± 5.0	0.060	negligible	0.868	negative
	2013	2.0 ± 2.0	3.0 ± 4.0	–0.226	small	0.430	negative
	2014	2.0 ± 3.0	0.5 ± 1.0	0.512	large	0.022	positive
	2015	1.0 ± 2.5	0.0 ± 2.5	0.132	negligible	0.629	negative
	2016	2.0 ± 2.5	1.0 ± 1.8	0.236	small	0.385	negative
	2017	2.0 ± 2.8	3.0 ± 1.8	–0.149	small	0.567	negative
TX	2012	3.0 ± 2.3	2.5 ± 4.0	–0.019	negligible	1.000	negative
	2013	1.5 ± 3.5	4.0 ± 3.0	–0.307	small	0.155	negative
	2014	3.0 ± 3.8	6.0 ± 6.8	–0.301	small	0.115	negative
	2015	1.0 ± 1.0	1.5 ± 2.0	–0.040	negligible	0.841	negative
	2016	1.0 ± 3.0	3.5 ± 3.5	–0.389	medium	0.103	suspended
	2017	2.0 ± 2.5	4.0 ± 4.0	–0.490	large	0.026	positive
VA	2012	6.0 ± 2.5	1.0 ± 2.3	0.857	large	0.018	positive
	2013	3.0 ± 3.0	3.0 ± 2.0	0.442	medium	0.078	suspended
	2014	3.0 ± 3.0	1.5 ± 2.0	0.469	medium	0.125	suspended
	2015	2.0 ± 1.8	2.0 ± 1.8	0.017	negligible	0.971	negative
	2016	2.0 ± 3.0	3.0 ± 2.0	–0.211	small	0.230	negative
	2017	2.0 ± 2.0	2.0 ± 1.8	–0.122	negligible	0.572	negative
WA	2012	3.0 ± 3.5	4.0 ± 8.0	–0.271	small	0.427	negative
	2013	4.0 ± 2.8	2.0 ± 4.0	0.298	small	0.270	negative
	2014	2.0 ± 4.0	3.0 ± 3.5	–0.130	negligible	0.579	negative
	2015	1.0 ± 2.0	2.5 ± 4.0	–0.305	small	0.139	negative
	2016	1.0 ± 2.0	4.0 ± 4.5	–0.482	large	0.061	suspended
	2017	2.5 ± 4.0	3.0 ± 4.0	–0.036	negligible	0.918	negative

Note: [1]The analysis was performed using "delta" measurements (difference between median Silver projects and bottom level for Silver level [50 Pt] and between median Gold projects and bottom level for Gold level [60 Pt]).

TABLE A2. LEED-NCv3 Silver-to-Gold category cross-certification category performance in California.

Category	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	*P*-value	NFSAs
SS (26 pt)	2012	19.0 ± 5.5	19.5 ± 6.0	−0.092	negligible	0.661	negative
	2013	15.0 ± 8.0	17.5 ± 7.8	−0.131	negligible	0.409	negative
	2014	16.0 ± 8.0	18.0 ± 5.5	−0.086	negligible	0.533	negative
	2015	18.0 ± 8.0	20.0 ± 10.5	−0.124	negligible	0.283	negative
	2016	19.0 ± 11.8	22.0 ± 5.5	−0.393	medium	0.057	suspended
	2017	17.0 ± 11.0	22.0 ± 3.5	−0.412	medium	0.073	suspended
WE (10 pt)	2012	2.0 ± 2.0	5.0 ± 5.0	−0.119	negligible	0.566	negative
	2013	5.0 ± 2.5	7.0 ± 2.0	−0.258	small	0.102	negative
	2014	5.0 ± 3.0	5.0 ± 3.5	0.064	negligible	0.639	negative
	2015	5.0 ± 2.0	6.0 ± 2.0	−0.101	negligible	0.382	negative
	2016	5.0 ± 4.0	7.0 ± 2.0	−0.292	small	0.013	suspended
	2017	5.0 ± 3.0	7.0 ± 3.0	−0.331	medium	0.010	suspended
EA (35 pt)	2012	14.0 ± 8.5	22.0 ± 7.0	−0.670	large	0.001	positive
	2013	15.0 ± 11.5	23.0 ± 9.8	−0.544	large	0.0006	positive
	2014	12.0 ± 9.0	22.5 ± 11.0	−0.717	large	0.0000001	positive
	2015	10.0 ± 8.0	20.0 ± 12.8	−0.582	large	0.0000005	positive
	2016	12.0 ± 7.0	22.0 ± 13.0	−0.636	large	0.00000006	positive
	2017	10.0 ± 8.3	18.5 ± 15.0	−0.587	large	0.000004	positive
MR (14 pt)	2012	6.0 ± 2.0	5.5 ± 2.0	0.211	small	0.304	negative
	2013	5.0 ± 3.0	6.0 ± 1.0	−0.226	small	0.152	negative
	2014	5.0 ± 3.0	5.0 ± 2.0	−0.043	negligible	0.757	negative
	2015	5.0 ± 3.0	5.0 ± 2.0	−0.208	small	0.072	negative
	2016	5.5 ± 2.0	6.0 ± 1.0	−0.064	negligible	0.591	negative
	2017	5.0 ± 3.0	5.0 ± 2.0	−0.124	negligible	0.334	negative
EQ (15 pt)	2012	9.0 ± 2.0	11.0 ± 2.0	−0.347	medium	0.089	suspended
	2013	8.0 ± 2.0	10.0 ± 1.0	−0.398	medium	0.012	suspended
	2014	9.0 ± 2.5	9.0 ± 2.0	−0.089	negligible	0.515	negative
	2015	9.0 ± 2.0	9.0 ± 2.0	−0.044	negligible	0.706	negative
	2016	9.0 ± 3.0	9.0 ± 3.0	0.076	negligible	0.523	negative
	2017	9.0 ± 2.3	9.0 ± 5.0	−0.159	small	0.214	negative
ID (6 pt)	2012	3.0 ± 2.0	4.5 ± 2.0	−0.469	medium	0.021	suspended
	2013	4.0 ± 1.0	4.0 ± 2.0	−0.071	negligible	0.655	negative
	2014	4.0 ± 2.0	4.0 ± 2.0	−0.127	negligible	0.353	negative
	2015	4.0 ± 2.0	4.0 ± 1.8	−0.129	negligible	0.267	negative
	2016	4.0 ± 4.0	5.0 ± 1.0	−0.227	small	0.054	negative
	2017	4.0 ± 2.0	5.0 ± 2.0	−0.221	small	0.084	negative

TABLE A3. LEED-NCv3 Silver-to-Gold category cross-certification performance in Florida.

Category	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	P-value	NFSAs
SS (26 pt)	2012	17.0 ± 11.8	18.0 ± 8.0	−0.010	negligible	0.974	negative
	2013	15.0 ± 7.5	19.0 ± 8.5	−0.285	small	0.161	negative
	2014	15.0 ± 9.0	22.0 ± 4.3	−0.588	large	0.003	positive
	2015	20.0 ± 9.0	23.0 ± 2.8	−0.546	large	0.014	positive
	2016	19.0 ± 11.8	22.0 ± 5.5	−0.393	medium	0.057	suspended
	2017	17.0 ± 11.0	22.0 ± 3.5	−0.412	medium	0.073	suspended
WE (10 pt)	2012	7.0 ± 3.8	6.0 ± 1.5	−0.158	small	0.436	negative
	2013	4.0 ± 2.5	8.5 ± 5.0	−0.689	large	0.0007	positive
	2014	6.5 ± 4.0	6.0 ± 3.5	−0.107	negligible	0.594	negative
	2015	6.0 ± 4.0	7.0 ± 2.5	−0.055	negligible	0.820	negative
	2016	5.0 ± 2.0	5.0 ± 2.0	−0.053	negligible	0.808	negative
	2017	8.0 ± 4.0	8.0 ± 5.0	−0.082	negligible	0.734	negative
EA (35 pt)	2012	9.0 ± 5.3	29.0 ± 13.0	−0.780	large	0.00009	positive
	2013	13.0 ± 5.3	18.0 ± 9.0	−0.535	large	0.008	positive
	2014	12.0 ± 6.5	14.0 ± 5.3	−0.063	negligible	0.757	negative
	2015	10.0 ± 6.5	14.0 ± 5.5	−0.341	medium	0.126	suspended
	2016	10.0 ± 6.8	15.0 ± 11.0	−0.496	large	0.016	positive
	2017	11.5 ± 5.0	18.0 ± 9.3	−0.511	large	0.025	positive
MR (14 pt)	2012	6.0 ± 2.0	5.0 ± 1.5	0.161	small	0.426	negative
	2013	6.0 ± 2.3	6.0 ± 2.0	−0.045	negligible	0.836	negative
	2014	7.0 ± 2.0	7.0 ± 1.0	−0.096	negligible	0.633	negative
	2015	5.0 ± 2.5	6.0 ± 1.0	−0.350	medium	0.116	suspended
	2016	7.0 ± 1.0	7.0 ± 1.8	0.044	negligible	0.841	negative
	2017	6.0 ± 2.0	7.0 ± 1.3	−0.280	small	0.224	negative
EQ (15 pt)	2012	9.0 ± 2.8	11.0 ± 1.0	−0.345	medium	0.085	suspended
	2013	9.0 ± 2.3	10.0 ± 2.5	−0.279	small	0.171	negative
	2014	9.0 ± 2.5	10.0 ± 3.0	−0.379	medium	0.054	suspended
	2015	8.0 ± 2.5	10.0 ± 1.5	−0.400	medium	0.076	suspended
	2016	8.0 ± 4.0	9.0 ± 1.8	−0.123	negligible	0.558	negative
	2017	8.0 ± 4.0	11.0 ± 1.8	−0.506	large	0.027	positive
ID (6 pt)	2012	3.0 ± 2.0	5.0 ± 2.0	−0.273	small	0.174	negative
	2013	4.0 ± 1.5	5.0 ± 3.0	−0.250	small	0.220	negative
	2014	4.0 ± 2.0	5.0 ± 2.0	−0.472	large	0.016	positive
	2015	3.5 ± 2.5	6.0 ± 1.8	−0.327	small	0.142	negative
	2016	4.0 ± 2.0	5.0 ± 1.8	−0.314	small	0.129	negative
	2017	3.5 ± 3.0	5.0 ± 1.3	−0.302	small	0.190	negative

TABLE A4. LEED-NCv3 Silver-to-Gold category cross-certification performance in Illinois.

Category	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	*P*-value	NFSAs
SS (26 pt)	2012	11.5 ± 10.0	23.0 ± 6.0	−0.666	large	0.041	positive
	2013	15.0 ± 6.5	21.0 ± 6.3	−0.607	large	0.004	positive
	2014	19.0 ± 7.8	22.0 ± 7.5	−0.174	small	0.498	negative
	2015	22.0 ± 6.3	24.0 ± 5.5	−0.247	small	0.205	negative
	2016	19.0 ± 6.8	20.0 ± 6.5	−0.153	small	0.478	negative
	2017	21.0 ± 8.0	23.5 ± 8.0	−0.522	large	0.067	suspended
WE (10 pt)	2012	6.0 ± 4.0	6.0 ± 2.5	−0.083	negligible	0.833	negative
	2013	6.0 ± 4.0	7.0 ± 4.0	−0.113	negligible	0.604	negative
	2014	6.0 ± 1.0	6.0 ± 1.5	0.099	negligible	0.711	negative
	2015	4.0 ± 3.0	6.0 ± 2.5	−0.311	small	0.109	negative
	2016	6.0 ± 4.0	6.0 ± 2.0	−0.101	small	0.645	negative
	2017	7.0 ± 3.0	8.0 ± 1.0	−0.377	medium	0.192	negative
EA (35 pt)	2012	13.5 ± 5.0	19.5 ± 7.5	−0.583	large	0.076	suspended
	2013	8.0 ± 5.8	14.0 ± 6.0	−0.603	large	0.005	positive
	2014	11.0 ± 5.5	15.5 ± 8.5	−0.530	large	0.034	positive
	2015	8.0 ± 5.3	13.0 ± 5.0	−0.602	large	0.002	positive
	2016	10.0 ± 5.3	17.0 ± 14.5	−0.607	large	0.004	positive
	2017	12.0 ± 5.8	14.0 ± 7.0	−0.333	medium	0.254	negative
MR (14 pt)	2012	6.5 ± 2.0	5.0 ± 3.0	0.041	negligible	1.000	negative
	2013	6.0 ± 1.8	7.0 ± 0.3	−0.392	medium	0.065	suspended
	2014	6.0 ± 1.0	6.0 ± 2.0	0.023	negligible	0.951	negative
	2015	5.0 ± 2.0	7.0 ± 1.0	−0.418	medium	0.031	suspended
	2016	6.0 ± 2.5	6.0 ± 1.8	0.292	small	0.173	negative
	2017	6.0 ± 1.8	6.0 ± 0.0	−0.189	small	0.511	negative
EQ (15 pt)	2012	11.5 ± 2.0	9.0 ± 2.5	0.375	medium	0.286	negative
	2013	10.0 ± 3.8	11.0 ± 3.3	−0.113	negligible	0.604	negative
	2014	7.0 ± 3.5	9.5 ± 1.0	−0.447	medium	0.074	suspended
	2015	8.0 ± 2.0	9.5 ± 3.0	−0.452	medium	0.020	suspended
	2016	7.0 ± 3.0	9.0 ± 3.5	−0.413	medium	0.053	suspended
	2017	8.0 ± 1.8	8.5 ± 6.0	−0.100	negligible	0.737	negative
ID (6 pt)	2012	4.0 ± 3.0	4.5 ± 2.5	−0.041	negligible	0.935	negative
	2013	4.0 ± 2.0	5.0 ± 2.0	−0.211	small	0.327	negative
	2014	4.0 ± 1.8	5.0 ± 2.0	−0.401	medium	0.109	suspended
	2015	5.0 ± 1.0	5.0 ± 2.0	−0.276	small	0.156	negative
	2016	4.0 ± 2.0	4.0 ± 1.0	−0.182	small	0.398	negative
	2017	4.0 ± 0.0	4.5 ± 1.0	−0.244	small	0.421	negative

TABLE A5. LEED-NCv3 Silver-to-Gold category cross-certification performance in Massachusetts.

Category	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	P-value	NFSAs
SS (26 pt)	2012	20.0 ± 14.5	20.0 ± 7.0	–0.257	small	0.528	negative
	2013	15.0 ± 4.5	25.0 ± 2.5	–1.000	large	0.0006	positive
	2014	22.0 ± 3.3	22.0 ± 7.0	0.187	small	0.567	negative
	2015	17.0 ± 11.0	23.0 ± 4.0	–0.450	medium	0.069	suspended
	2016	21.0 ± 5.5	20.0 ± 9.0	0.033	negligible	0.884	negative
	2017	20.0 ± 5.8	23.0 ± 6.0	–0.341	medium	0.100	suspended
WE (10 pt)	2012	6.0 ± 3.8	5.0 ± 2.5	0.028	negligible	1.000	negative
	2013	5.0 ± 1.5	6.0 ± 1.5	–0.408	medium	0.203	suspended
	2014	6.0 ± 3.8	6.0 ± 3.8	–0.147	small	0.662	negative
	2015	6.0 ± 2.0	4.0 ± 4.0	0.314	small	0.209	negative
	2016	6.0 ± 2.8	6.5 ± 3.0	–0.192	small	0.362	negative
	2017	4.0 ± 2.0	5.0 ± 3.0	–0.144	small	0.492	negative
EA (35 pt)	2012	15.0 ± 7.0	23.0 ± 15.0	–0.314	small	0.432	negative
	2013	12.0 ± 2.8	12.0 ± 5.0	0.041	negligible	0.931	negative
	2014	7.0 ± 2.3	15.0 ± 5.0	–0.933	large	0.0006	positive
	2015	11.5 ± 8.0	17.5 ± 10.0	–0.557	large	0.024	positive
	2016	10.0 ± 3.8	12.5 ± 7.0	–0.692	large	0.0009	positive
	2017	11.0 ± 4.3	14.5 ± 7.0	–0.541	large	0.009	positive
MR (14 pt)	2012	6.0 ± 4.5	7.0 ± 2.5	–0.229	small	0.550	negative
	2013	5.0 ± 1.8	6.0 ± 1.5	–0.184	small	0.627	negative
	2014	6.0 ± 1.3	6.0 ± 1.8	0.106	negligible	0.766	negative
	2015	5.0 ± 1.0	5.5 ± 2.0	–0.186	small	0.463	negative
	2016	4.0 ± 2.8	6.0 ± 2.0	–0.406	medium	0.051	suspended
	2017	5.0 ± 2.0	6.0 ± 2.0	–0.148	small	0.480	negative
EQ (15 pt)	2012	9.0 ± 3.5	10.0 ± 3.5	–0.314	small	0.404	negative
	2013	8.0 ± 0.8	11.0 ± 0.8	–0.816	large	0.008	positive
	2014	9.0 ± 0.3	10.0 ± 2.0	–0.600	large	0.053	suspended
	2015	9.5 ± 3.0	9.0 ± 3.0	0.079	negligible	0.770	negative
	2016	9.0 ± 3.0	11.0 ± 2.0	–0.353	medium	0.090	suspended
	2017	7.0 ± 3.8	9.5 ± 2.0	–0.389	medium	0.060	suspended
ID (6 pt)	2012	3.0 ± 1.3	4.0 ± 1.5	–0.571	large	0.053	suspended
	2013	6.0 ± 2.5	6.0 ± 1.8	–0.061	negligible	0.918	negative
	2014	5.0 ± 1.0	4.0 ± 1.8	0.120	negligible	0.634	negative
	2015	4.0 ± 2.0	6.0 ± 1.0	–0.721	large	0.003	positive
	2016	3.0 ± 3.5	6.0 ± 1.0	–0.447	medium	0.031	suspended
	2017	3.0 ± 2.8	5.5 ± 2.0	–0.656	large	0.001	positive

TABLE A6. LEED-NCv3 Silver-to-Gold category cross-certification performance in Ohio.

Category	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	P-value	NFSAs
SS (26 pt)	2012	16.5 ± 6.0	18.0 ± 4.0	–0.119	negligible	0.705	negative
	2013	17.0 ± 4.0	21.0 ± 3.8	–0.357	medium	0.208	negative
	2014	17.0 ± 5.3	17.5 ± 4.0	–0.164	small	0.494	negative
	2015	16.5 ± 4.3	18.0 ± 6.3	–0.637	large	0.019	positive
	2016	17.0 ± 7.0	22.0 ± 1.0	–0.786	large	0.001	positive
	2017	18.0 ± 10.0	18.0 ± 6.8	–0.075	negligible	0.782	negative
WE (10 pt)	2012	6.0 ± 2.0	6.5 ± 4.0	–0.238	small	0.456	negative
	2013	6.0 ± 3.0	8.0 ± 1.8	–0.560	large	0.048	positive
	2014	6.0 ± 3.3	7.5 ± 1.0	–0.429	medium	0.064	suspended
	2015	5.0 ± 3.0	7.0 ± 1.8	–0.560	large	0.039	positive
	2016	6.0 ± 3.0	6.0 ± 2.0	–0.314	small	0.232	negative
	2017	5.0 ± 3.8	7.0 ± 2.5	–0.193	small	0.463	negative
EA (35 pt)	2012	12.5 ± 5.0	16.5 ± 6.0	–0.583	large	0.041	positive
	2013	10.0 ± 3.5	16.0 ± 6.5	–0.786	large	0.004	positive
	2014	13.0 ± 5.3	16.0 ± 2.0	–0.429	medium	0.067	suspended
	2015	14.0 ± 7.5	12.0 ± 5.5	0.044	negligible	0.892	negative
	2016	11.0 ± 4.5	13.0 ± 3.5	–0.271	small	0.306	negative
	2017	13.0 ± 3.8	17.0 ± 7.0	–0.683	large	0.005	positive
MR (14 pt)	2012	6.5 ± 2.0	8.0 ± 3.0	–0.476	large	0.102	suspended
	2013	7.0 ± 2.0	7.0 ± 2.8	–0.131	negligible	0.654	negative
	2014	6.0 ± 3.3	8.0 ± 1.0	–0.512	large	0.025	positive
	2015	6.0 ± 2.3	6.0 ± 1.0	–0.308	small	0.267	negative
	2016	6.5 ± 1.5	7.0 ± 2.3	–0.236	small	0.368	negative
	2017	7.0 ± 2.0	6.0 ± 2.8	0.099	negligible	0.717	negative
EQ (15 pt)	2012	8.5 ± 4.0	9.5 ± 3.0	–0.191	small	0.540	negative
	2013	9.0 ± 2.5	9.0 ± 2.3	0.000	negligible	1.000	negative
	2014	9.0 ± 4.0	10.0 ± 3.0	–0.094	negligible	0.697	negative
	2015	9.0 ± 2.5	10.0 ± 3.8	–0.253	small	0.378	negative
	2016	9.0 ± 3.5	9.0 ± 2.5	–0.236	small	0.378	negative
	2017	9.0 ± 3.8	10.0 ± 2.8	–0.317	small	0.217	negative
ID (6 pt)	2012	3.5 ± 2.0	4.5 ± 2.0	–0.179	small	0.551	negative
	2013	5.0 ± 2.0	4.0 ± 2.0	0.310	small	0.271	negative
	2014	4.0 ± 3.0	5.0 ± 2.0	–0.376	medium	0.102	negative
	2015	3.0 ± 2.0	5.0 ± 2.0	–0.604	large	0.029	positive
	2016	4.0 ± 2.5	4.0 ± 2.0	–0.050	negligible	0.903	negative
	2017	3.0 ± 2.0	5.0 ± 3.8	–0.280	small	0.269	negative

TABLE A7. LEED-NCv3 Silver-to-Gold category cross-certification performance in Texas.

Category	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	P-value	NFSAs
SS (26 pt)	2012	16.0 ± 8.8	16.5 ± 8.5	–0.192	small	0.602	negative
	2013	15.5 ± 6.0	18.0 ± 6.5	–0.144	negligible	0.511	negative
	2014	14.0 ± 5.8	19.0 ± 2.5	–0.793	large	0.00003	positive
	2015	13.5 ± 7.5	18.0 ± 4.0	–0.558	large	0.004	positive
	2016	12.0 ± 7.5	17.0 ± 7.3	–0.395	medium	0.083	suspended
	2017	20.0 ± 3.3	22.0 ± 0.8	–0.449	medium	0.048	suspended
WE (10 pt)	2012	5.0 ± 2.3	8.5 ± 4.0	–0.673	large	0.052	suspended
	2013	4.5 ± 2.0	5.0 ± 2.0	–0.011	negligible	0.971	negative
	2014	6.0 ± 3.0	5.0 ± 2.8	0.000	negligible	1.000	negative
	2015	6.0 ± 3.5	5.0 ± 2.0	0.232	small	0.230	negative
	2016	6.0 ± 3.8	6.0 ± 3.5	–0.218	small	0.342	negative
	2017	4.0 ± 4.0	8.0 ± 1.8	–0.811	large	0.0003	positive
EA (35 pt)	2012	13.0 ± 10.5	16.0 ± 14.0	–0.307	small	0.395	negative
	2013	12.5 ± 5.0	20.0 ± 13.0	–0.621	large	0.004	positive
	2014	13.0 ± 6.0	16.0 ± 4.3	–0.371	medium	0.051	suspended
	2015	13.0 ± 4.0	17.0 ± 3.0	–0.477	large	0.013	positive
	2016	13.0 ± 7.5	14.0 ± 5.0	–0.119	negligible	0.609	negative
	2017	12.0 ± 4.5	16.0 ± 4.8	–0.440	medium	0.053	suspended
MR (14 pt)	2012	7.0 ± 3.0	5.5 ± 1.5	0.327	small	0.387	negative
	2013	7.0 ± 2.0	8.0 ± 2.8	–0.455	medium	0.034	suspended
	2014	6.0 ± 2.8	7.0 ± 2.0	–0.011	negligible	0.963	negative
	2015	7.0 ± 1.0	7.0 ± 2.0	–0.268	small	0.165	negative
	2016	6.0 ± 2.0	7.0 ± 1.5	–0.124	negligible	0.596	negative
	2017	6.0 ± 3.5	5.0 ± 1.8	0.033	negligible	0.898	negative
EQ (15 pt)	2012	9.0 ± 2.3	10.0 ± 2.5	–0.461	medium	0.200	suspended
	2013	10.0 ± 1.5	9.0 ± 1.8	0.239	small	0.270	negative
	2014	10.0 ± 1.0	12.0 ± 2.0	–0.751	large	0.00008	positive
	2015	9.0 ± 2.0	10.0 ± 2.0	–0.477	large	0.013	positive
	2016	9.0 ± 1.0	12.0 ± 1.8	–0.568	large	0.012	positive
	2017	9.0 ± 2.0	9.0 ± 3.0	–0.066	negligible	0.784	negative
ID (6 pt)	2012	4.0 ± 2.5	5.5 ± 1.5	–0.403	medium	0.252	suspended
	2013	3.0 ± 2.5	5.0 ± 2.8	–0.132	negligible	0.545	negative
	2014	5.0 ± 1.8	6.0 ± 0.3	–0.569	large	0.003	positive
	2015	4.0 ± 2.0	6.0 ± 1.0	–0.633	large	0.001	positive
	2016	5.0 ± 2.0	5.0 ± 1.3	–0.280	small	0.221	negative
	2017	4.0 ± 1.5	6.0 ± 0.0	–0.700	large	0.002	positive

TABLE A8. LEED-NCv3 Silver-to-Gold category cross-certification performance in Virginia.

Cate–gory	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	*P*-value	NFSAs
SS (26 pt)	2012	18.0 ± 3.5	17.0 ± 4.5	0.035	negligible	0.988	negative
	2013	14.5 ± 5.0	19.0 ± 4.5	–0.408	medium	0.113	suspended
	2014	19.0 ± 5.5	15.5 ± 9.0	0.042	negligible	0.894	negative
	2015	19.0 ± 5.5	20.0 ± 3.3	–0.365	medium	0.220	suspended
	2016	18.0 ± 6.0	21.0 ± 6.8	–0.407	medium	0.043	suspended
	2017	17.0 ± 10.3	21.0 ± 3.0	–0.514	large	0.018	positive
WE (10 pt)	2012	7.0 ± 3.5	8.0 ± 3.0	0.000	negligible	1.000	negative
	2013	7.0 ± 0.0	9.0 ± 1.5	–0.508	large	0.048	positive
	2014	7.0 ± 3.5	9.5 ± 5.0	–0.417	medium	0.145	suspended
	2015	6.0 ± 2.8	7.0 ± 2.8	–0.313	small	0.304	negative
	2016	7.0 ± 4.0	7.0 ± 3.5	0.074	negligible	0.719	negative
	2017	6.0 ± 4.8	7.0 ± 3.8	–0.304	small	0.162	negative
EA (35 pt)	2012	15.5 ± 8.5	12.0 ± 3.8	0.178	small	0.685	negative
	2013	11.5 ± 4.0	15.5 ± 9.5	–0.391	medium	0.129	suspended
	2014	7.5 ± 4.5	13.5 ± 5.0	–0.729	large	0.007	positive
	2015	9.0 ± 3.0	11.0 ± 8.0	–0.426	medium	0.147	suspended
	2016	10.0 ± 5.0	15.0 ± 6.5	–0.720	large	0.0003	positive
	2017	10.0 ± 3.8	14.0 ± 4.8	–0.538	large	0.013	positive
MR (14 pt)	2012	5.5 ± 2.0	8.0 ± 3.3	–0.929	large	0.018	positive
	2013	7.0 ± 1.0	7.0 ± 2.0	0.092	negligible	0.741	negative
	2014	7.0 ± 2.0	7.0 ± 2.0	–0.031	negligible	0.956	negative
	2015	6.0 ± 1.0	7.0 ± 2.3	–0.539	large	0.061	positive
	2016	7.0 ± 1.0	7.0 ± 2.0	–0.126	negligible	0.533	negative
	2017	7.0 ± 1.8	7.0 ± 0.8	–0.008	negligible	0.985	negative
EQ (15 pt)	2012	7.0 ± 4.5	10.0 ± 2.3	–0.571	large	0.181	suspended
	2013	10.0 ± 3.0	9.0 ± 3.0	0.150	small	0.575	negative
	2014	10.0 ± 2.5	11.0 ± 4.0	–0.312	small	0.289	negative
	2015	9.0 ± 2.8	11.0 ± 2.0	–0.408	medium	0.161	suspended
	2016	9.0 ± 3.0	10.0 ± 3.0	–0.120	negligible	0.557	negative
	2017	10.0 ± 2.5	10.0 ± 3.0	–0.044	negligible	0.854	negative
ID (6 pt)	2012	4.0 ± 3.5	5.0 ± 0.0	–0.286	small	0.394	negative
	2013	4.5 ± 2.0	4.0 ± 2.0	0.042	negligible	0.895	negative
	2014	5.0 ± 1.5	5.5 ± 1.0	–0.437	medium	0.132	suspended
	2015	4.0 ± 2.8	4.0 ± 2.0	–0.052	negligible	0.896	negative
	2016	4.0 ± 3.0	5.0 ± 1.8	–0.313	small	0.119	negative
	2017	4.0 ± 3.5	4.0 ± 2.8	–0.123	negligible	0.581	negative

TABLE A9. LEED-NCv3 Silver-to-Gold category cross-certification performance in Washington.

Cate-gory	Year	Median ± IQR		Inferential statistics			
		Silver level	Gold level	Cliff's, δ	Effect size	P-value	NFSAs
SS (26 pt)	2012	17.0 ± 9.5	15.5 ± 9.0	0.083	negligible	0.829	negative
	2013	18.0 ± 3.5	22.0 ± 7.5	–0.442	medium	0.097	suspended
	2014	18.0 ± 4.8	21.0 ± 5.3	–0.463	medium	0.046	positive
	2015	22.0 ± 4.0	17.5 ± 9.0	0.251	small	0.229	negative
	2016	18.5 ± 2.0	23.0 ± 2.8	–0.863	large	0.0009	positive
	2017	20.0 ± 8.0	21.5 ± 5.0	–0.298	small	0.327	negative
WE (10 pt)	2012	5.5 ± 1.5	7.0 ± 3.0	–0.312	small	0.370	negative
	2013	5.0 ± 2.3	5.5 ± 3.5	–0.269	small	0.324	negative
	2014	6.0 ± 4.0	4.0 ± 3.3	0.082	negligible	0.737	negative
	2015	5.0 ± 2.0	8.0 ± 4.0	–0.624	large	0.003	positive
	2016	5.0 ± 2.0	5.0 ± 1.0	–0.036	negligible	0.916	negative
	2017	4.5 ± 2.0	6.1 ± 1.0	–0.345	medium	0.222	negative
EA (35 pt)	2012	12.5 ± 9.5	24.0 ± 14.0	–0.646	large	0.047	positive
	2013	13.0 ± 5.0	16.5 ± 8.0	–0.423	medium	0.115	suspended
	2014	10.0 ± 6.8	17.0 ± 4.5	–0.585	large	0.012	positive
	2015	10.0 ± 3.8	18.0 ± 7.0	–0.767	large	0.0002	positive
	2016	8.0 ± 5.0	16.0 ± 4.3	–0.891	large	0.0006	positive
	2017	10.5 ± 4.0	18.0 ± 5.0	–0.773	large	0.005	positive
MR (14 pt)	2012	5.5 ± 3.0	6.0 ± 1.0	–0.062	negligible	0.908	negative
	2013	6.0 ± 2.3	6.0 ± 1.5	0.058	negligible	0.863	negative
	2014	6.0 ± 2.0	5.0 ± 1.3	0.087	negligible	0.721	negative
	2015	5.0 ± 3.0	5.0 ± 2.0	0.192	small	0.362	negative
	2016	5.0 ± 1.0	5.0 ± 2.0	0.255	small	0.341	negative
	2017	5.0 ± 2.0	4.5 ± 3.0	0.405	medium	0.159	negative
EQ (15 pt)	2012	9.0 ± 3.3	10.5 ± 3.0	–0.583	large	0.088	suspended
	2013	9.0 ± 3.3	10.0 ± 3.0	–0.192	small	0.479	negative
	2014	8.0 ± 2.8	11.0 ± 4.3	–0.521	large	0.025	positive
	2015	7.0 ± 1.8	12.0 ± 1.0	–0.940	large	0.000006	positive
	2016	9.5 ± 2.0	10.0 ± 2.0	–0.355	medium	0.180	negative
	2017	8.5 ± 3.0	9.5 ± 4.0	–0.167	small	0.591	negative
ID (6 pt)	2012	4.0 ± 2.0	4.0 ± 1.0	0.208	small	0.658	negative
	2013	3.0 ± 1.3	4.5 ± 2.5	–0.442	medium	0.095	suspended
	2014	4.0 ± 1.0	5.0 ± 2.3	–0.217	small	0.356	negative
	2015	5.0 ± 3.8	6.0 ± 2.0	–0.188	small	0.371	negative
	2016	5.0 ± 2.0	5.0 ± 2.0	–0.155	small	0.573	negative
	2017	4.0 ± 2.0	4.5 ± 2.0	–0.369	medium	0.198	negative

REFERENCES

Beninger, P. G., Boldina, I., and Katsanevakis, S. (2012). "Strengthening statistical usage in marine ecology." *Journal of Experimental Marine Biology and Ecology* 426–427(1 September), 97–108, doi:10.1016/j.jembe.2012.05.020.

Bickel, D. R. (2003). "Robust and efficient estimation of the mode of continuous data: the mode as a viable measure of central tendency." *Journal of Statistical Computation and Simulation*, 73(12), 899–912, DOI: 10.1080/0094965031000097809.

Cheng, J. C. P. and Ma, L. J. (2015). "A non-linear case-based reasoning approach for retrieval of similar cases and selection of target credits in LEED projects." *Building and Environment*, 93(2), 349–361, DOI 10.1016/j.buildenv.2015.07.019.

Cliff, N. (1993). "Dominance statistics: ordinal analyses to answer ordinal questions." *Psychological Bulletin*, 114(3), 494–509, DOI 10.1037/0033-2909.114.3.494.

Cohen, J. (1992). "A power primer." *Psychological Bulletin*, 112(1), 155–159, DOI 10.1037//0033-2909.112.1.155.

Crandall C. S. and Sherman, J. W. (2016). "On the scientific superiority of conceptual replications for scientific progress." *Journal of Experimental Social Psychology*, 66(SEP), 93–99, DOI: 10.1016/j.jesp.2015.10.002.

Durlak, J. A. (2009). "How to select, calculate, and interpret effect sizes." *Journal of Pediatric Psychology*, 34(9), 917–28, DOI:10.1093/jpepsy/jsp004.

Fagerland, M. W. and Sandvik, L. (2009). "The Wilcoxon–Mann–Whitney test under scrutiny." *Statistics in Medicine*, 28(10), 1487–1497, DOI 10.1002/sim.3561.

Fay, M. P. and Proschan, M. A. (2010). "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules." *Statistics Surveys*, 4, 1–39, DOI: 10.1214/09-SS051.

Fuerst, F. (2009). "Building momentum: an analysis of investment trends in LEED and Energy Star-certified properties." *Journal of Retail and Leisure Property*, 8(4), 285–97, DOI 10.1057/rlp.2009.18.

Hurlbert, S. H. and Lombardi, C. M. (2009). "Final collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian." AnZF, 46(5), 311–49, DOI 10.5735/086.046.0501.

Hurlbert, S. H. and Lombardi, C. M. (2012). "Lopsided reasoning on lopsided tests and multiple comparisons." Aust New Zeal J Stat, 54(1), 23–42, DOI 10.1111/j.1467-842X.2012.00652.x.

Illankoon, I.M.C.S., Tam, V.W.Y.T., and Le, K.N. (2017b). "Environmental, Economic, and Social Parameters in International Green Building Rating Tools." *Journal of Professional Issues in Engineering Education and Practice*, 143(2), 05016010-1-05016010-8, doi.org/10.1061

Illankoon, I.M.C.S., Tam, V.W.Y.T., Le, K.N., and Shen, L. (2017a). "Key credit criteria among international green building rating tools." *Journal of Cleaner Production* 164, 209–220, doi.org/10.1016/j.jclepro.2017.06.206

LEED-NC 2009. LEED for new construction and major renovations rating system. http://www.eyal-zipuim.co.il/uploadimages/greenproduct.pdf.

Ma, J. and Cheng, J. C. P. (2016). "Data-driven study on the achievement of LEED credits using percentage of average score and association rule analysis." *Building and Environment*, 98(MAR), 121–132, DOI 10.1016/j.buildenv.2016.01.005.

Mann, H. B. and Whitney, D. R. (1947). "On a test of whether one of two random variables is stochastically larger than the other." *The Annals of Mathematical Statistics*, 18(1), 50–60, URL: http://www.jstor.org/stable/2236101.

Marx, A., Backes, C., Meese, E., Lenhof, H. P., and Keller, A. (2016). "EDISON-WMW: exact dynamic programing solution of the Wilcoxon–Mann–Whitney test." *Genomics Proteomics Bioinformatics*, 14(1), 55–61, DOI 10.1016/j.gpb.2015.11.004.

Meehl, P. E. (1990). Appraising and amending theories: "The strategy of Lakatosian defense and two principles that warrant it." *Psychological Inquiry*, 1, 108–141, DOI 10.1207/s15327965pli0102_1.

Nakagawa, S. and Cuthill, I.C. (2007). "Effect size, confidence interval and statistical significance: a practical guide for biologists." *Biological Reviews*, 82(4), 591–605, DOI 10.1111/j.1469-185X.2007.00027.x.

Picquelle, S. J. and Mier, K. L. (2011). "A practical guide to statistical methods for comparing means from two-stage sampling." *Fisheries Research*, 107(1–3), 1–13, DOI 10.1016/j.fishres.2010.09.009.

Pushkar, S. (2018). "Sacrificial Pseudoreplication in LEED Cross-Certification Strategy Assessment: Sampling Structures." *Sustainability*, 10(5), 1353, doi.org/10.3390/su10051353.

Pushkar, S. and Shaviv, E. (2016). "Using shearing layer concept to evaluate green rating systems" *Architectural Science Review*, 59(2), 114–125, doi.org/10.1080/00038628.2014.966051.

Pushkar, S. and Verbitsky, O. (2018). "LEED-NCv3 Silver and Gold certified projects in the US: An observational study." *Journal of Green Building*, 13(2), 67-83, doi.org/10.3992/1943-4618.13.2.67.

Pushkar, S. and Verbitsky, O. (2019). "Silver and Gold LEED Commercial interiors: certified projects." *Journal of Green Building*, 14, accepted.

Sandoval, P. and Prakash, A. (2016). "The Gold Rush: The popularity of the Gold Tier in LEED Certification." *Environmental Policy and Governance*, 26(6), 543–555, DOI 10.1002/eet.1724.

Sullivan, G. M. and Feinn, R (2012). "Using Effect Size—or Why the P Value Is Not Enough." *Journal of Graduate Medical Education*, (September), 279–282, DOI: 0.4300/JGME-D-12-00156.1.

USGBC website. http://www.usgbc.org/projects/.

USGBC, LEED for New Construction Projects Directory. http://www.usgbc.org/projects/new-construction.

Volker, M. A. (2006). "Reporting effect size estimates in school psychology research." *Psychology in the Schools*, 43(6), 653–672, DOI 10.1002/pits.20176.

Wu, P., Mao, C., Wang, J., Song, Y. Z., and Wang, X. Y. (2016). "A decade review of the credits obtained by LEED v2.2 certified green building projects." *Building and Environment*, 102(JUN), 167–178, DOI 10.1016/j.buildenv.2016.03.026.

Wu, P., Song, Y., Shou, W., Chi, H., Chong, H. Y, & Sutrisna, M. (2017). "A comprehensive analysis of the credits obtained by LEED 2009 certified green buildings." *Renewable and Sustainable Energy Reviews*, 68(Part 1, FEB), 370–379, DOI 10.1016/j.rser.2016.10.007.

Zhang, F., Mockus, A., Keivanloo, I., and Zou, Y. (2016). "Towards building a universal defect prediction model with rank transformed predictors." *Empirical Software Engineering*, 21(5), 2107-2145, 10.1007/s10664-015-9396-2.